STATISTICAL MODELLING FOR THE PROGNOSTIC CLASSIFICATION OF PATIENTS WITH PANCREATIC CANCER FOR OPTIMISATION OF TREATMENT ALLOCATION

by

DEBORAH DAWN STOCKEN

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Health and Population Sciences

College of Medicine and Dentistry

The University of Birmingham

March 2010

ABSTRACT

Pancreatic cancer is a common cause of cancer death and is difficult to diagnose and treat. A prognostic index is a tool that can be used in clinical practice to predict survival. Thirty six prognostic factor studies were identified but the size and statistical methods were inappropriate. Valid statistical analyses are essential to make best use of data and optimise clinical application.

Continuous variables are often simplified incorrectly by i) assuming linear relationships between predictors and log-hazard or ii) using dichotomisation. Non-linearity is addressed for the first time in this disease site using restricted cubic spline and fractional polynomial functions, ideal for smooth curved relationships. Multivariable models containing non-linear transformations gave a substantially better fit. Important effects of some covariates were unrecognised under simplistic assumptions. The fitted functions generated by the two methods were similar. A direct comparison of these strategies was based on a novel approach assessing the difference in the AIC values by calculating a sampling distribution in multiple bootstrap resamples. Model validation is also addressed for the first time in this disease and suggested minimal over-fitting with reproducible prognostic information when fitted to external data.

This thesis provides the first validated prognostic tool in advanced pancreatic cancer developed using appropriate statistical methodology. Four risk-sets identified by the model could help clinicians target treatments to patients more appropriately and have an impact on future trial design and analysis.

ACKNOWLEDGEMENTS

I am grateful to Cancer Research UK (CRUK) for their support for this research project under a Population and Behavioural Sciences project grant (reference: C19491/A6150, Lead investigator: D.D. Stocken). I am indebted to my supervisors Professor Nick Freemantle and Professor Philip Johnson and my mentor Professor Cindy Billingham, University of Birmingham, for their wisdom and belief on a daily basis. I am grateful for the support of staff at the CRUK Clinical Trials Unit, University of Birmingham, where I am proud to work and especially the encouragement from the Biostatistics Team.

I am grateful to Professor Bass Hassan and Professor Doug Altman, University of Oxford, who initiated the project and negotiated with British Biotech in providing individual patient data. The clinical input of Bass and Philip has been crucial to the understanding and clinical applicability of the statistical models. I am grateful to the Liverpool Cancer Trials Unit for providing the GEMCAP individual patient data, specifically Professor John Neoptolemos, University of Liverpool, Professor David Cunningham, Royal Marsden Hospital, London, and Dr Ian Chau, Guys Hospital, London. I am grateful to Mr Ben Bridgewater, South Manchester University Hospital, for providing individual patient data from the cardiac surgical databases of QuORU and NWQIP. I am grateful all patients who took part in these trials.

Finally, this thesis would not have been undertaken or completed without the encouragement of my family and close friends, who have supported me both emotionally and practically over the duration. This thesis is dedicated to my parents, who have always encouraged me to pursue my qualifications (but I think, Dad, it stops here …. maybe!), and my children, who have kept tabs on my progress chapter by chapter. Together with my friends, they have all kept me sane throughout! Thank you all!

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RESPONSIBILITIES

All chapters in this thesis were written by DD Stocken and reviewed by N Freemantle, PJ Johnson and LJ Billingham. DD Stocken, AB Hassan, DG Altman, PJ Johnson, N Freemantle, LJ Billingham contributed to the initial research concept, presented in a successful Cancer Research UK Population and Behavioural Sciences grant application (reference: C19491/A6150).

Chapters 1-6 contributed to a clinical paper (Stocken *et al.*, 2008) and two poster presentations (Pancreatic Society of Great Britain and Ireland annual meeting 2005, National Cancer Research Institute annual meeting 2007), authors DD Stocken, AB Hassan, DG Altman, LJ Billingham, SR Bramhall, PJ Johnson and N Freemantle with the following responsibilities: AB Hassan and DG Altman had the initial concept for the research project and approached British Biotech for the individual patient data used in the model development phase; DD Stocken was responsible for undertaking all the analysis and preparing the paper; N Freemantle was statistical supervisor for the paper, DG Altman and LJ Billingham were statistical advisors for the paper; AB Hassan, PJ Johnson and SR Bramhall were responsible for clinical interpretation of the statistical analyses. All authors contributed and approved the final version of the paper.

Chapters 5-8 contributed to a statistical paper (submitted), an oral presentation (International Society for Clinical Biostatisticians annual meeting 2010) and a poster presentation (International Society for Clinical Biostatisticians annual meeting 2009), authors DD Stocken, LJ Billingham, PJ Johnson, AB Hassan, B Bridgewater, N Freemantle with the following responsibilities: DD Stocken was responsible for undertaking all the analysis and preparing the paper; N Freemantle was statistical supervisor for the paper; LJ Billingham was statistical advisor; PJ Johnson, AB Hassan and B Bridgewater were responsible for clinical interpretation of the statistical analyses. All authors contributed and approved the final version of the paper.

CHAPTER 1: AIMS AND OBJECTIVES

The overall aim of this research was to identify groups of patients with advanced pancreatic cancer with differing prognoses and to define the factors, at the time of diagnosis, which may predict the group to which a patient best belongs. In this way, treatments could then be targeted at specific groups who have the ability to cope with and possibly benefit from further toxic treatments. There is currently no prognostic tool in routine clinical use for prediction of survival from advanced pancreatic cancer. Such a tool would help clinicians identify subgroups of patients and help in their decision making regarding appropriate treatment strategies. Historically, randomised controlled trials have stratified patients at randomisation by disease stage but generally patients are not clinically separated into prognostic groups for consideration of treatment except surgically.

Pancreatic cancer has an aggressive biological phenotype that appears to be largely resistant to therapy. As the effect of treatment is often minimal, large numbers of patients are needed to provide robust statistical evidence but sample sizes in pancreatic cancer trials tend to be small so a significant problem is how to compare treatments across trials due to case mix. Statistical models ideally need validating on external data to ensure portability to new datasets.

This research investigates potential prognostic factors for survival in patients with advanced pancreatic cancer through the construction and validation of statistical models evaluating clinical, histological and laboratory factors whilst investigating the underlying assumptions, appropriateness and suitability of such models. Based on methodologically strong statistical modelling, an attempt to identify and classify patients into defined clinical prognostic subgroups of

differing risk was carried out to i) predict disease outcome ii) identify possible factors for stratification of patients in future prospective randomised trials and iii) improve the targeting of therapeutic modalities. Recommendations for the design and analysis of future randomised controlled trials have been made, specifically identification of possible important stratification factors and the appropriateness of different statistical modelling approaches including dealing with different data types. At the time of the conception of this research project in 2003, the National Institute for Clinical Excellence recommended gathering further evidence for the role of gemcitabine as a potential new therapy in pancreatic cancer patients (NICE, 2010). As such, it was timely that significant effort be placed on the design of future studies, including the stratification of patients prior to allocation of treatments based on identification of specific risk-sets.

Prognostic modelling is based on multivariable analysis of time to event data and as such it is important to describe the statistical methods used to analyse time to event data as well as presenting clinical results. This thesis is organised into 10 chapters: Chapter 2 presents the clinical setting and background to pancreatic cancer, Chapter 3 introduces the standard statistical methods for analysis of time to event data, Chapters 4 and 5 provide an introduction to prognostic indices (Chapter 4) and the statistical methods for prognostic modelling (Chapter 5), Chapters 6 and 7 discuss the underlying assumption of linearity between a predictor and outcome (Chapter 6) and presents a comparison of methods for dealing with non-linear relationships (Chapter 7), Chapter 8 presents further investigation of these methods in a much larger dataset of cardiac surgery patients, Chapter 9 provides an external validation of identified prognostic factors in pancreatic cancer and finally Chapter 10 concludes with discussion, recommendations and direction of further research.

CHAPTER 2: CANCER OF THE PANCREAS

- **SUMMARY**
- The role of the pancreas is to make enzymes necessary to digest food and to produce insulin to enable the body to use glucose
- Pancreatic cancer is the eighth most common cause of cancer death in the world, and fifth in the western world with poor long-term survival
- Surgical resection is the only 'curative' treatment performed in approximately 10%-15% patients
- For the majority of patients with un-resectable tumours, standard treatment is the administration of gemcitabine chemotherapy
- The aim of this research is to identify prognostic factors and prognostic subgroups of patients with advanced pancreatic cancer for appropriate targeting of therapies
- Identification of important prognostic factors is desirable to enable accurate stratification of patients in randomised controlled trials

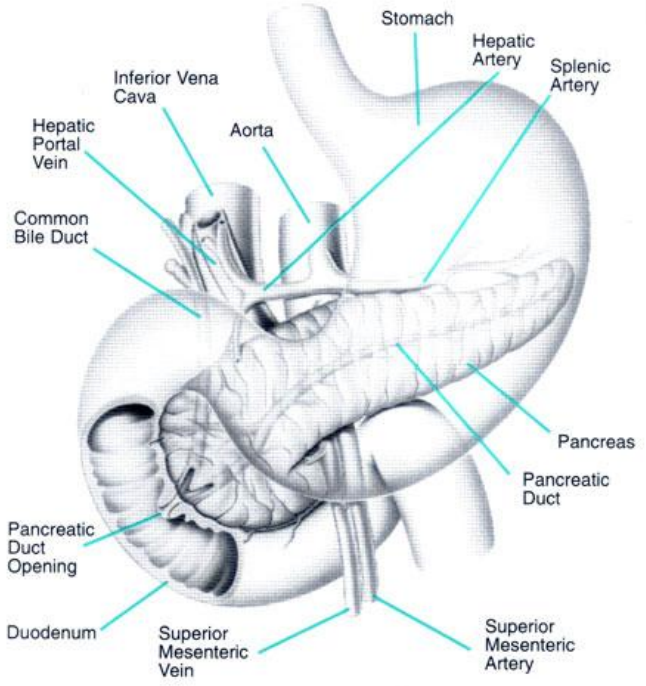
2.1. Introduction

The aim of this chapter is to provide a background to pancreatic cancer including world-wide incidence, survival and treatment. Known risk factors and treatments for pancreatic cancer are summarised and the example dataset is introduced.

2.2. The Pancreas

The pancreas is a solid gland which is 20 to 25cm in length, 4 to 6cm in width and 3 to 4cm in depth (Figure 2.1, provided by JP Neoptolemos). It is divided into 5 parts: the head, uncinate process, neck, body and tail and is attached in the back of the abdominal cavity behind the stomach.

Figure 2.1: Location of the Pancreas



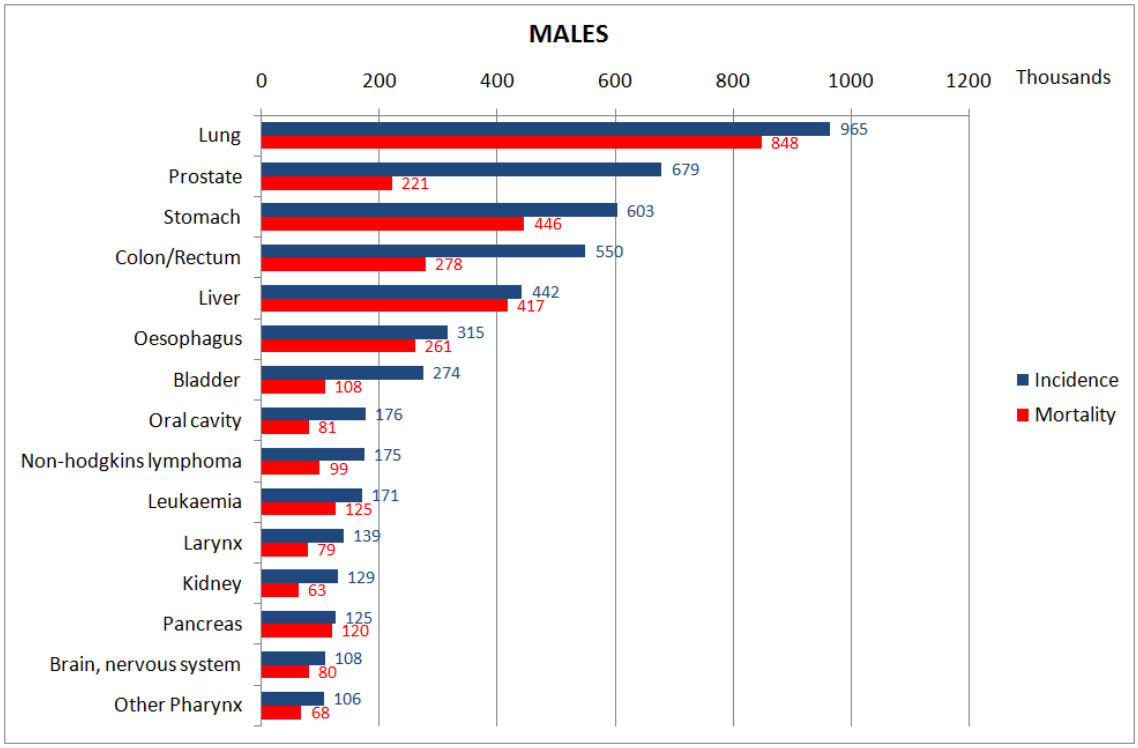
The pancreas has two roles: i) to make enzymes necessary to digest food (carbohydrates, proteins and fats) in the intestines and ii) to produce insulin to enable the body to use glucose. Digestive enzymes and insulin are made by different parts of the pancreas. There are more than 30 enzymes produced by groups of glands in the pancreas each responsible for breaking down different food types. The main enzymes are amylase (carbohydrate digestion), trypsin (digesting proteins) and lipase (digesting fat). Enzymes are collected in the main pancreatic duct and released into the duodenum where they are activated by duodenal juice. The pancreatic duct and the bile duct join together so pancreatic juice and bile can be emptied together since pancreatic enzymes and bile acid are both needed for fat digestion. Bile acid is produced by the liver and stored by the gall bladder and disperses fat before pancreatic enzymes break it down. If a pancreatic tumour blocks the main bile duct then bile acid is not released into duodenum but dispersed into blood system and out of the body through kidneys, which results in the eyes and

skin turning yellow due to a build up of bilirubin, known as obstructive jaundice. Tumours which block the pancreatic duct lead to poor digestion and weight loss. Insulin is a hormone to control the level of sugar in the blood by allowing glucose to pass into the body's cells. Glucose is required by all of the body's cells as a source of energy. Without insulin, glucose does not pass into the cells but remains in the blood causing diabetes mellitus. Insulin is made in cells mainly in the tail of the pancreas. Enzyme and insulin production are made by distinct populations of pancreatic cells.

2.3. Cancer of the Pancreas

Pancreatic cancer is one of the less common cancers ranking thirteenth in incidence worldwide, eleventh most common cancer in the UK (Cancer Research UK, 2010) with peak incidence being in the 65-75 age group (Parkin *et al.*, 2005). Pancreatic cancer accounted for 232,306 cases and 227,023 deaths in 2002 worldwide (Figure 2.2 (Parkin *et al.*, 2005), adapted). In the UK there were 7,660 new cases in 2006 and 7,727 deaths in 2007 (Table 2.1 (Cancer Research UK, 2010), adapted). As such, pancreatic cancer is the eighth most common cause of cancer death in the world and fifth in the western world (Jemal *et al.*, 2003;Parkin *et al.*, 2005). Generally pancreatic cancer presents in older patients, only about 10% of patients present aged less than 50 years.

Figure 2.2: Worldwide Incidence and Mortality by Sex and Site, 2002



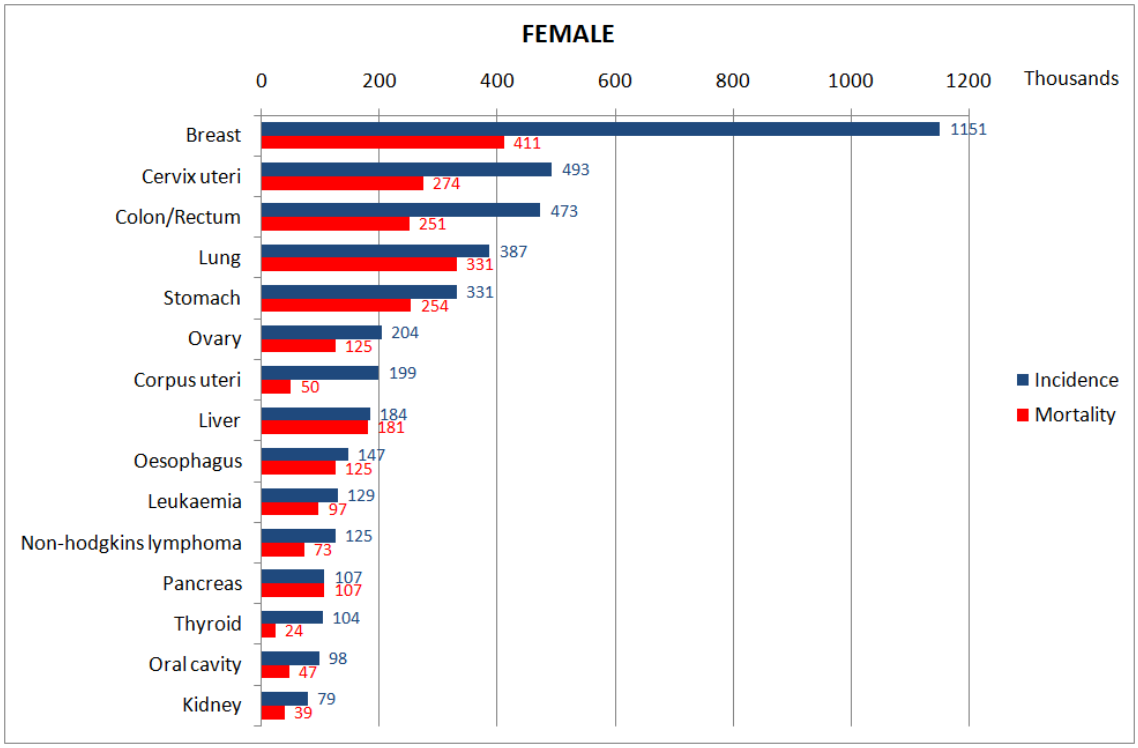


Table 2.1: Pancreatic Cancer Statistics – Key Facts

Pancreatic Cancer UK	Males	Females	Persons
Number of new cases (UK 2006)	3731	3929	7660
Rate per 100,000 population*	10.3	8.1	9.1
Number of deaths (UK 2007)	3742	3985	7727
Rate per 100,000 population*	10.0	7.9	8.9
One-year survival rate (patients diagnosed 1998 to 2001, England)	13%	13%	NA
Five-year survival rate (patients diagnosed 1998 to 2001, England)	3%	2%	NA

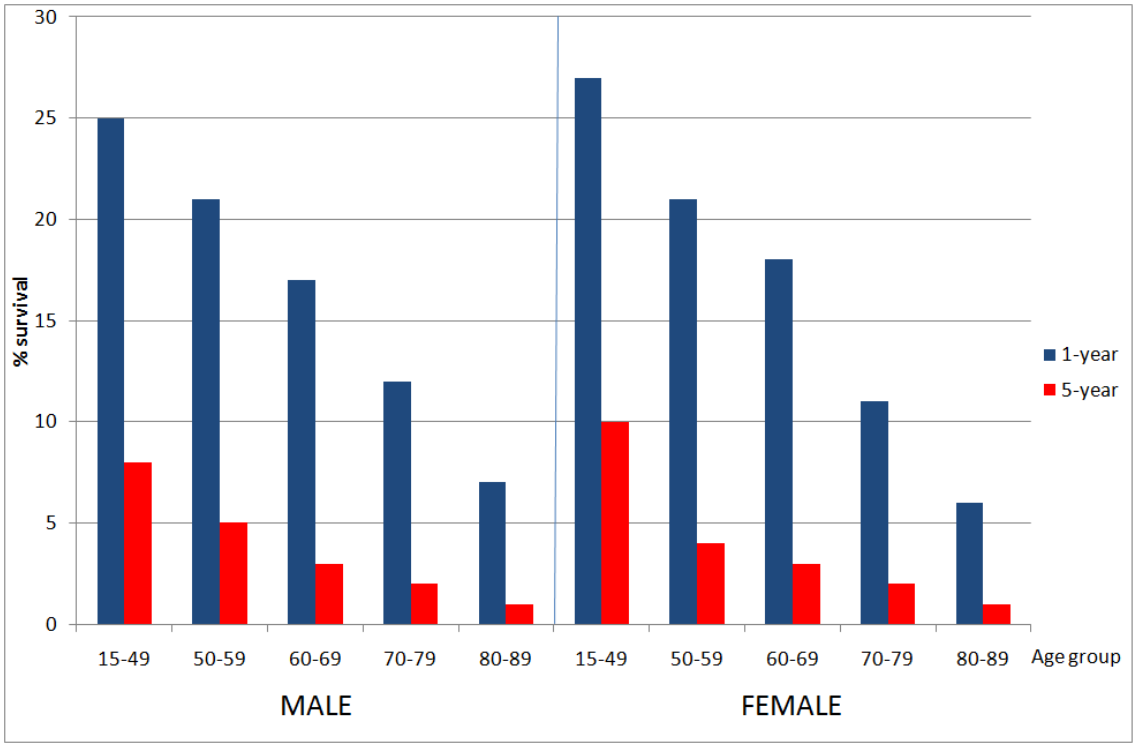
* age-standardised to the European population; NA=not available

Approximately 80% of patients present with a tumour located in the head of the pancreas. The most common type is the ductal adenocarcinoma arising from small ducts of the pancreas and accounting for 90% of all pancreatic cancers. The prognosis of pancreatic cancer patients is primarily based on resectability of a solitary tumour. Patients undergoing 'curative resection' of a single primary tumour are associated with improved survival but curative resection is only possible in approximately 10 to 15% of patients (Alexakis *et al.*, 2004;Sener *et al.*, 1999). Significant improvements in surgical outcome have been obtained with increasing specialisation (Birkmeyer *et al.*, 1999;Neoptolemos *et al.*, 1997). Most failures following surgery are due to local recurrence or liver metastases which occur within one or two years. In over 85% of patients, pancreatic cancer is diagnosed at a stage of disease too advanced for curative surgery when the disease is largely resistant to conventional cancer treatments (chemotherapy, radiotherapy).

Long-term survival remains poor with the reported overall 5-year survival rates of 0.4% to 4% (Bramhall *et al.*, 1995;Carr *et al.*, 1999;Jemal *et al.*, 2003) and 5 to 25% for patients undergoing resection (Beger, 1995;Bramhall *et al.*, 1995;Carr *et al.*, 1999;Neoptolemos *et al.*, 2004;Stocken

et al., 2005;Yeo, 1998) with higher rates achieved in patients receiving treatment at specialised centres (Gordon *et al.*, 1995;Sosa *et al.*, 1998). For patients diagnosed in England and Wales, approximately 13% of patients survive longer than 12 months from diagnosis and only 2 to 3% survive longer than five years (Figure 2.3 (Cancer Research UK, 2010), adapted). Further benefits in survival are anticipated by identifying risk groups by which to target treatments.

Figure 2.3: One- and five- year Relative Survival by Age and Sex, Patients Diagnosed with Pancreatic Cancer, England, 1998 to 2001, followed to the end of 2003



2.4. Risk Factors

Tobacco smoking is the single most defined risk factor and is said to account for up to one third of pancreatic cancers: smokers having double the risk compared to non-smokers (Doll *et al.*, 1994;Lin *et al.*, 2002;Silverman *et al.*, 1994) based on a dose-response relationship. Two benign

diseases: pancreatitis and diabetes, are associated with increased risk of pancreatic cancer. Patients with hereditary pancreatitis have a cumulative risk of pancreatic cancer at age 70 of 40% increasing to 75% with paternal inheritance (Howes *et al.*, 2004;Lowenfels *et al.*, 1997). Gene mutations have been identified which now allows screening of these families. Onset of type II diabetes has been associated with onset of pancreatic cancer and is present in over 50% of pancreatic cancer patients, the majority being diagnosed within two years (Everhart and Wright, 1995;Huxley *et al.*, 2005). The risk of developing pancreatic cancer in patients having diabetes for less than five years was doubled (odds ratio=2.1, CI=1.9 to 2.3) (Huxley *et al.*, 2005).

2.5. Surgical Treatment

Surgical resection of a solitary tumour is the only 'curative' treatment which can improve the outlook for patients (Wagner *et al.*, 2004) but only 10% to 15% patients are eligible for the procedure due to advanced disease. The first successful resection was performed by German surgeon Kausch in 1909 as a two-stage operation. In 1942 Whipple perfected the procedure into a one-stage resection, similar to the operation performed today. The Kausch-Whipple resection is associated with post-operative weight loss. The modified pylorus-preserving Whipple resection preserves the stomach, pylorus and part of the duodenum and although less radical, is not associated with decreased survival (Bassi *et al.*, 2005). Removal of a pancreatic cancer by resection is a major procedure with low surgical mortality, 2% if performed by experienced surgeon (Cameron *et al.*, 1993) but with a high complication rate of 40% (Bassi *et al.*, 2005).

When resection is not possible, palliative surgery (biliary bypass) may relieve biliary or duodenal obstruction which can cause jaundice and pain. Older patients who may not be suitable, may have their jaundice relieved by inserting a tube (stent) through the tumour. If surgery is not performed

then symptoms of jaundice, pain, weight loss, pancreatic insufficiency, fatigue and depression need to be palliated medically.

2.6. Chemotherapy and Radiotherapy Treatments

Only two chemotherapy single agents have produced reproducible survival benefit: 5-fluouroracil (5-FU) and gemcitabine. These single agents have also been studied in combination with other drugs, notably folinic acid (FA) with 5-FU and platin based agents with gemcitabine. For those patients who undergo resection of their tumour, standard treatment in the UK is administration of adjuvant (postoperative) 5FU/FA chemotherapy, following the first large (n=289) randomised controlled trial (ESPAC-1) investigating the roles of both chemotherapy and radiotherapy treatments (Neoptolemos *et al.*, 2004;Neoptolemos *et al.*, 2001a;Stocken *et al.*, 2005). Final analysis of the ESPAC-1 trial reported a 29% decreased risk of death with chemotherapy (hazard ratio=0.71, 95%CI: 0.55 to 0.92) (Neoptolemos *et al.*, 2004), a benefit confirmed in a subsequent meta-analysis (hazard ratio=0.75, 95%CI: 0.64 to 0.90) (Stocken *et al.*, 2005) and in analysis of composite data with the continuation ESPAC-3 trial (hazard ratio=0.70, 95%CI: 0.55 to 0.80) (Neoptolemos *et al.*, 2009b). For the majority of patients with un-resectable tumours, standard treatment in the UK is the administration of gemcitabine (NICE, 2010), following a randomised controlled trial comparing gemcitabine and 5-FU (Burris *et al.*, 1997). Clinical benefit response (a composite of measurements of pain, performance status and weight) was used as the primary measure of efficacy and the trial concluded gemcitabine to be more effective in alleviating disease related symptoms in a total of 126 patients.

Recently two large, international, phase III trials compared gemcitabine (the standard treatment from the advanced setting) as an experimental treatment in the adjuvant setting; one compared

with a control arm (Oettle *et al.*, 2007) and one compared against the standard 5FU/FA treatment (Neoptolemos *et al.*, 2009a). A significant benefit for adjuvant chemotherapy compared with control was seen in 368 patients based on disease free survival (DFS) (Oettle *et al.*, 2007). Median DFS for adjuvant gemcitabine was 13.4 (95%CI: 11.4 to 15.3) months compared with a median DFS of 6.9 (95%CI: 6.1 to 7.8) months for control, affirming the need for a trial comparing 5FU/FA and gemcitabine. ESPAC-3 concluded that there was no significant survival difference between adjuvant gemcitabine and adjuvant 5FU/FA in 1088 patients (hazard ratio=0.94, 95%CI: 0.81 to 1.08) but that the toxicity profile was more favourable for gemcitabine (Neoptolemos *et al.*, 2009a). This highlights how evidence for treatment in the advanced setting can pave the way for treatment to be investigated in the adjuvant setting and thus the need for all evidence to be robust and trustworthy. Similarly, a recent, large, Cancer Research UK trial (GEMCAP) including a meta-analysis comparing standard gemcitabine treatment in the advanced setting with combination gemcitabine and capecitabine treatment has concluded an advantage for the combination treatment (meta-analysis hazard ratio=0.86, 95%CI: 0.75 to 0.98) (Cunningham *et al.*, 2009) paving the way for the ongoing Cancer Research UK adjuvant trial (ESPAC-4) comparing the gemcitabine and capecitabine combination against gemcitabine alone.

Use of adjuvant radiotherapy is not universal due to results of the first large (n=289) randomised controlled trial (ESPAC-1) showing a 28% increased risk of death in patients receiving adjuvant chemo-radiotherapy (hazard ratio=1.28, 95%CI: 0.99 to 1.66) (Neoptolemos *et al.*, 2004). Initially defined in a ground-breaking study, the north American gastrointestinal study group (GITSG) reported a survival benefit for adjuvant chemo -radiotherapy in a study of 43 patients where median survival was 20 months in patients receiving chemo-radiotherapy compared with 11 months without (Kalser and Ellenberg, 1985), hazard ratio not presented but estimated as 0.54

(95%CI: 0.27 to 1.05) (Stocken *et al.*, 2005). Though the trial has been criticised for being small and with uncertainty surrounding the estimated treatment effect (wide confidence interval surrounding the estimated hazard ratio), it was regarded as sufficient evidence for the American National Cancer Institute (NCI) to recommend surgery plus chemo-radiotherapy as standard treatment in the United States. However the study required confirmation before being adopted as standard of care in Europe but the first large (n=289) randomised controlled trial has since shown no survival benefit for chemo-radiotherapy (Neoptolemos *et al.*, 2004;Neoptolemos *et al.*, 2001a).

Neoadjuvant treatment is treatment given prior to any planned surgery to improve resectability of the primary tumour and minimising any remaining residual disease. Neoadjuvant treatment could be chemotherapy, radiotherapy or a combination but to date early phase studies (Palmer *et al.*, 2007) have not led to any large phase III trial.

2.7. Example Dataset in Advanced Pancreatic Cancer

This thesis is predominantly based on analysis of data from patients with advanced pancreatic cancer. Initially, statistical modelling was carried out on a dataset collated by British Biotech combining data from two international studies (Bramhall *et al.*, 2002;Bramhall *et al.*, 2001). External validation was carried out using an external dataset from the Cancer Research UK GEMCAP trial (Cunningham *et al.*, 2009). However, during the research, an additional dataset of cardiac surgery patients became available (Pagano *et al.*, 2009) and due to suitability further methodologically driven analyses were able to be carried out on these data. The initial dataset from 653 advanced pancreatic cancer patients collated by British Biotech is described below.

Two international phase III randomised controlled trials (BB128 (Bramhall *et al.*, 2001) and BB193 (Bramhall *et al.*, 2002)) randomised 414 and 239 patients with advanced pancreatic cancer respectively: BB128 randomised patients between marimistat and gemcitabine; BB193 randomised patients between marimistat with gemcitabine and gemcitabine alone. The studies had similar eligibility criteria: histologically or cytologically unresectable pancreatic cancer, within 8 weeks of diagnosis or disease recurrence and Karnofsky performance status (KPS) of >50% (BB128) or >60% (BB193). Previous therapy for metastatic or locally advanced disease was an exclusion criteria. The primary outcome measure in both studies was survival time calculated from the date of randomisation to date of death from any cause. Randomisation was stratified by cancer stage (stage I/II, III or IV), KPS (50 to 70%, 80 to 100%), sex and study centre. On average patients in the two trials were randomised 20 and 15 days after diagnosis and started treatment the day following randomisation. The average age of patients was 63 years (range 29 to 89), 368 (56%) were male, 439 (68%) had cancer stage IV disease, 436 (67%) presenting with metastases and 251 (39%) had lymph node involvement (Table 2.2). No significant survival benefit for marimastat over gemcitabine was seen in the BB128 trial (p=0.19) (Bramhall *et al.*, 2001). Similarly no significant survival benefit was seen for combination gemcitabine and marimistat over gemcitabine alone in the BB193 trial (p=0.95) (Bramhall *et al.*, 2002).

Table 2.2: Patient Characteristics by Trial

Variable		BB128 N=414 (63%)	BB193 N=239 (37%)	TOTAL N=653 (100%)
DEMOGRAPHICS				
* Age at entry (years)	Median	63	62	63
	Range	29 to 89	32 to 85	29 to 89
Ethnic race	White	364 (88%)	226 (95%)	590 (90%)
	Black	27 (6%)	8 (3%)	35 (6%)
	Oriental	7 (2%)	0	7 (1%)
	Other	15 (4%)	5 (2%)	20 (3%)
	Missing	1	0	1
Sex	Male	228 (55%)	140 (59%)	368 (56%)
	Female	186 (45%)	99 (41%)	285 (44%)
Treatment	Gemcitabine	103 (25%)	119 (50%)	222 (34%)
	Marimistat	311 (75%)	120 (50%)	431 (66%)
TUMOUR INFORMATION				
Cancer stage	I	19 (4%)	13 (5%)	32 (5%)
	II	45 (11%)	27 (11%)	72 (11%)
	III	76 (19%)	28 (12%)	104 (16%)
	IV	268 (66%)	171 (72%)	439 (68%)
	Missing	6	0	6
Distant Metastases	M0	129 (31%)	65 (27%)	194 (30%)
	M1	265 (64%)	171 (72%)	436 (67%)
	Missing	20 (5%)	3 (1%)	23 (3%)
Regional lymph nodes	N0	153 (37%)	90 (38%)	243 (37%)
	N1	164 (40%)	87 (36%)	251 (39%)
	Missing	97 (23%)	62 (26%)	159 (24%)
Primary tumour T stage	T0	5 (1%)	3 (1.5%)	8 (1.5%)
	T1	114 (30%)	44 (20%)	158 (26%)
	T2	90 (24%)	54 (25%)	144 (24%)
	T3	167 (44%)	113 (53%)	280 (47%)
	T4	6 (1%)	1 (0.5%)	7 (1.5%)
	Missing	32	24	56
SERUM CHEMISTRY and HAEMATOLOGY				
Laboratory Variables		Median (Range), Missing	Median (Range), Missing	Median (Range), Missing
* AST (SGOT) IU/L		24 (6 to 365), 17	26 (9 to 538), 12	25 (6 to 538), 29
* TOTAL BILIRUBIN μmol/L		13.7 (3.4 to 277.0), 16	13.7 (3.0 to 135.1), 8	13.7 (3.0 to 277.0), 24
* ALK PHOSPHATASE IU/L		136 (36 to 1660), 16	157 (35 to 2064), 8	140 (35 to 2064), 24
* ALBUMIN g/L		38 (22 to 47), 17	38 (24 to 47), 8	38 (22 to 47), 25
* LDH IU/L		163 (77 to 1074), 21	169 (29 to 1495), 11	164 (29 to 1495), 32
* BUN mmol/L		9.2 (2.9 to 34.3), 17	9.3 (4.3 to 27.9), 16	9.3 (2.9 to 34.3), 33
* CA19/9 KU/l		686 (5 to 1,000,000), 17	800 (8 to 1,000,000), 30	710 (5 to 1,000,000), 47
* HAEMOGLOBIN g/dL		12.5 (5.5 to 16.1), 28	12.4 (8.3 to 19.1), 13	12.4 (5.5 to 19.1), 41
* WBC 10^9/L		7.6 (2.3 to 31.6), 28	8.3 (2.4 to 23.7), 13	7.9 (2.3 to 31.6), 41
OUTCOME				
Event indicator	Alive	22 (5%)	19 (8%)	41 (6%)
	Dead	392 (95%)	220 (92%)	612 (94%)
* Follow-up of alive (months)	Median	20.1	19.4	20.7
	Range	0.9 to 24.6	1.9 to 23.3	0.9 to 24.6

* = continuous measurements

2.8. Conclusion

The British Biotech dataset provides a representative dataset of patients with advanced pancreatic cancer. The dataset combined two randomised controlled trials both based on similar eligibility criteria and both showing no differences in survival of the experimental treatment group. The dataset contains eighteen baseline clinical, histological, biochemical and demographic variables (including trial and treatment variables) considered appropriate for analysis, eight of which are categorical variables and 10 which are continuous measurements. It is a robust dataset containing a high event rate and long follow-up of alive patients. Appropriate statistical methods for the analysis of this kind of time to event data are described in the next chapter.

CHAPTER 3: SURVIVAL AND HAZARD FUNCTIONS

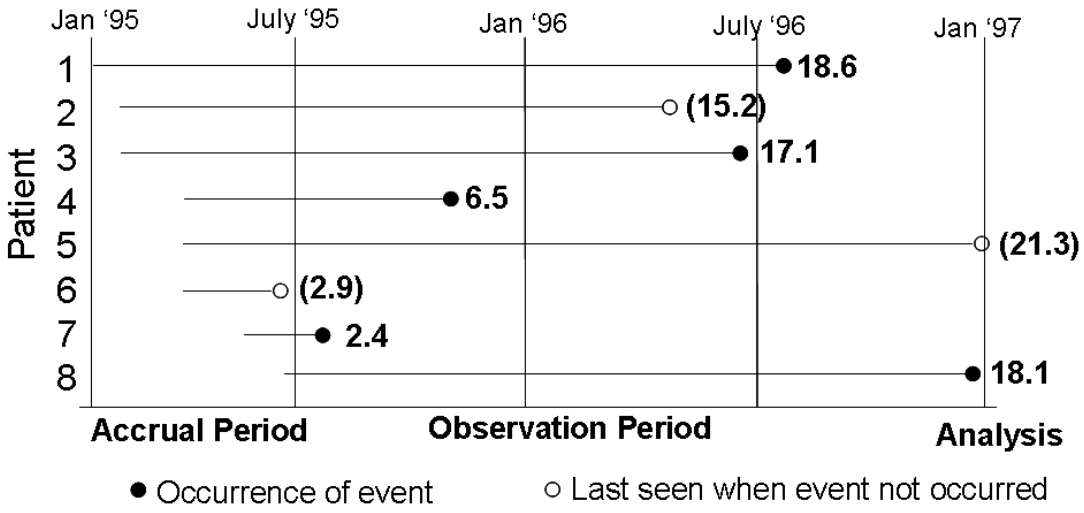
- **SUMMARY**
- Time to event data differs from continuous data due to censoring
- Statistical analyses should utilise methods which account for censored data
- Survival data can be summarised through estimates of the survivor function and the hazard function estimated from observed data
- Kaplan-Meier estimates of survival account for death times and censored survival times and can be compared using the log-rank non-parametric test
- The hazard ratio is the ratio of hazard rates and is used as a summary of the risk in one group of patients compared to another
- Many survival analysis methods assume the hazard ratio is constant over time (proportional hazards)

3.1. Introduction

The aim of this chapter is to introduce the basic statistical methods used to analyse time to event (or survival) data. Typically patients are entered into a study over a period of time and followed post-treatment for assessment of the endpoint of interest, such as death, disease progression and so on. Time to event is usually defined as the time from entry into the study to time of the endpoint of interest. Not all patients in the study will experience the endpoint of interest and it is impractical and unethical to delay analysis until all patients have experienced this endpoint. Some patients may not have experienced the event since they i) have become 'lost to follow-up' in the study through withdrawing from the study, ii) are deemed to be 'cured' or iii) are entered later in recruitment phase with shorter but unbiased follow-up. Patients who are lost to follow-up or who do not experience the event of interest are (right) censored in the analysis and are treated differently than just patients with missing data. Censored patients are known to have 'survived'

event-free for a certain length of time but the time of their event is unknown and is beyond the date they were last seen event free, known as the censor date (Figure 3.1 (Machin *et al.*, 2006), adapted):

Figure 3.1: Censored Survival Data



Characteristically time to event data differs from continuous data due to this censoring. Censored survival times hold information about a patient's survival time up to the date of censor and should not be ignored just because the patient has not experienced the event of interest. Statistical analysis should utilise methods which account for censored data to avoid losing this important information and this makes analysis of time to event data more complex. Most methods assume the reason for censoring is non-informative, which assumes that censoring is not associated with outcome, else informative censoring is said to have occurred and standard statistical methods are invalidated. In this research, the reason for censoring is assumed to be non-informative. Of the 653 patients randomised, 41 patients (6%) were alive with median follow-up 20.7 months (Table 2.2). Minimum follow-up was 0.9 months and as such all patients had some follow-up (ranging to a

maximum of 24.6 months). Some drop-out in clinical trials is expected. A minimal number of advanced pancreatic cancer patients (5 patients(<1%)) were censored before 12-months of follow-up, a clinically relevant and reported time-point in advanced pancreatic cancer. Reasons for drop-out of the censored patients were not available. However, it is likely that drop-out is related to the fact that these are poorly patients, receiving toxic chemotherapy treatment with average age 63 years, but not related to death and as such does not invalidate the assumption of non-informative censoring.

3.2. Survivor and Hazard Functions

The distribution of survival data is generally described by two functions: the survivor function and the hazard function (Collett, 1994). The survivor function *S(t)* is the probability that a patient survives from the time or origin (entry to trial) to sometime beyond time *t* :

$$S(t) = p(T > t) = 1 - F(t) = 1 - \int_{u=0}^{t} f(u)du$$

where *T* is a random variable associated with survival time with a probability density function *f(t)* and associated cumulative distribution function *F(t)*, defined as the probability that survival time is less than some value *t* .

The hazard function *h(t)* is approximately the probability of death between time *t* and *t+δt* , conditional on having survived up to time *t* :

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t \mid T \ge t)}{\delta t} \right\}$$

where the probability that random variable *T* lies between time *t* and *t+δt* is conditional on *T* being $\geq t$. *h(t)* then represents the instantaneous rate of death .

The survivor, hazard and probability density functions can be estimated from each other and as such the hazard function can be presented as a as a conditional failure rate and as a function of survival:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\{\log S(t)\}$$

so $S(t) = \exp\{-H(t)\}$

where *H(t)* is the cumulative hazard $H(t) = \int_{u=0}^{t} h(u)du$

and $H(t) = -\log S(t)$

Survival data can be summarised through estimates of the survivor function and the hazard function and can be estimated from observed data using non-parametric methods if the form of *f(t)* is not specified, or else parametrically.

3.3. Estimation of the Survivor Function

The overall probability of survival to time *t* (Machin et al., 2006) is:

$S(t) = p_1 \times p_2 \times p_3 \times \ldots \times p_t$

where p_1 is the probability of surviving at least to time 1, p_2 is the conditional probability of surviving time 2 having survived time 1 and so on where time could be in days, months, years depending on the disease in question. For any time *t*:

p_t = *[number of patients followed for at least (t-1) and who survive time t]* / *[number of patients alive at the end of time (t-1)]*

Deaths occur at distinct times $t_1 < t_2 < t_3 \ldots t_n$ where t_n is the maximum death time observed. It then follows:

$p_t = \frac{(n_t - d_t)}{n_t}$ *or* $p_t = 1 - \frac{d_t}{n_t}$

where n_t denotes the number of patients alive at the start and at risk of death during short interval (t, t+1), d_t denotes the number of deaths during the interval hence the number of patients surviving being n_t-d_t. As such the value of *S(t)* only changes when $d_t \neq 0$ so:

$$S(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right)$$

S(1), S(2), …., S(t) are known as the Kaplan-Meier estimates of survival, also known as the product-limit estimates (Kaplan and Meier, 1958). As shown, the Kaplan-Meier estimate of survival accounts not only for death times but also censored survival times utilising all information. A useful summary of survival data is a plot of the estimates of survival *S(t)* against *t* from which the median (where S(t)=0.5) and other percentiles of the distribution of survival times can be estimated. Survival curves are plotted as step functions since estimates of survival are assumed constant between death times. The number of patients remaining at risk should be indicated on the plot to provide information regarding the reliability of the survival estimates over time which are influenced by large proportions of censored patients. Confidence intervals for the survivor function may provide an indication of reliability of the survival estimates but are not a very efficient way of assessing the data being estimated at fixed time points. 95% confidence intervals can be calculated assuming a Normal distribution as $S(t) \pm 1.96 * se[S(t)]$. The standard error (se) of the survivor function can be calculated using different methods (Collett, 1994;Machin *et al.*, 2006). The LIFETEST procedure in (SAS Institute Inc., 1999) calculates Kaplan-Meier estimates of survival with standard errors calculated using Greenwood's formula (Kalbfleisch and Prentice, 1980) producing summary statistics and plots of survival estimates.

3.4. Comparison of Survivor Functions

A test statistic measures the extent to which observed data depart from a null hypothesis, usually of no difference, across groups. The probability of obtaining data as extreme, or more extreme, when the null hypothesis is true provides the strength of evidence known as the p-value. A large p-value indicates it is likely to obtain the observed data if the null hypothesis, of no difference between groups, was true. A smaller p-value indicates increasing evidence against the null hypothesis to a point where the null hypothesis could be rejected. Although the p-value provides a measure of evidence to reject the null hypothesis, often the important information in prognostic factor studies is the estimation of prognostic ability of covariates including the actual direction of the estimated effect which can obtained from descriptive statistics (e.g. median survival estimates) and associated confidence intervals.

The log-rank non-parametric test (also called the Mantel-Cox test) is the most widely used test to quantify the extent of any difference between groups of survival data containing censored observations. The advantage of this test is that it compares survival estimates across the duration of the curve not at a single particular time point, which can be misleading and inefficient. The log-rank test statistic (Collett, 1994) is based on comparing expected number of deaths under the null hypothesis of no difference between groups with the observed number of deaths at each of the successive distinct death times $t_1 < t_2 < t_3 \ldots t_n$ and is calculated as:

$$\chi^2_{\log-rank} = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}$$

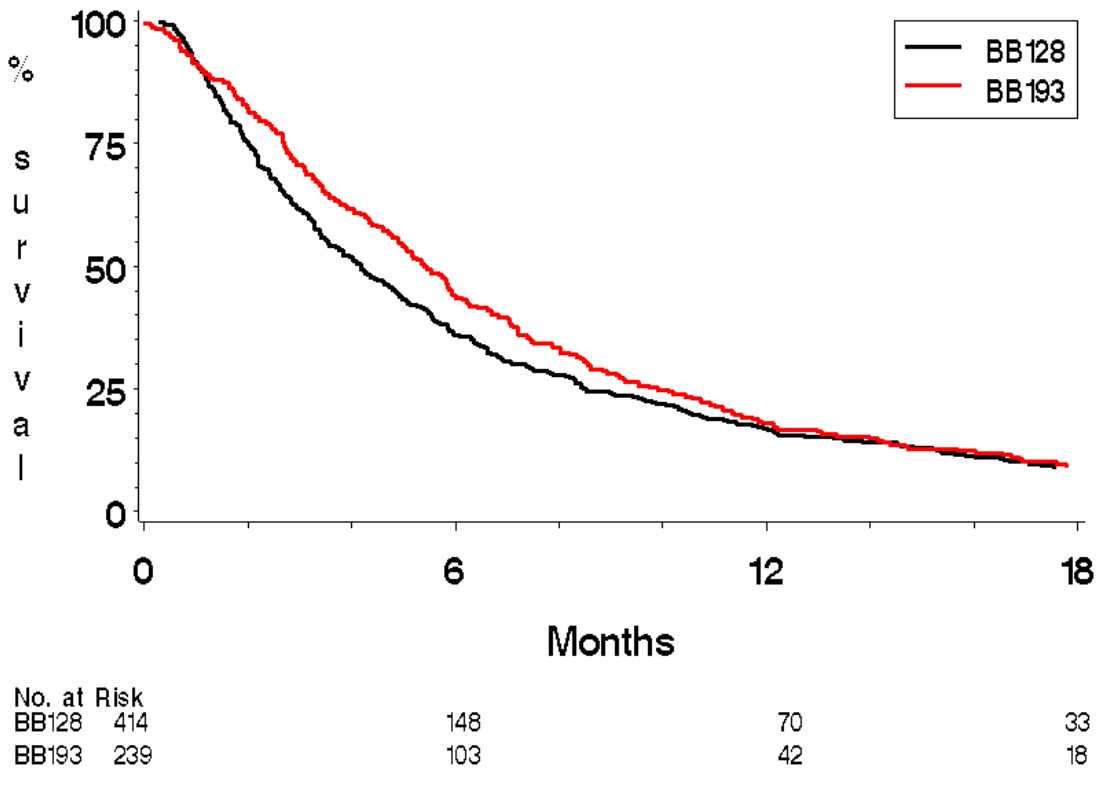
where g is the number of groups being compared. The log-rank test statistic is compared against percentage points of a chi-square distribution with g-1 degrees of freedom and determined to be significantly different or not at a particular significance (error) level, usually 5%. Other alternative

versions of the log-rank test are available when the number of events is small or the assumption of proportional hazards is violated (Machin *et al.*, 2006).

3.5. Estimation of the Survivor Function in Advanced Pancreatic Cancer

A total of 653 patients with advanced pancreatic cancer were randomised. The majority of patients (612, 94%) had died by the time of analysis with a median follow-up time of 21 months for the 41 patients still alive. The median survival estimate for the overall group was 4.7 months (95%CI: 4.2 to 5.1) with 12-month survival estimate of 17.4% (95%CI: 14.5% to 20.3%). Median survival estimates were 4.2 (95%CI: 3.6 to 4.8) and 5.4 (95%CI: 4.8 to 6.0) for BB128 and BB193 respectively with corresponding 17.0% (95%CI: 13.4% to 20.6%)and 17.9% (95%CI: 13.0% to 22.8%)12-month survival estimates. Survival estimates did not differ significantly at the 5% level across the two trials (χ^2_{LR}=2.28, p=0.13, Figure 3.2):

Figure 3.2: Survival Function by Trial



3.6. Estimation of the Hazard Function

The risk of death at a particular time point *t*, or hazard rate *h(t)*, can be estimated within specific small time intervals and is the risk of death within the interval given that a patient has survived to the start of that interval. It is estimated as a ratio of the number of deaths divided by the number at risk of death within the interval based on individual patient follow-up. There are various methods but the Kaplan-Meier method again defines time intervals based on actual death times at distinct times $t_1 < t_2 < t_3 \ldots t_n$ where t_n is the maximum death time observed. The Kaplan-Meier estimate of the hazard function (Collett, 1994) at time *t* is:

$$h(t) = \frac{d_j}{n_j \tau_j}$$

where d_j is the number of deaths in the interval $t_j \leq t \leq t_{j+1}$, n_j is the number of patients at risk at time t_j and τ_j is the width of the interval $t_j \leq t \leq t_{j+1}$ assuming the hazard function is constant between successive death times as with the Kaplan-Meier estimate of the survivor function.

3.7. The Hazard Ratio

At any time *t* the hazard ratio for two groups of patients is the ratio of the hazard rates at that time. The overall hazard rate over the whole follow-up period for a specific group of patients can be calculated using the observed and expected values calculated using the log-rank method. The hazard ratio (HR) is simply a ratio of the two hazard rates estimated as:

$$HR = \frac{O_A / E_A}{O_B / E_B}$$

where O_A and O_B are the observed numbers of deaths in groups A and B respectively and E_A and E_B are the expected numbers of deaths in groups A and B calculated using the log-rank method. A hazard ratio = 1 indicates equal hazard rates in the two groups. A hazard ratio < 1 indicates the

hazard of death is less for a patient in group A relative to a patient in group B. For the situation where there are only two groups, the log-rank test is testing the null hypothesis that the hazard ratio = 1. 95% confidence intervals for the hazard ratio cannot be calculated assuming a Normal distribution since possible values of the *HR* range from 0 to ∞. Log *HR* is approximately normally distributed which enables a 95% confidence interval to be based on: *log HR* $\pm$ *1.96*se[log HR]*. The standard error (se) of log *HR* can be calculated using:

$$se[\log HR] = \sqrt{\left(\frac{1}{E_A} + \frac{1}{E_B}\right)}$$

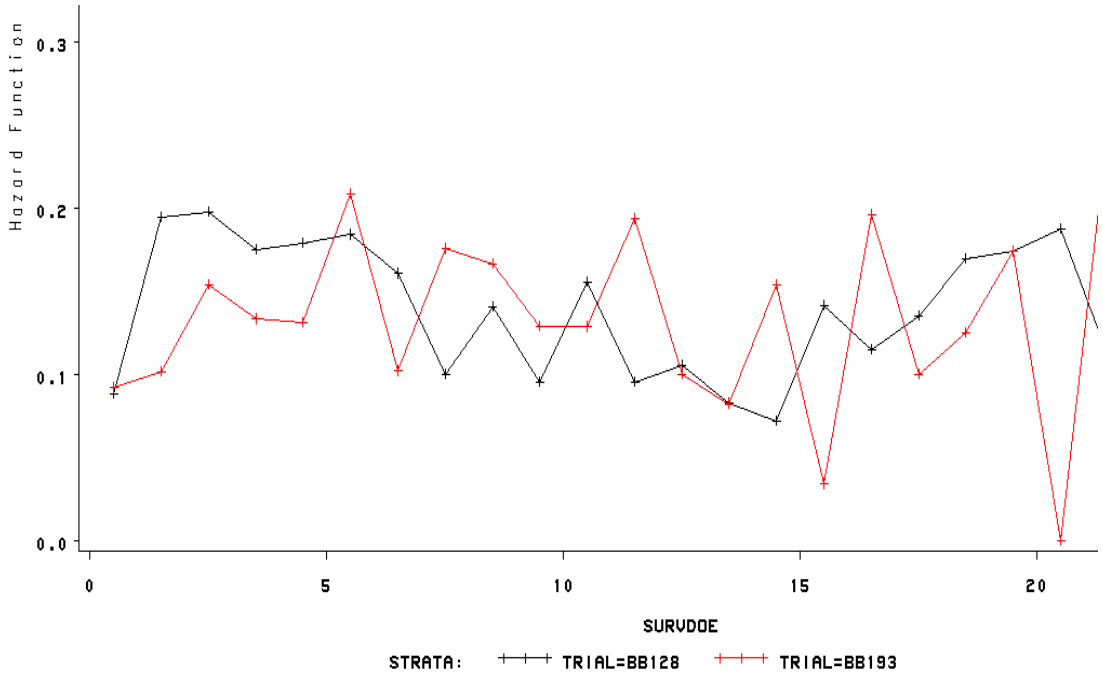
A confidence interval spanning unity would indicate a non-significant or borderline significant difference in hazard rates between the two groups. As with the log-rank test, other alternative versions of the hazard ratio and confidence intervals are available when the number of events is small (Machin *et al.*, 2006).

Many survival analysis methods are based on the assumption of proportional hazards, that the hazard ratio is constant over time. The validity of this can be examined using a log cumulative hazard plot: a plot of *log(-log[S(t)])* vs. *log t* where *S(t)* is the Kaplan-Meier estimate of the survivor function. With proportional hazards the curves for the groups should be approximately parallel. The PHREG procedure in SAS (SAS Institute Inc., 1999) calculates hazard ratios with 95% confidence intervals. The LIFETEST procedure in SAS plots log cumulative hazard curves.

3.8. Estimation of the Hazard Function in Advanced Pancreatic Cancer

Hazard rates were estimated for one-monthly time intervals to 18 months from trial entry and appeared similar for both trials and reasonably constant over time (Figure 3.3):

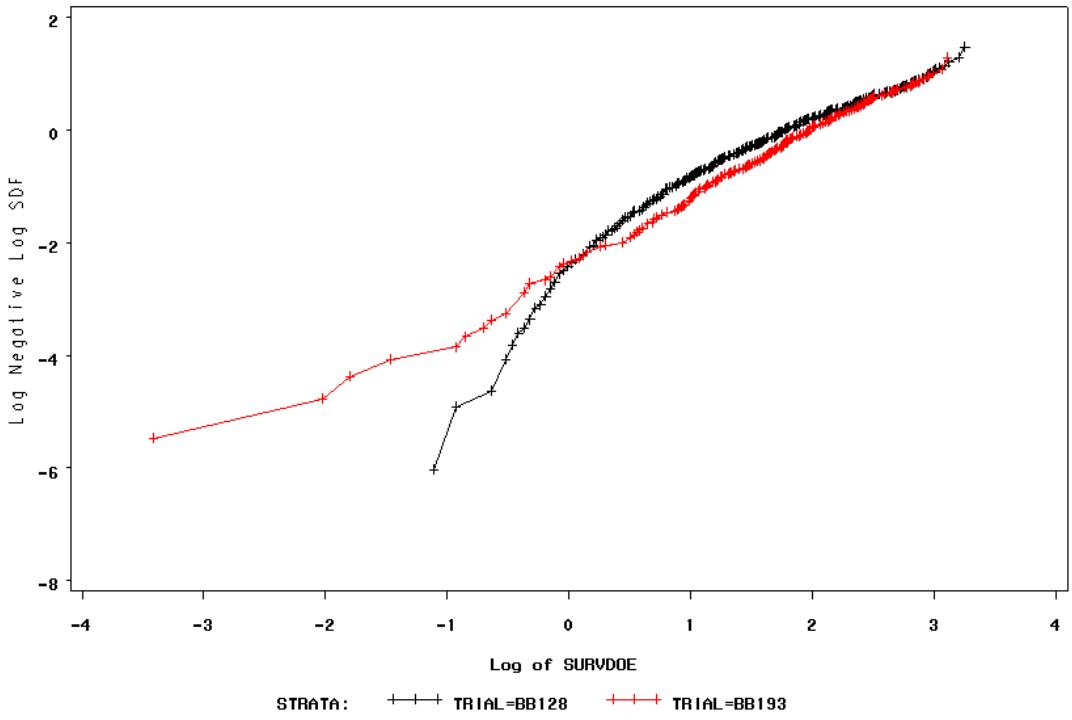
Figure 3.3: Monthly Hazard Rates by Trial



The hazard ratio (HR) of the relative risk of death across the two studies was estimated HR=0.88 for BB193 compared to BB128, indicating an overall 12% reduction in the risk of death in the 239 patients in the BB193 trial compared to the 414 patients in the BB128 trial. The 95% confidence interval was (95%CI: 0.75 to 1.04) indicating that 95% of calculated intervals will contain the true HR, estimated in this sample to lie somewhere between a 25% reduction and a 4% increase in the risk of death for BB193 patients compared to BB128 patients. The confidence interval provides a range of values for the true HR estimated from the available data but with increasing uncertainty at the limits of the interval. The fact that the confidence interval spans unity also confirms the lack of any statistically significant difference between the two trials as previously reported by the log-rank test statistic.

The assumption of proportional hazards was investigated using a log cumulative hazard plot and showed curves for the two trials to be approximately parallel (Figure 3.4):

Figure 3.4: Log Cumulative Hazard Plot by Trial



3.9. Conclusion

Prognostic modelling, described in Chapter 4, is based on multivariable analysis of time to event data, described in Chapter 5, so it is important to describe the basic statistical methods used to analyse time to event data which are more complex due to censored data. The Kaplan-Meier estimates of survival for this dataset (12-month survival rate estimated as 17%) are similar to those reported in Chapter 1 (12-month survival rate for England and Wales reported as 13%). Alive patients have been followed for almost two years and the 2-year survival estimate (not plotted) is estimated as 3.6% (5-year rates reported as 0.4 to 4%, Chapter 1). As such, this

dataset provides data from a representative sample of patients with advanced pancreatic cancer, based on estimated survival distributions. The survival distributions do not appear to be significantly different across the two trials. The hazard ratio does not indicate a significantly different hazard rate for either trial. The hazard function appears constant over time and similar across the two trials. As such, there does not appear to be any significant factor why the trial data from the two studies may not be merged for investigating potential prognostic factors. Statistical methods should always account for the fact that patients come from two trials and also that patients have been randomised to different treatment groups, even if these effects appear non-significant. The assumption of proportional hazards does not appear to be violated. Methods for assessing the prognostic ability of variables and developing prognostic indices are described in the next chapter. The statistical methods for identifying prognostic factors in the multivariable setting are described in Chapter 5.

CHAPTER 4: PROGNOSTIC INDICES

- **SUMMARY**
- Prognostic factor studies are intended to help clinicians in their decision making
- A prognostic index is a clinical tool to predict the survival of patients after diagnosis
- Evidence is also relevant for the appropriate design and analysis of cancer research
- 36 studies reported 34 possible prognostic factors for advanced pancreatic cancer, the majority questionable in terms of sample size and statistical methods
- Published studies are often contradictory and as such few prognostic markers are in use
- Prognostic studies should adopt a multivariable approach
- Model validation is important to assess the extent of over-fitting and thus external validity and where possible should be conducted using an independent dataset
- As such, model validation can be seen as a continuing process with the availability of new data
- Internal model validation methods (the non-parametric bootstrap) allow assessment of the validity and stability of selected prognostic factors in the absence of independent data
- Complete case analysis can result in biased or imprecise regression coefficients if missing data are not missing at random
- Analysis of a complete dataset based on multiple imputation could be informative

4.1. Introduction

The aim of this chapter is to discuss the important methodological issues surrounding identification of prognostic factors and the development of a prognostic index including model validation techniques and dealing with missing data.

The goal of prognosis is to predict the future. A prognostic index is a tool used by clinicians to predict the survival of patients after diagnosis. Prognosis may change over time but in terms of patient management is frozen at a specific time (often time of diagnosis) to enable treatment plans

to be put in place. Prognosis may be modified by therapy and assessment of possible prognostic factors is often the first step in determining a patient treatment plan. Comprehensive prognosis (Gospodarowicz *et al.*, 2001) reflects the expected outcome of all potential competing risks not just those associated with the illness in question. Medical decision making is carried out on a daily basis based on comprehensive prognosis which is most relevant to the individual patient. Prognostic factor studies are intended to help clinicians in their decision making. Evidence of prognostic factors is also relevant for the appropriate design and analysis of evaluative research. Identification of important prognostic factors can contribute to the eligibility criteria and stratification of patients to ensure balance of these important factors across each treatment group in randomised controlled trials. The more the variability in outcome due to prognostic factors can be minimised, the more confident research can be of the true efficacy of a treatment. Similarly determination of prognostic factors can identify which subgroups of patients could be explored for their response to specific treatments and highlight poorer prognostic groups suitable for more experimental or palliative treatments. The purpose of prognostic factor studies was summarised in an evaluation of prognostic factors in breast cancer (Altman, 2001;Altman and Lyman, 1998) and include (Table 4.1):

Table 4.1: Purpose of Prognostic Factor Studies

- To guide clinical decision making, including treatment selection and patient counselling
- To improve understanding of the disease process
- To improve the design and analysis of clinical trials (for example, risk stratification)
- To assist in comparing outcome between treatment groups in non-randomised studies allowing adjustment for case mix
- To define risk groups based on prognosis
- To predict disease outcome more accurately or parsimoniously

Prognostic factor studies tend to fall into two categories: exploratory or confirmatory (Gospodarowicz *et al.*, 2001). Most are exploratory studies, usually with no pre-specified hypotheses involving multiple data-driven analyses of multiple factors, endpoints and subgroups, but reporting only those which show statistical significance. Poorly designed studies often yield the most dramatic but probably incorrect and misleading results yet may achieve a wide audience in inadvertently published in reputable journals. As such these exploratory studies tend to have misleading conclusions and are often inconsistent and contradictory, limiting their potential application and as a consequence few prognostic markers are in use. Confirmatory studies, to the contrary, are prospective studies with pre-defined hypotheses, endpoints and subgroups resulting in limited analyses. At least one confirmatory study should be carried out before a new prognostic factor or prognostic index is considered for use in clinical practice. Important characteristics of a confirmatory prognostic factor study (Simon and Altman, 1994) include (Table 4.2):

Table 4.2: Characteristics of a Confirmatory Prognostic Factor Study

- Treatment should be standardised or randomised and accounted for in the analysis
- Hypotheses should be stated in advance, including specification of prognostic factors, coding of prognostic factors, endpoints and subsets of patients and treatments
- Sample size and number of events should be sufficiently large that statistically reliable results are obtained. There should be at least 10 events per prognostic factor examined per subset analysed
- Analyses should assess whether new factors add predictiveness after adjustment for or within subsets determined by known standard prognostic factors
- Consideration should be given to the number of analyses conducted
- Cut-off values for prognostic factors should be pre-specified or preferably avoided

Both the American Joint Cancer Committee (AJCC) and International Union Against Cancer (UICC) encourage only statistically significant and independent prognostic factors be considered

for use in clinical practice based on multivariable analyses of large prospective datasets with sufficient evidence to detect small but clinically relevant differences in outcome. However, both acknowledge few studies exist resulting in their alternative approach grouping tumour, host and environment related factors (Section 4.6, Table 4.4). Sample size calculations for prognostic factor studies are not straightforward and often not carried out since the data are often retrospective or derived for another primary purpose, such as a clinical trial. Data from randomised controlled trials can be used to study prognosis but if the treatment under investigation is effective then a 'treatment' variable should be included as a separate factor in the multivariable model (Machin *et al.*, 2006;Moons *et al.*, 2009;Simon and Altman, 1994). There are no straightforward methods for estimating sample size for multivariable prognostic studies. Power formulae in survival analysis are based on the relatively simple case of the comparison of randomised treatment groups using hypothesis testing. In prognostic studies it is an accurate estimate of the size of the effect of a set of covariates that is important. Possible prognostic covariates may be correlated with treatment so existing power formulae are not valid (Schmoor *et al.*, 2000). Prognostic factors often display a larger effect than treatment effects suggesting prognostic factor studies could be smaller than randomised controlled trials. However adjustment for correlation again increases the sample size and a variance inflation factor has been developed for the situation of one correlated factor (Schmoor *et al.*, 2000). In survival studies it is the number of events rather than the sample size that is important. The number of regression coefficients being estimated should be less than the number of events. Ideally, studies require at least several hundred outcome events (Moons *et al.*, 2009). As a guide, for each potential prognostic variable under consideration there should be at least 10 events to obtain reasonably stable estimates for each regression coefficient (Peduzzi *et al.*, 1995). In the advanced pancreatic cancer dataset there are 612 events which should be

adequate for stable estimation of the prognostic ability of the eighteen possible independent factors under investigation.

A pathway to obtaining high quality evidence regarding prognostic ability of a factor (Riley *et al.*, 2009) and guidelines for reporting tumour marker prognostic studies (McShane *et al.*, 2005) were published in an attempt to encourage large, protocol-driven prospective studies with transparent and complete reporting. Thus enabling sensible judgement of the usefulness of the reported data and explaining why so few clinically useful markers emerge from hundreds of tumour marker reports in oncology. Poor study design and analysis and inadequate reporting of studies were identified as major barriers.

4.2. Statistical Development of a Prognostic Index

In oncology, baseline data collected from identified studies at the time of randomisation into a randomised controlled trial may be utilised to develop a prognostic model for survival. Clinicians implicitly use multiple factors to estimate an individual patient's prognosis. As such, prognostic studies should adopt a multivariable approach to the design and analysis (Moons *et al.*, 2009). Multivariable regression methods are used for the simultaneous and flexible assessment of multiple baseline factors. Although individual variables may appear to be statistically significant prognostic factors, two or more factors may be aliased with one another and as such explain similar variability in the data. Multivariable regression methods aim to select those which appear to influence directly the outcome of interest and discard those which do not, over and above those already selected as important in the regression model. Multivariable regression methods can also take account and allow inclusion of already recognised prognostic factors.

Prognostic indices are derived from the final predicted regression model yielding a prognostic score for each patient. Prognostic scores can be grouped to form risk groups derived from the prognostic index based on cut-points of the index and survival curves can be calculated for each of these risk groups. Survival curves for these predicted risk groups could be biased (overfitting) since the same data are used for selecting the prognostic model as for estimating the effect of the model. Overfitting can occur in the development of prognostic models including large numbers of covariates when some of the structure of the data has been fitted to random noise in the data and is a measure of poor external validity. One method to assess of the extent of overfitting and achieve valid accurate estimates of prognostic value is to validate the model on an independent external set of data. However this is often not a possibility or may be inefficient if preserving an independent set requires withholding data from the model building process thus compromising estimation of the statistical functions. An alternative method is internal model validation based on statistical re-sampling techniques.

4.3. The Nonparametric Bootstrap

In the absence of an external dataset, internal model validation methods based on statistical re-sampling simulation techniques allow assessment of the accuracy of statistical estimates and of the stability of a set of selected prognostic factors. The nonparametric bootstrap (Efron and Tibshirani, 1993) is a way of using the available data to simulate new datasets had the study been repeated multiple times with a different sample of patients. The new simulated datasets are the bootstrap samples created by sampling one patient at a time with replacement from the original dataset up to the same size of the original set of data. Using random selection with replacement means that an individual patient from the original data set can be included repeatedly within a bootstrap sample while others may not be included at all, called bootstrap replication. This

process is repeated usually until a maximum number (often 200) of simulated bootstrap samples have been created. Since the bootstrap samples are derived from the original data, the amount of variability in a particular parameter estimate across all bootstrap samples should then provide an unbiased estimate of the stability of that estimate from the analysis of original data. Bootstrapping can also be used to assess the stability of a selected set of variables (Altman and Anderson, 1989).

A bootstrap depends on the original sample of data being representative. If the sample is not representative, conclusions will be inappropriate hence it is important that a sample is a good representation of the population of patients under study. Bootstrap validation methods should be seen as a complimentary tool, not a replacement for external validation methods. Internal validation methods are used to assess overfitting when modelling the advanced pancreatic cancer data. An external validation set was expected to become available during the project and as such it was anticipated that the final model could also undergo external validation. As such, data-splitting the original dataset was avoided thus maximising the available data to detect significant prognostic factors, using internal validation methods until the independent, external dataset became available.

4.4. Missing Data and Imputation

Incomplete prognostic data is common (Burton and Altman, 2004) but problematic. Statistical analytical methods, including maximum likelihood estimation, generally exclude patients with missing prognostic data. Analyses performed on a complete case basis with a subsequent reduction in available data for analysis, can impede construction of reliable models and can result in biased or under- or overestimation of regression coefficients and overestimation of variance if

missing data are not missing at random (MAR) affecting prognostic ability of the covariates (Burton and Altman, 2004). Missing completely at random (MCAR) data are data missing for reasons unrelated to outcome. Missing at random data are data missing for reasons related to other explanatory data but not outcome. Most statistical methods for dealing with missing data assume data are missing at random. Guidelines for reporting prognostic studies with missing covariate data (Burton and Altman, 2004) are based on assessment of the i) quantification of completeness of covariate data, ii) approaches for handling missing covariate data, recommending multiple imputation and iii) exploration of missing data, providing any known reasons. Methods for handling missing data include multiple imputation methods.

Imputing missing values and exploratory analysis of a complete imputed dataset could be informative compared with excluding subjects with incomplete data (Harrell, 2001) but computation needs to account for the imputation. Multiple imputation methods (Rubin, 1987;Schafer, 1997) are based on random selection from the conditional distribution of the variable in question in light of other variables. Imputation is repeated *m* times where $m \geq 3$ to account for variability related to the imputed values. Each *m* repeated imputation results in a complete dataset analysed using standard methods. Parameter estimates are averaged over the multiple imputations and variability across imputed datasets is a component when deriving estimates of the variance, thus including an estimate of uncertainty derived from the observed missingness and providing better estimates than a single imputation. Multiple imputation is viewed as the principal analysis by some (Harrell, 2001).

In the advanced pancreatic cancer dataset, the majority of variables had $\leq$5% missing values (Table 2.2) and reasons for missing data were not available. Tumour stage, CA19-9 and WBC

were missing in less than 10% of patients. The missing biochemical data is likely to be linked to 'centre'. Missing tumour stage is directly related to cancer stage (99% complete data) and again is likely to be linked to 'centre'. Lymph node status was missing for 24% of patients. In this sample of metastatic pancreatic cancer patients, missing lymph node status is likely to represent a group of patients where status was not deemed applicable to assess and hence was considered in the analysis as classification variable using 'negative' as a reference level and including 'missing' as a classification level. As such, the assumption of MAR does not appear to be violated. Principal analysis was based on complete cases (85% with complete data on the factors of interest) but multiple imputation was explored to provide valid, inferential alternative results.

4.5. Summary of Adopted Modelling Strategy

A proposed strategy (Harrell *et al.*, 1996) for maximising model accuracy and avoiding poorly fitted and over-fitted regression models in the development of multivariable prognostic models is summarised in Table 4.3.

The strategy is adapted for the analysis of the advanced pancreatic cancer dataset. Specifically, data on 653 patients, a large dataset in this disease, were collected at the time of randomisation and quality assured by British Biotech to minimise data errors and missing data. It is a clinically reduced dataset, reduced to those variables which are clinically meaningful and available within an NHS outpatient clinic. The data are robust for statistical modelling with a 94% event (death) rate and an average follow-up of alive patients of almost two years. Patients with limited response data (lost to follow-up) are retained in the analysis but censored out at the date last seen alive.

Table 4.3: Summary of Modelling Strategy

1. Assemble accurate data and as large a sample as possible with sufficient follow-up to capture events
2. Formulate focused clinical hypotheses, expected relationships and possible interactions
3. Discard observations with missing response data if missing at random
4. Assess missing covariate data, exclude if small % patients otherwise impute missing data
5. Data reduction techniques if number of covariates is large compared with outcome data
6. Use entire sample in model development
7. Check linearity assumption and make possible transformations
8. Check additivity assumptions and investigate pre-specified clinically motivated plausable interaction terms
9. Check for overly influential observations, may indicate over fitting, a need for rescaling or truncating highly skewed variables or highlight data errors
10. Check distributional assumptions (stratification or time-dependent covariates if proportional hazards assumption is violated, distributional assumption of parametric models)
11. Do limited backwards step-down variable selection – possible loss of information through stepwise techniques as do not address over fitting
12. This is the 'final' model
13. Validate model for calibration and discrimination ability, preferably using bootstrapping. Steps 7 to 11 repeated approximately for each bootstrap sample.
14. If using stepwise selection present a summary table of variability of important factors over bootstrap samples
15. Estimate likely shrinkage of predictions from model, consider shrinking predictions accordingly.

Multiple imputation was considered and presented as a supportive analysis to a complete case analysis, since the proportion of missing data is small. The assumption of 'linearity' refers to the functional form of the relationship between each continuous covariate and log-hazard, which is assumed to be linear. This assumption is discussed further in Chapter 5 and was investigated for each of the ten continuous variables in the advanced pancreatic cancer dataset comparing and presenting different methods of transformation, including dichotomisation. This research was focused on the main effects of eighteen possible prognostic factors investigating appropriate

transformations for the correct modelling of these main effects. As such, interactions of identified main effects especially where main effects are more complex non-linear transformations, is part of planned further research. Approximately four times as many patients are required for sufficient evidence to detect interaction terms (Schmoor *et al.*, 2000) and it is debatable if 612 events would be enough evidence. The proportional hazards assumption was investigated in Chapter 3 and was not considered to be substantially violated after accounting for unavoidable early deaths. Variable selection methods are discussed in Chapter 5, backward elimination being the adopted method for the analysis in the advanced pancreatic cancer dataset. Internal validation methods were used to avoid data-splitting the main dataset, maximising available data to detect significant prognostic factors until the independent, external dataset became available.

4.6. Reported Prognostic Factors in Pancreatic Cancer

Generally, ductal adenocarcinomas are separated from other pancreatic tumours, such as peri-ampullary tumours, which have better prognoses. Prognosis for tumours of the head of pancreas has been favoured compared with tumours of the body or tail which are generally diagnosed at a later stage of disease due to lack of symptoms. A review by the International Union Against Cancer (UICC) (Roder and Ott, 2001) reported prognostic factors for pancreatic cancer as 'essential', 'additional', 'new and promising' (Table 4.4). Essential prognostic factors are predominantly tumour related, TNM stage being the most important staging system based on size and location of the primary tumour: T1/T2 tumours are limited to the pancreas, T3 tumours invade local tissue (duodenum, bile duct, peripancreatic tissue) and T4 invade adjacent organs (stomach, spleen, colon, adjacent large vessels) (Roder and Ott, 2001).

Table 4.4: Prognostic Factors for Pancreatic Cancer

Prognostic Factors	Tumor Related	Host Related	Environment Related
Essential	TNM classification Lymph-node metastases Infiltration of peripancreatic tissue and organs Lymphatic vessel invasion *Tumor size* *Curative resection (R0)*		Possibility of resection High volume centers
Additional	Disseminated tumor cells in the bone marrow Disseminated tumor cells in the peritoneal cavity	Performance status	Adjuvant therapy
New and promising	DNA content Genetic instability Tumor suppressor genes *p53*, p16, DPC4, Bcl-2 Protooncogenes: Ki-67 Oncogenes K-ras Growth factors VEGF EGF/EGFR TGF-beta FDG-PET		Neoadjuvant concepts Antiangiogenesis Cytokine-secreting pancreatic adeno-carcinoma vaccine

In resectable patients the most important prognostic factor may be involvement of the resection margins (Neoptolemos *et al.*, 2001b). The first large randomised controlled trial (ESPAC-1) also reported grade of disease, lymph node status and tumour size as prognostic factors for survival following surgery (Neoptolemos *et al.*, 2004).

Due to the short survival of patients with advanced pancreatic cancer, it is possible there could be limited prognostic factors. A search of the WSci, EmBase and Ovid publication databases conducted in 2004 (Stocken *et al.*, 2008) identified 36 published prognostic factor studies

(excluding posters and presentations) reporting a total of 34 possible prognostic factors for advanced pancreatic cancer patients (Table 4.5), grouped as surgical, clinical, laboratory or demographic:

Table 4.5a: Published Prognostic Factor Studies

Type (number) of studies	Number of patients per study	Prognostic Factors reported (frequency of reporting)		Univariate (U) / Multivariable (MV) analysis
Randomised Controlled Trial (n=4)	207 to 322	Surgical: Clinical: Laboratory:	Metastases (3) Tumour location (1) Performance status (3) Treatment (1) Alkaline phosphatase (1)	4 MV
Consecutive Series >500 patients (n=3)	782 to 2380	Surgical: Clinical: Laboratory: Demographic:	Metastases (1) Stage of disease (1) Operation (1) Performance status (1) Diabetes (1) Pain (1) Appetite/weight (1) Jaundice (1) Treatment (1) Albumin (1) Age (1) Specialist centre (1)	2 MV 1 U

Table 4.5b: Published Prognostic Factor Studies

Type (number) of studies	Number of patients per study	Prognostic Factors reported (frequency of reporting)			Univariate (U) / Multivariable (MV) analysis
Consecutive Series 100 to 500 patients (n=14)	102 to 450	Surgical:	Metastases	(4)	13 MV 1 U
			Stage of disease	(2)	
			Tumour location	(1)	
			Operation	(2)	
			Tumour size	(1)	
			Duodenal invasion	(1)	
			Peri dissemination	(1)	
			Ascites	(1)	
		Clinical:	Performance status	(2)	
			Diabetes	(1)	
			Pain	(1)	
			Appetite/weight	(3)	
			Symptom onset	(1)	
			Treatment	(2)	
		Laboratory:	CA242	(2)	
			CA19-9	(2)	
			Leukocytes	(1)	
			Gamma GT	(1)	
			Albumin	(1)	
			LDH	(1)	
			CRP	(3)	
			Iron	(1)	
		Demographic:	Age	(1)	
Consecutive Series <100 patients (n=15)	28 to 95	Surgical:	Metastases	(1)	8 MV 7 U
			Stage of disease	(1)	
			Grade of disease	(1)	
			Nodal status	(1)	
			Operation	(1)	
			Tumour size	(2)	
			Fibrosis	(1)	
		Clinical:	Performance status	(4)	
			Inflammation	(1)	
			Appetite/weight	(1)	
			Treatment	(1)	
		Laboratory:	CA19-9	(7)	
			VEGF	(1)	
			CEA	(1)	
			Phase angle BIA	(1)	
			SCA	(1)	

The majority of these studies were questionable in terms of sample size and statistical methods, most based on small retrospective studies sometimes with inadequate analyses leaving the significance of specific prognostic factors unclear. Some variables are related (e.g. small tumours are more likely to be lymph node negative) hence the need to assess the importance of each factor in light of all other variables using multivariable statistical analysis. Thirty-two studies (Bachmann *et al.*, 2003;Cubiella *et al.*, 1999;Engelken *et al.*, 2003;Falconer *et al.*, 1995;Friedman and Vandeneeden, 1993;Fujino *et al.*, 2003;Halm *et al.*, 2000;Ikeda *et al.*, 2001;Ishii *et al.*, 1996;Karayiannakis *et al.*, 2003;Kuhlmann *et al.*, 2004;Lundin *et al.*, 1995;Micke *et al.*, 2003;Ni *et al.*, 2005;Ohigashi *et al.*, 2003;Paillaud *et al.*, 2003;Ridwelski *et al.*, 2001;Rothenberg *et al.*, 1996;Saad *et al.*, 2002;Shibamoto *et al.*, 1996;Stemmler *et al.*, 2003;Storniolo *et al.*, 1999;Talar-Wojnarowska *et al.*, 2003;Tas *et al.*, 2001;Terwee *et al.*, 2000;Trigui *et al.*, 2000;Tsuruta *et al.*, 2001;Ueno *et al.*, 2000;Watanabe *et al.*, 2004;Yasue *et al.*, 1994;Ziske *et al.*, 2003;Gupta *et al.*, 2004) were based on consecutive series of patients, often retrospective, often single-centre, of which 15 studies were based on fewer than 100 patients (Halm *et al.*, 2000;Ikeda *et al.*, 2001;Ishii *et al.*, 1996;Karayiannakis *et al.*, 2003;Micke *et al.*, 2003;Ohigashi *et al.*, 2003;Paillaud *et al.*, 2003;Rothenberg *et al.*, 1996;Saad *et al.*, 2002;Stemmler *et al.*, 2003;Talar-Wojnarowska *et al.*, 2003;Tsuruta *et al.*, 2001;Yasue *et al.*, 1994;Ziske *et al.*, 2003;Gupta *et al.*, 2004). One study was a summary of five observational studies with varied inclusion criteria, inconsistent results and no prospective verification (Terwee *et al.*, 2000). The largest series (2380 patients) identified factors based on univariate analyses and data containing a large proportion (57%) of censored patients (Storniolo *et al.*, 1999). Four studies (Berlin *et al.*, 2002;Ducreux *et al.*, 2002;Johnson *et al.*, 2001;Maisey *et al.*, 2002) were randomised controlled trials and reported five prognostic factors based on multivariable analyses: metastases, tumour site, performance status, alkaline phosphatase and treatment.

Six of the papers reported in the literature review (Table 4.6) claimed to have developed a prognostic index suitable for use in clinical practice when making therapeutic decisions and suitable for use when designing and analysing future clinical trials. The largest (Terwee *et al.*, 2000) was a study pooling 1020 patients with unresectable pancreatic cancer from five studies. The strength of the study is the large sample size but the authors admit there were inconsistencies in the results in the individual studies included with varied survival figures unexplained by the prognostic variables. These differences could be due to the exclusion of important risk factors since the study was limited to only three prognostic factors available across all five studies. Three of the remaining studies were retrospective and five were based consecutive series of patients, the largest including a total of 166 patients, the smallest including only 55 patients (47 deaths). None of these reported indices has been adopted into clinical practice in the UK and as such, there is currently no prognostic tool in routine use for prediction of survival in advanced pancreatic cancer by which clinicians could identify subgroups of patients who may or may not benefit from further treatment.

Table 4.6: Summary of Reported Prognostic Indices

Study	Single/ Multicentre	Patients (Deaths)	Variables (Continuous)	Analysis of Continuous Covariates	Prognostic Factors Identified	Validated	Comments
1 (Terwee *et al.*, 2000)	Pooling of 5 Multicentre studies	1020 (?)	3 (1) for 5 studies 6 (1) for 4 studies	Non-linearity investigated	Age Sex Metastases	No	Prospective Limited factors Included study and treatment
2 (Trigui *et al.*, 2000)	25 centres	166 (?)	17 (5)	Dichotomised	Pain Ascites Weight loss Metastases	No	Prospective Excluded centre and Treatment
3 (Cubiella *et al.*, 1999)	Single	134 (134)	34 (18)	Dichotomised at median	Performance status Metastases	No	Retrospective Excluded treatment
4 (Ueno *et al.*, 2000)	Single	103 (95)	20 (10)	Dichotomised	CRP Performance status CA19-9	No	Retrospective Excluded treatment
5 (Ishii *et al.*, 1996)	Single	65 (?)	? (?)	Dichotomised	Performance status CEA Metastases	No	Retrospective Excluded treatment
6 (Ikeda *et al.*, 2001)	Single	55 (47)	19 (8)	Dichotomised	Performance status Lymph node swelling CA19-9	No	Prospective Excluded treatment

4.7. Clinical Trial Design

The usual endpoint for efficacy of cancer treatments is all cause mortality and comparison of overall survival is often the primary outcome measure in randomised controlled trials in cancer. Other endpoints include progression free survival (time to progression or death whichever is first, also known as disease free survival), objective response rates, toxicity of experimental treatments and patient self-reported quality of life. Symptom assessment from the patient perspective (health related quality of life) is increasingly important in clinical trials where small survival differences in treatments are expected (Carter *et al.*, 2009) and indeed clinical benefit response (a composite of measurements of pain, performance status and weight) was used as the primary measure of treatment efficacy in the licensing trial of gemcitabine in the advanced disease setting (Burris *et al.*, 1997).

To compare survival rates adequately across treatment groups and provide sufficient evidence of a survival difference requires large studies. Randomised controlled trials in pancreatic cancer usually require larger number of patients to enable detection of small expected treatment differences and as such may require multi-centre/ country participation. Random allocation of patients to treatment groups should be used to minimise selection bias and ensure patient and tumour characteristics are adequately balanced across the different treatment arms. Identification of important prognostic factors is desirable to enable accurate stratification of patients at the time of randomisation. Stratification factors are factors which are known to have a significant effect on outcome and as such should be balanced equally across treatment groups. Stratification factors used in pancreatic clinical trials in the adjuvant setting often include surgical centre and resection margin status (R0 patients are those with negative resection margins at surgery who have a better prognosis than patients with residual tumour left at surgery (R1) (Butturini *et al.*,

2008;Neoptolemos *et al.*, 2001b)). Stratification factors used in pancreatic clinical trials in the advanced disease setting include oncology centre, stage of disease (locally advanced patients are those with no distant metastases who have a better prognosis than patients with metastases) and performance status (patients with performance status 0 or 1 have better performance status associated with better prognosis than patients with performance status 2) (Cunningham *et al.*, 2009).

4.8. Conclusions

In summary, the advanced pancreatic cancer dataset contains 653 patients with 612 deaths and should enable stable estimation of the prognostic ability of the eighteen identified possible prognostic factors. Six previously reported prognostic indices included one large study of 1020 patients but with only three common variables to investigate. The remaining five studies were based on $\leq$166 patients, four were single centre studies, three were retrospective. None of the six studies included any validation. A total of 36 prognostic factor studies were identified reporting a total of 34 prognostic factors and of these 11 were available for analysis in the advanced pancreatic cancer dataset. Of the 23 factors not collected, 15 had only been reported once. The most frequently occurring factors, not available in the advanced pancreatic dataset were performance status (reported in 10 studies) and appetite/ weight loss (reported in 5 studies). The functional form of the relationship between each continuous covariate and log-hazard was usually not investigated, usually being dichotomised. Non-linearity is addressed for the first time in this disease site and was investigated for each of the ten continuous variables in the advanced pancreatic cancer dataset. Model validation is addressed for the first time in this disease site; internal validation was carried out to assess the degree of overfitting, external validation was carried out based on an independent, external dataset. Multiple imputation methods were used as

a supportive analysis to the principal complete case analysis, to assess the impact of missing data which again had not been addressed previously in the literature.

CHAPTER 5: STATISTICAL METHODS FOR PROGNOSTIC MODELLING

- **SUMMARY**
- A multivariable approach to analysis allows investigation of the relationship between survival and one or more possible factors
- Non-parametric methods model the hazard function with no distributional assumptions
- Akaike's Information Criteria is valid for comparing non-nested as well as nested models
- Multivariable regression modelling can be used with categorical and continuous variables
- Dichotomisation of continuous data is common but is inefficient and unnecessary
- Continuous variables are often simplified by assuming a linear relationship between predictor and log-hazard which may not be appropriate
- Assessment of model performance and validity is an integral aspect of model development

5.1. Introduction

The aim of this chapter is to describe the statistical methods for assessing treatment effects and identification of prognostic factors in the multivariable setting including constructing prognostic indices. Kaplan Meier and log-rank methods can be used in the analysis of a single possible prognostic factor but in reality many underlying factors may have an impact on length of survival and all factors need to be investigated and accounted for in the analysis. A multivariable modelling approach to the analysis of survival data allows investigation of the relationship between survival and one or more possible factors. A variety of statistical methods are available for multivariable modelling of survival data but generally most clinical studies, including randomised cancer clinical trials, use the proportional hazards regression model (Cox, 1972) to investigate possible prognostic factors and treatment effects. Cox proportional hazards regression and methods for assessing the underlying assumptions are discussed, specifically the assumption of proportional

hazards. The underlying linearity assumption for continuous covariates is introduced and discussed. Issues surrounding variable selection, comparison of different regression models and issues relating to missing covariate data are discussed.

5.2. Cox Proportional Hazards Regression

In parametric models there is some underlying assumption about the distribution of the hazard rate. For example, in a Weibull regression model the baseline hazard function is non-constant but rather increases or decreases monotonically over time, although does not change direction:

$$h_0(t) = \lambda\gamma t^{\gamma-1}$$

where λ is a scale parameter and γ is a shape parameter, both >0, denoted W(λ, γ). Weibull is often used when the survival distribution is assumed to be positively skewed i.e. with a heavy left tail, early events. Indeed when γ=1 then the hazard function is constant λ and conforms back to the exponential distribution. Estimates of log λ (intercept) and γ (slope) can be taken from a log cumulative hazard plot where non-parallel lines indicate different shape parameter γ across groups indicating non-proportionality. In pancreatic cancer clinical trials, especially those in the adjuvant setting where all patients undergo 'curative' resection, patients may be more at risk of death post-surgery with diminishing risk as time increases. In this case the hazard rate may be dependent on follow-up time. The purpose of statistical modelling is to relate non-constant hazards to influential covariates. In contrast to the parametric models, non-parametric methods model the hazard function with no distributional assumption for survival times. Since parameters are estimated for each of the potential covariates, the Cox proportional hazards regression is described as a semi-parametric approach.

Cox proportional hazards regression models the hazard function directly and aims to detect the extent to which potential factors affect the form of the hazard function. The only assumption is that the hazards across groups are proportional to one another and thus the ratio of hazard functions does not vary with time. For two treatment groups the proportional hazards model (Collett, 1994) is:

$h_A(t) = \psi h_B(t)$

where $h_A(t)$ and $h_B(t)$ are the hazards of death at time t for treatments A and B respectively and ψ is a constant value of the ratio of the hazards (the hazard ratio) where $\psi = h_B(t) / h_A(t)$. An alternative more general expression is:

$$h_i(t) = e^{\beta X_i} h_0(t)$$

where $h_i(t)$ is the hazard function for the ith of n patients, $h_0(t)$ is the baseline hazard function where all covariate values X_i equal zero and $e^{\beta X} = \psi$ (the hazard ratio) and can be rewritten as:

$$h_i(t) = \exp(\sum_{j=1}^{p} \beta_j X_{ji}) h_0(t) \qquad \text{(Equation 1)}$$

where j=1 to p explanatory variables X_i and β_j are regression coefficients for expressing the relationship between the hazard of death and each covariate X_j. Then $\log \psi$ (the log hazard ratio) = βX. For a vector of explanatory variables X the hazard ratio is interpreted as the hazard at time t for a patient relative to the hazard for a patient with all variable values of zero. The general proportional hazards model is then:

$$h_i(t) = \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi}) h_0(t)$$

and can be rewritten as an additive model on the log-hazard scale:

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

where $\beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} = \eta_i$ is known as the linear predictor of the model, also known as risk score or prognostic index. The proportional hazards model can also be regarded as the linear model for the logarithm of the hazard ratio.

Inferences about the effects of the explanatory variables X_i on the log hazard ratio (estimating β_p coefficients) in the proportional hazards model can be made without making any assumptions regarding the form of the baseline hazard function $h_0(t)$ although estimates of the β_p coefficients can be used to estimate $h_0(t)$ if needed. The unknown estimates of the β_p coefficients are computed using maximum likelihood estimation. The likelihood of the data is the joint probability of the observed data as a function of observed survival times and unknown β parameters. The estimates of β are values which are most likely given the observed data or maximum likelihood estimates which are values which maximise the likelihood function. For computation purposes, it is simpler to maximise the logarithm of the likelihood function using an iterative procedure such as the Newton-Raphson with Breslow's approximation for multiple tied events at a specific time. Computation of the standard errors for each β estimate (se[β]) enable a 95% confidence interval confidence interval for β to also be estimated based on: $\beta \pm 1.96 * se[\beta]$. Since the distribution of the logarithm of the hazard ratio is more closely approximated with the normal distribution and $\log \psi = \beta X$ then a 95% confidence interval confidence interval for the true hazard ratio ψ can calculated by exponentiating the confidence interval calculated for β.

If the confidence interval for β does not include zero then this provides evidence that the value of β is non-zero. The null hypothesis, that $\beta_p=0$ in the presence of other terms in the model, can be tested using the Wald test which compares $\beta/se(\beta)$ against percentage points of a chi-square

distribution at a particular significance (error) level, often 5%. Generally individual estimates of β_p in are not completely independent of one another so testing separate hypotheses about β parameters may not be interpretable. Alternative methods for comparing Cox proportional hazards models are based on the log-likelihood ratio statistic.

5.3. Comparing Alternative Models

How well a model fits to a specific set of survival data can be calculated using a statistic known as -2logL (-2*log-likelihood) which is a summary measure of the agreement between the model and the data. Since –2logL is dependent upon the number of observations in the data then its value is dataset specific and is increased with increased observations. As such it cannot be used to compare models across different datasets. Comparison of -2logL can be used to compare two nested proportional hazards models developed on the same set of data however formal testing is inadvisable since -2logL does not have a chi-squared distribution (Collett, 1994). Since only nested models should be compared using -2logL, Akaike's Information Criteria (AIC) (Akaike, 1974) is valid for comparing non-nested as well as nested models by penalising the log-likelihood for model complexity by adding at least twice the number of parameters estimated:

$AIC=-2logL + \alpha k$

where α is a predetermined constant between 2 and 6 (determined according to chosen significance level where q=2 or 3 approximates a 5% significance level) and k is the number of unknown β parameters in the model. Choosing a model based on the AIC is a successful strategy producing simplified models (Ambler *et al.*, 2002).

Smaller values of the AIC indicate a better model but it is unclear whether AIC statistics from non-nested models can be compared with hypothesis testing. Akaike suggested (Akaike, 1974) that in

some circumstances testing could be used: "When there are different families approximating the true likelihood equally well the situation will at least locally be approximated by the different parameterizations of one and the same family. For these cases the significance of the difference of AIC's between two models will be evaluated by comparing it with the variability of a chi-square variable with the degree of freedom equal to the difference of the k's of the two models". Using a fixed level of significance for the comparison of AIC is not appropriate i) in nested models with different k as it does not account for increased variability of estimates due to increased number of parameters and ii) in non-nested models with the same k. The AIC can be seen as a method of ranking models, where the model with the smallest AIC is ranked 'best', the next smallest being 'next best' and so on. In this way, a 'best' model is declared but which may be more complex than and not as parsimonious as the 'next best' model but with almost identical AIC statistics. Hypothesis tests are based on specific distributional assumptions. A way to avoid making distributional assumptions is to calculate AIC statistics and the difference in AIC's between the different models in multiple bootstrap resamples of the data. This provides a distribution for the differences in AIC between different models which requires no distributional assumptions and which can then be summarised appropriately.

There are different variable selection techniques based on forward selection, backward elimination and stepwise selection procedures: forward selection begins with a null model and includes significant variables one at a time into the model, backward elimination begins with a full model of covariates and excludes non-significant variables one at a time, stepwise selection is based on forward or backwards selection but also considers all other variables for exclusion/ inclusion at each step, although previously deleted variables are normally permitted to re-enter the model only once. These procedures are often automated in statistical software packages but should be used

with caution since they may provide results which are dependent upon the selection technique used and may not account for any known hierarchy in the data. The 'full' model including all variables can be difficult to interpret with large numbers of variables if many are not important. Forward and backward elimination selection methods are advantageous in selecting only variables with greater influence with high probability and eliminating variables with little influence resulting in a more general and useful model not substantially inferior to the 'full' model (Sauerbrei *et al.*, 2007). Selection procedures can be unstable in small datasets but instability can be addressed in an analysis of models from bootstrap replication as an internal validation in the absence of independent data for external validation. A recommended selection strategy is summarised (Table 5.1, (Collett, 1994),adapted):

Table 5.1: Strategy for Model Selection

1. Fit models that contain each variable one at a time. Compare with the null model to determine which variables on their own significantly reduce the value of -2logL.
2. Important variables from Step 1 are fitted together. Variables which do not significantly increase -2logL when omitted are discarded. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined
3. Variables which were not important on their own are added one at a time and any that significantly reduce -2logL are retained in the model.
4. A final check is made to ensure that no term in the model can be omitted without significantly increasing -2logL and that no term not included significantly reduces -2logL

Forced inclusion refers to the inclusion of factors that were deemed important at the study design stage 'forcing' into each model to account for the structure of the data forced through the study design. Possible candidate factors to consider would include stratification factors at randomisation including centre effect and in the case of prognostic modelling accounting for the underlying structure according the randomisation of treatment groups. It is recommended (Machin *et al.*,

2006;Moons *et al.*, 2009;Simon and Altman, 1994) that for randomly allocated treatments, treatment group should be forced into the model and the corresponding hazard ratio reported even if it turns out to be a non-significant variable, and equally so for variables used in any stratified randomisation procedure. Stratification factors at randomisation in the advanced pancreatic dataset were trial, cancer stage, and sex.

5.4. Format of Variables

Variables for consideration of inclusion in a multivariable model may be a mix of categorical (including binary) and continuous data. A strength of multivariable regression modelling is that it can be used with both categorical and continuous variables. It has been shown that the hazard ratio $\psi = h_B(t) / h_A(t)$ for the binary case of two treatment groups and that $\log \psi = \beta X$. For categorical variables with greater than two groups, individual categories are grouped into pairs as in the binary case known as dummy, indicator or 'classification' variables. In the advanced pancreas data set, tumour stage is coded as either 0, 1, 2, 3, 4 for each patient. Providing there are sufficient observations in each individual category then classification variables could be created as (Table 5.2):

Table 5.2: Possible Classification Variables for Tumour Stage

	Class T1	Class T2	Class T3	Class T4
Tumour stage 0	0	0	0	0
Tumour stage 1	1	0	0	0
Tumour stage 2	0	1	0	0
Tumour stage 3	0	0	1	0
Tumour stage 4	0	0	0	1

In this way, each classification variable reduces to a binary covariate and is modelled as previously described ensuring all the classification variables created from one specific categorical variable are included in the model simultaneously. Each categorical variable with k categories will have k-1 classification variables with estimation of k-1 corresponding parameter estimates.

Continuous covariates are often categorised (frequently dichotomised) for clinical ease of interpretation. However this implies a step in hazard of $\exp(\beta)$ at a cut point which is may not be realistic, more likely is a smooth change in hazard. Dichotomisation of continuous data is common practice but is problematic and unnecessary. Since the variability in outcome within groups is ignored by categorisation then the variability between groups may be significantly underestimated since patients close to the cut point are analysed as being very different rather than being very similar, resulting in a serious reduction of statistical power to detect relationships between predictors and outcome, residual confounding and serious bias (Altman and Royston, 2006;Royston *et al.*, 2006a). The risk of bias occurs when the choice of cut point is data driven based on investigating multiple cut points but reporting only that which is most significant. The use of different cut points across multiple studies also hinders direct comparisons. Continuous variables are often simplified by assuming a linear relationship between predictor and log-hazard i.e. the log risk increases or decreases linearly as the value of the factor increases, which also may not be appropriate. Indeed the effects of important prognostic factors may go unrecognised due to these simplistic assumptions made in statistical modelling (Stocken *et al.*, 2008). It is more preferable to retain the continuous nature of the variable whilst allowing some form of non-linearity accomplished by the application of a transformation to the independent variable (Box and Tidwell, 1962). Misspecification of the functional form may lead to inappropriate conclusions but has not been previously investigated in pancreatic cancer studies.

The PHREG (experimental version TPHREG in SAS v9.1 now combined into PHREG in SAS v9.2) procedures in SAS (SAS Institute Inc., 1999) carry out Cox proportional hazards regression creating k-1 classification variables for categorical covariates with associated type-3 χ^2 test based on k-1 degrees of freedom and also creating additional levels of a classification variables to account for missing data, if required.

5.5. Model Performance

After a model has been fitted to a set of observations, adequacy of that model can be investigated using residuals. Residuals are useful as they are calculated for each individual patient in the dataset and can be plotted to study for apparent patterns indicating lack of fit. Deviance residuals are residuals which should be symmetrically distributed about zero and as such can help identify outlying patients. An alternative approach is using Martingale residuals which can help assess functional form. Deviance residuals are Martingale residuals transformed to be more symmetric and can be output from the PHREG procedure in SAS (SAS Institute Inc., 1999) using the RESDEV statement. Prognostic risk groups can be derived from the individual patient prognostic scores and the survival distributions described for each of these risk groups. Internal model validation investigating the stability and external validity of a set of selected prognostic factors can be carried out by estimating the extent of model optimism (overfitting) in a model using statistical re-sampling simulation techniques. This can be carried out using the 'validate' statement in the Design Library of the programming language R (The R Foundation for Statistical Computing, 2008).

5.6. Statistical Development of a Prognostic Model in Advanced Pancreatic Cancer

A strategy based on avoiding over-fitted regression models in the development (Harrell 1996, Collett) and reporting (McShane 2005) of multivariable prognostic models was followed. The first stage of data reduction in the advanced pancreatic cancer dataset was considering only factors that were clinically relevant and available within an NHS out-patient clinic. Eighteen baseline clinical, histological, biochemical and demographic variables (including trial and randomised treatment group) were considered appropriate for analysis as possible prognostic factors of survival (Table 2.2). The majority of variables had $\leq$5% missing values. Tumour stage, CA19-9 and WBC were missing in less than 10% of patients and lymph node status was missing for 24% of patients. Metastases or lymph node status were considered in the analysis as classification variables using 'negative' as a reference level. Tumour stage and cancer stage were considered as binary variables based on clinical confirmation and small numbers of patients within categories. Both variables were also investigated retaining the original ordered categories but this did not alter the results. The principal analysis was based on complete cases where patients had complete data on all of the prognostic factors of interest. Multiple imputation is considered in the following chapter following development of the final prognostic model.

Initial analysis was based on standard methodology comparing Kaplan-Meier survival estimates using the log-rank test and estimating univariate hazard ratios for levels of each factor (Table 5.3, (Stocken *et al.*, 2008), adapted):

Table 5.3: Univariate Log-Rank Analyses

		Patient	Death	12-month Survival	Median Survival (95% CI)	χ^2_{LR}, p (χ^2_W, p)	HR (95% CI)
Overall Survival		653	612	17%	4.7 (4.2 to 5.1)	-	-
Trial	**BB128**	414	392	17%	4.2 (3.6 to 4.8)	2.28, p=0.13	1.0
	BB193	239	220	18%	5.4 (4.8 to 6.0)		0.88 (0.75 to 1.04)
DEMOGRAPHICS							
Age group	**<=63 yrs**	343	320	21%	5.1 (4.3 to 5.8)	4.42, p=0.036	1.0
	>63	310	292	13%	4.3 (3.5 to 4.9)	(7.55, p=0.006)	1.18 (1.01 to 1.39)
Ethnic group	**White**	590	554	18%	4.6 (4.1 to 5.1)	0.20, p=0.65	1.0
	Other	62	57	12%	5.2 (3.5 to 6.0)		1.06 (0.80 to 1.41)
Sex	**Female**	285	267	18%	4.9 (4.2 to 5.8)	0.73, p=0.39	1.0
	Male	368	345	17%	4.5 (3.9 to 5.1)		1.07 (0.91 to 1.26)
Treatment	**Gemcitbine**	222	204	18%	5.5 (4.7 to 5.9)	2.79, p=0.095	1.0
	Marimistat	431	408	17%	4.2 (3.5 to 4.9)		1.15 (0.98 to 1.36)
TUMOUR INFORMATION							
Cancer Stage	**Early(I/II)**	104	92	26%	6.8 (5.7 to 8.2)	14.72, p<0.001	1.0
	Late (III/IV)	543	514	16%	4.1 (3.5 to 4.7)		1.53 (1.26 to 1.86)
Metastases	**M0**	194	176	30%	6.8 (5.9 to 8.4)	35.47, p<0.001	1.0
	M1	436	414	12%	3.5 (3.2 to 4.0)		1.69 (1.43 to 1.99)
	Missing	23	22	17%	5.5 (4.9 to 7.5)		1.35 (0.88 to 2.09)
Lymph nodes	**N0**	243	226	20%	5.5 (4.8 to 6.0)	6.73, p=0.035	1.0
	N1	251	240	18%	4.5 (3.5 to 5.4)		1.19 (1.00 to 1.43)
	Missing	159	146	11%	3.8 (3.2 to 4.9)		1.29 (1.04 to 1.59)
Tumour Stage	**Early(0/1/2)**	310	291	16%	4.3 (3.7 to 4.9)	1.63, p=0.44	1.0
	Late (3/4)	287	268	18%	4.9 (4.2 to 5.8)		0.91 (0.77 to 1.08)
	Missing	56	53	18%	5.8 (3.5 to 7.9)		0.87 (0.66 to 1.16)
SERUM CHEMISTRY and HAEMATOLOGY							
AST	**Normal**	538	499	19%	5.1 (4.6 to 5.7)	14.17, p<0.001	1.0
	Abnormal	86	84	12%	2.8 (2.2 to 3.9)	(5.99, p=0.014)	1.55 (1.18 to 2.04)
BILIRUBIN	**Normal**	464	429	20%	5.1 (4.7 to 5.7)	9.32, p=0.002	1.0
	Abnormal	165	159	11%	3.8 (3.3 to 4.4)	(6.27, p=0.012)	1.32 (1.09 to 1.61)
ALK PHOS	**Normal**	442	411	20%	5.5 (5.0 to 6.1)	20.20, p<0.001	1.0
	Abnormal	187	177	13%	3.1 (2.6 to 3.5)	(56.05, p<0.001)	1.49 (1.23 to 1.81)
ALBUMIN	**Normal**	583	544	19%	5.1 (4.6 to 5.6)	31.37, p<0.001	1.0
	Abnormal	45	43	7%	1.5 (1.0 to 2.7)	(74.34, p<0.001)	2.36 (1.49 to 3.72)
LDH	**Normal**	543	505	20%	5.2 (4.8 to 5.8)	37.05, p<0.001	1.0
	Abnormal	78	75	5%	2.1 (1.5 to 2.8)	(36.16, p<0.001)	2.08 (1.50 to 2.88)
BUN	**Normal**	407	382	20%	5.1 (4.3 to 5.7)	3.43, p=0.064	1.0
	Abnormal	213	199	13%	4.4 (3.5 to 5.1)	(5.28, p=0.022)	1.17 (0.98 to 1.40)
CA19/9	**Normal**	98	86	28%	6.3 (4.8 to 8.0)	7.74, p=0.005	1.0
	Abnormal	508	481	16%	4.6 (4.0 to 5.1)	(4.84, p=0.028)	1.38 (1.12 to 1.70)
Hb	**Normal**	79	77	8%	3.7 (3.3 to 5.1)	6.88, p=0.009	1.0
	Abnormal	533	495	20%	4.9 (4.4 to 5.6)	(10.64, p=0.001)	0.73 (0.55 to 0.95)
WBC	**Normal**	483	446	21%	5.5 (4.9 to 5.9)	34.36, p<0.001	1.0
	Abnormal	129	126	8%	2.9 (2.4 to 4.0)	(46.52, p<0.001)	1.78 (1.40 to 2.26)

LR=Log-Rank Statistic, W=Wald Chi-square Statistic under 'linear' assumption, HR=Hazard Ratio for categories

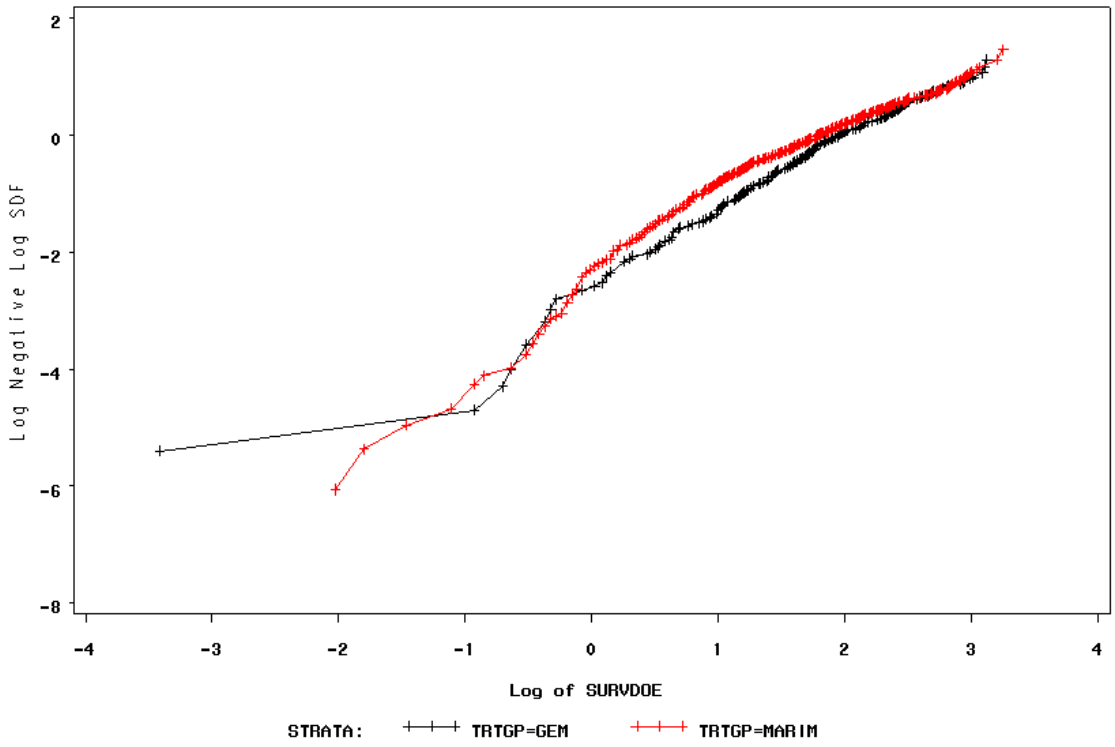
Ten of the eighteen possible prognostic factors were collected as continuous measurements. Continuous data were initially investigated based on i) a step functional relationship with outcome using dichotomised covariates (normal/ abnormal based on central laboratory reference ranges) or ii) under the assumption of linearity with log-hazard. Step-functional relationships are affected by the cut-point used which can be pre-defined or data-driven. In these trial data, cut-points were pre-defined as normal/ abnormal (according to central laboratory reference ranges) as used in clinical practice, grouping abnormally low and abnormally high measurements together.

Log-rank analyses indicated that potentially important factors were age, cancer stage (I/II vs. III/IV), metastases and laboratory measures AST, alkaline phosphatase, albumin, LDH, WBC, bilirubin, CA19-9 and haemoglobin. BUN and nodal status held borderline statistical significance at the 5% level. Trial, treatment (gemcitabine vs. marimistat), race (white vs. not-white), sex and tumour stage (T0,1,2 vs. T3,4) did not appear to be significantly related to survival.

5.7. Assumption of Proportional Hazards

For valid interpretation of the regression coefficients from a Cox proportional hazards model, the assumption of proportional hazards must hold. The proportional hazards assumption was investigated for each covariate using log cumulative hazard plot (Collett, 1994) which should give approximately parallel lines when proportional hazards are observed. The log cumulative hazard plot by trial (Figure 3.4) showed curves for the two trials to be approximately parallel. The same can be seen in the log cumulative hazard plot by treatment group (Figure 5.1):

Figure 5.1: Log Cumulative Hazard Plot by Treatment Group



The shape of the log cumulative hazard curves are subject to greater measurement error in the early stages of follow-up and as such justifies the use of an appropriate test, rather than reviewing and interpreting plots. An appropriate test of proportionality is the inclusion and assessment of a time dependent covariate which is a method used to capture non-proportionality should it exist. A time dependent covariate is a variable X whose values change over time t. Since the values of X depend on t then the relative hazard with the baseline hazard is also dependent on time and the model no longer has proportional hazards. When proportional hazards are observed, the influence of a time dependent covariate X (X=factor*(ln(survival)-ln(mean survival))) should be minimal and non-significant in the proportional hazards model. The mean observed time in the 556 advanced pancreatic cancer patients with used in the complete case analysis was 6.8 months. Time

dependent covariates were calculated for factors of trial and treatment and were non-significant (p=0.24 and p=0.08 respectively).

5.8. Cox Proportional Hazards Regression Models

The hazard of death was assessed in the multivariable setting using Cox proportional hazards regression modelling based on the strategy presented in Table 5.1 (Collett, 1994). Variable reduction was based on a manual backward elimination method using a nominal significance level of 0.05 for elimination and including trial, sex, cancer stage (stratification factors at randomisation) and randomised treatment as 'forced' variables in each model. A final check was made to ensure that no term in the model could be omitted and that none of the omitted terms could be included without significantly affecting the model. All variables omitted from the model based on backward elimination were considered and re-instated if they significantly improved the model. Continuous variables were investigated based on i) dichotomisation ('categorical' model) or ii) under the assumption of linearity ('linear' model) in this initial look at the data.

The final 'categorical' model (Table 5.4, (Stocken *et al.*, 2008), adapted) was based on 556 patients (520 deaths) with complete data on the prognostic factors of interest and identified six highly significant prognostic factors: LDH, albumin, metastases, WBC, CA19-9 and bilirubin:

Table 5.4: Dichotomising Continuous Variables

FINAL 'CATEGORICAL' MODEL (n=556 patients, 520 deaths)				
AIC=5583.2	**Variable**	**χ^2**	**p-value**	**HR (95% CI)**
Stratification factors	TRIAL	7.0	0.008	0.77 (0.64 to 0.94)
	*CANCER STAGE	0.9	0.36	1.16 (0.85 to 1.58)
	SEX	1.8	0.18	0.89 (0.74 to 1.06)
	TRT	3.4	0.065	1.19 (0.99 to 1.44)
Independent factors	ALBUMIN g/L	24.2	<0.001	2.30 (1.65 to 3.21)
	LDH IU/L	25.1	<0.001	2.05 (1.55 to 2.72)
	WBC 10^9/L	10.3	0.001	1.44 (1.15 to 1.79)
	+METS	11.9	<0.001	1.54 (1.21 to 1.97)
	BILIRUBIN μmol/L	8.4	0.004	1.34 (1.10 to 1.64)
	CA199 KU/l	9.4	0.002	1.48 (1.15 to 1.89)

HR=hazard ratio

* Stage (I/II vs. III/IV)

+ Metastases (negative vs. positive): missing data included in analysis as a separate 'classification' variable using lower level as the reference level

The final 'linear' model, assuming a linear relationship between independent predictors and log-hazard (Table 5.5, (Stocken *et al.*, 2008), adapted) was based on 556 patients (520 deaths) with complete data on the prognostic factors of interest and identified five highly significant prognostic factors: albumin, alkaline phosphatase, LDH, WBC and metastases:

Table 5.5: Assumption of Linearity of Continuous Variables

(shaded areas highlight factors identified in the 'Categorical' model)

FINAL 'LINEAR' MODEL (n=556 patients, 520 deaths)				
AIC=5557.1	**Variable**	**χ^2**	**p-value**	**HR (95% CI)^**
Stratification factors	TRIAL	8.1	0.005	0.76 (0.63 to 0.92)
	*CANCER STAGE	0.008	0.93	1.01 (0.74 to 1.38)
	SEX	3.8	0.051	0.84 (0.70 to 1.00)
	TRT	3.2	0.073	1.19 (0.98 to 1.44)
Independent factors	ALBUMIN g/L	41.0	<0.001	0.72 (0.65 to 0.79)
	LDH IU/L	13.3	<0.001	1.12 (1.05 to 1.19)
	WBC 10^9/L	11.7	<0.001	1.24 (1.10 to 1.41)
	+METS	10.5	0.001	1.50 (1.17 to 1.92)
	ALKPHOS IU/L	21.1	<0.001	1.13 (1.07 to 1.19)

HR=hazard ratio

* Stage (I/II vs. III/IV)

+ Metastases (negative vs. positive): missing data included in analysis as a separate 'classification' variable using lower level as the reference level

^ HR based on 5 unit increase in albumin g/L, WBC 10^9/L and 100 unit increase in LDH IU/L, alkaline phosphatase IU/L

Models were adjusted by stratification factors at randomisation (trial, sex, cancer stage) and randomised treatment group. 'Metastases' was selected as a significant independent covariate in both the final 'categorical' and 'linear' models. The inclusion of 'metastases' reduces the significant effect of 'cancer stage' seen at univariate analysis (Table 5.3) since these two variables are highly correlated explaining similar variability in the data.

The final selected 'linear' model (Table 5.5) has a substantially lower AIC of 5557.1 compared with the final selected 'categorical' model (Table 5.4) with an AIC of 5583.2. As such, the 'linear' model has a better fit even with one less prognostic factor based on five significant prognostic factors compared to six selected in the 'categorical' model. Both models considered albumin, LDH, metastases and WBC to be highly significant ($p<0.001$) independent prognostic factors. The 'linear' model considered alkaline phosphatase to also be important ($p<0.001$) but not in the categorical model, which considered CA19-9 ($p=0.002$) and bilirubin ($p=0.004$) to be important.

5.9. Conclusions

A multivariable approach, based on Cox proportional hazards modelling using backward elimination variable selection, was used to investigate multiple possible prognostic factors. These methods have the ability to include categorical (binary and classification variables) and continuous covariates. In previous published analyses of advanced pancreatic cancer data, continuous variables are usually dichotomised (categorised as normal/ abnormal based on central laboratory reference ranges) creating a step functional relationship with outcome. Continuous variables in the advanced pancreatic cancer dataset were dichotomised but since dichotomisation is unnecessary, they were also investigated under the assumption of linearity (a linear relationship with log-hazard). The strength of evidence for factors selected as prognostic differs between the two

'categorical and 'linear' models due to the underlying assumptions made regarding functional form. The final selected 'linear' model had a substantially lower AIC and hence was a better fit to the data. However, the underlying assumption of linearity may not be correct but has not been investigated previously in advanced pancreatic cancer. The relationships of continuous variables and outcome should be investigated further since alternative formats may be more appropriate and thus efficient. Non-linearity can be addressed through transformation of the variable in question. One transformation method based on fractional polynomial transformations has been used to address non-linearity in other cancer datasets, such as breast and renal, but not in pancreas. Chapter 6 investigates non-linearity of the continuous variables in the advanced pancreatic cancer dataset and investigates the applicability of non-linear fractional polynomial transformations. Addressing any non-linearity should lead to more accurate and efficient prognostic models which can then be internally validated, model performance assessed and the impact of missing data investigated further.

CHAPTER 6: FRACTIONAL POLYNOMIAL FUNCTIONS

- **SUMMARY**
- Correct modelling of non-linearity can be critical to estimation of prognostic factors
- Fractional polynomial regression models are particularly suitable for modelling smooth curved relationships between response and predictor
- Second-degree fractional polynomial functions (FP2) have a maximum of one turning point and degrees >2 are rarely required
- Fractional polynomial transformations can be implemented using SAS macros
- Model performance for a model based on fractional polynomials is confirmed

6.1. Introduction

The proportional hazards model assumes a linear relationship between independent covariates and log-hazard. The linearity assumption implies that each unit increase in variable X results in the same increase in risk of event (e.g. death) but the risk may increase to a certain limit then plateau or even decrease hence violating the assumption of linearity. A more appropriate analysis should retain the continuous nature of the variable allowing for some form of non-linear relationship with the response variable. Polynomials have been used to correct for non-linearity but low order polynomials are often too limited and high order polynomials often do not fit well at extreme data points. It is important to account for non-linearity especially in the prognostic factor setting since it has been shown (Royston *et al.*, 2006a) that a treatment effect can be underestimated unless a strongly prognostic non-linear covariate was modelled correctly using a suitable non-linear function, showing that correct modelling of non-linearity can be critical. Generally inclusion of an important mis-modelled non-linear covariate may cause unimportant variables correlated with it to enter the model spuriously (Royston and Sauerbrei, 2008). Also, important mis-modelled non-linear covariates may be excluded unless non-linearity is taken into

account (Stocken *et al.*, 2008). Transformations such as log X, √X or 1/X have been used in an attempt to improve model fit. If log X is defined as X^0 then these transformations can be seen to be different power functions X^p for different values of p.

6.2. Fractional Polynomial Functions

Fractional polynomial regression models provide a flexible parametric method for modelling curved relationships using few parameters and are particularly suitable for modelling smooth curved relationships between response and predictor (Royston and Altman, 1997). They were developed since smoothing splines were computationally intensive, do not provide predictive equations and do not extrapolate well (Royston and Altman, 1994). Fractional polynomials can be seen as an extension of Box and Tidwells' procedure for applying transformations in the ordinary least squares regression setting (Box and Tidwell, 1962). Fractional polynomials are an extended family of curves defining functions of the form $\beta_0+\beta_1 X^p$ where *p* is taken from the set {-2, -1, -0.5, 0, 0.5, 1, 2, 3} with non-integer and negative power terms as well a positive integer powers used in conventional polynomials, producing fractional polynomial models of degree 1 (FP1) with an additive predictor which is linear. A natural extension to the FP1 model is the FP*m* function where m is an integer ≥ 1 and are generalisations of quadratic functions. First-degree fractional polynomials (FP1) are simple transformations providing monotonic curves encompassing conventional polynomials defined as:

$$\varphi^*_1(X;p) = \beta_0+\beta_1 X^p = \beta_0+ \varphi_1(X;p)$$

Second-degree fractional polynomial functions (FP2) have a maximum of one turning point. The definition of an FP*m* function with power terms $p=(p_1, p_2)$ is:

$$\varphi^*_2(X;p) = \beta_0 + \beta_1 X^{p1} + \beta_2 X^{p2} = \beta_0 + X^p\beta = \beta_0 + \varphi_2(X;p)$$

With the set of *p* powers taken as {-2, -1, -0.5, 0, 0.5, 1, 2, 3} provides eight FP1 transformations, 28 FP2 transformations with distinct powers ($p_1 \neq p_2$) and eight FP2 transformations with equal powers ($p_1 = p_2$). The general definition of an FP*m* function is:

$$\varphi^*{}_m(X;p) = \beta_0 + \varphi_m(X;p) = \sum_{j=0}^{m} \beta_j H_j(X)$$

where X is a single covariate, m is a positive integer representing the degree of the fractional polynomial, p_j is a vector of powers, β_j is a vector of coefficients and:

$$H_j(X) = \begin{cases} X^{(p_j)} & \text{if } p_j \neq p_{j-1} \\ H_{j-1}(X)\log X & \text{if } p_j = p_{j-1} \text{ (Box-Tidwell transformation for repeated powers)} \end{cases}$$

where the expression $H_j(X)$ can be substituted for $\beta_j X_{ji}$ in the proportional hazards model (Equation 1, Chapter 5). The best powers *p* are those associated with the highest likelihood, hence lowest deviance (-2log-likelihood), chosen from the small preselected set p=[-max(3,m), …., -2, -1, -0.5, 0, 0.5 , 1, 2, .., max(3,m)] where 0 represents the natural logarithmic transformation $\log_e(X)$ and 1 represents no transformation (linear assumption). In health sciences, $m \leq 2$ provide enough flexibility and m>2 are rarely required and are associated with increased instability (Royston and Altman, 1994;Royston and Sauerbrei, 2008). Degree m is selected on *a priori* grounds or increasing m until no further significant improvement in fit is achieved. Models are fitted using maximum likelihood with the 'best' power vector being that with significantly lowest deviance (-2log-likelihood). The advantages of the fractional polynomial method are that implementation is simple and practical and that the resulting transformation may hold outside of the range of the observed data. There is a certain amount of inflexibility for manual model building in the multivariable setting, there are constraints in fitting a full model since variables may not be selected in the automated process.

Fractional polynomial transformations are implemented in SAS (SAS Institute Inc., 1999) using the %MFP8 macro (Meier-Hirmer *et al.*, 2003) for each continuous covariate, based on a closed test function selection procedure (Ambler and Royston, 2001) to maintain approximately the type 1 error rate. At the initial cycle all variables are sorted based on p-values of the full linear model. The best fitting fractional polynomial function is determined for the first covariate based on a difference in -2log-likelihood with one degree of freedom for each fractional polynomial term plus one for each power term. The functional form is retained and the process repeated for each covariate. Covariates with a larger p-value than the set alpha are excluded in this cycle but reinvestigated in the following cycle. The next cycle investigates each covariate in turn based on an initial functional format retained from the previous cycle. The macro terminates when the functional form does not change (convergence) based on differences in the -2log-likelihood, usually within 5 cycles.

6.3. Modelling with Fractional Polynomials in Advanced Pancreatic Cancer

The 'best' functional form of each continuous covariate was assessed in the univariate setting based on comparison of the AIC. Smaller values of the AIC indicate a better model but the 'best' model may be based on a complex transformation with almost identical AIC statistic to a less complex transformation. As such, the AIC had to be substantially lower to permit a complex transformation compared to a simple transformation. Substantial was indicated as a reduction in AIC of 4, based on an arbitrary decision rule taken from similar values as the chi-square distribution with 1 degree of freedom. If a simple log transformation reduced the AIC compared to the untransformed variable, then more complex fractional polynomial transformations were investigated. The use of log function as a screening tool for more complex functions may risk missing functions with a turning point but is a simple, appropriate, unbiased modelling strategy

adopted throughout. Second degree fractional polynomial transformations were investigated with a set of predefined powers {-2, -1, -0.5, 0, 0.5, 1, 2, 3} using a 5% level of significance. Each continuous covariate was then included in the multivariable analysis under either a linear, log or non-linear univariate fractional polynomial transformation.

Univariate analysis of the 10 continuous variables identified that non-linear transformations were appropriate for three variables in their relationship with survival: bilirubin μmol/L and LDH IU/L both as log transformations and CA19-9 KU/l as a second degree fractional polynomial transformation [CA19-$9^{0.5}$ + (CA19-$9^{0.5}$ x log(CA19-9))]. The seven remaining continuous covariates were analysed assuming a linear relationship with log-hazard (Table 6.1):

Table 6.1: Univariate AIC statistics

(shaded is selected transformation)

Variable	Linear	Log	FP
Age	6870.4	6870.8	Na
Albumin	6488.4	6493.7	Na
AlkPhos	6526.0	6532.4	Na
AST	6499.1	6498.5	Na
Bilirubin	6563.4	6555.6	6554.5
BUN	6462.9	6465.8	Na
CA19-9	6288.6	6233.6	6223.8
Haemoglobin	6347.0	6347.8	Na
LDH	6441.5	6425.4	6425.4
WBC	6318.5	6322.0	Na

FP=Fractional Polynomial, na=not applicable since log transformation not substantial improvement over linear assumption

As before, all variables were considered in the multivariable setting in their most appropriate functional form using Cox proportional hazards regression based on a backward elimination selection method using a 5% significance level for exclusion of variables and including trial, sex, cancer stage (stratification factors at randomisation) and randomised treatment as 'forced' variables in each model. A final check was made to ensure that no term in the model could be omitted and that none of the omitted terms could be included without significantly affecting the model. All variables omitted from the model based on backward elimination were considered and re-instated if they significantly improved the model. The final model based on transformed covariates ('FP' model) was based on 556 patients (520 deaths) with complete data on the prognostic factors of interest and identified seven prognostic factors. Five factors were highly significant with $p<0.001$: albumin, CA19-9, WBC, alkaline phosphatase and LDH with AST and BUN being more borderline in the model ($p=0.022$ and 0.027) (Table 6.2):

Table 6.2: Model based on Fractional Polynomial Transformations

(shaded highlight factors identified by both 'categorical' and 'linear' models)

FINAL 'FP' MODEL (n=556 patients, 520 deaths)				
AIC=5510.7	**Variable**	**χ^{2+}**	**p-value**	**HR (95% CI)^**
Stratification factors	TRIAL	14.5	<0.001	0.69 (0.57 to 0.83)
	*CANCER STAGE	1.8	0.18	1.19 (0.92 to 1.53)
	SEX	3.1	0.079	0.85 (0.71 to 1.02)
	TRT	3.0	0.082	1.19 (0.98 to 1.43)
Independent factors Linear	ALBUMIN g/L	42.3	<0.001	0.71 (0.64 to 0.79)
1st degree FP	Log(LDH)IU/L	14.5	<0.001	1.76 (1.31 to 2.34)
Linear	WBC 10^9/L	11.1	<0.001	1.25 (1.10 to 1.42)
Linear	ALKPHOS IU/L	14.2	<0.001	1.13 (1.06 to 1.20)
Linear	AST IU/L	5.3	0.022	0.91 (0.84 to 0.99)
Linear	BUN mmol/L	4.9	0.027	1.15 (1.02 to 1.30)
2nd degree FP	$CA199^{0.5}$ KU/l	51.5	<0.001	NA
2nd degree FP	$CA199^{0.5}$ KU/l x log(CA199)KU/l			NA

HR=hazard ratio, FP= Fractional Polynomial, NA=not appropriate

* Stage (I/II vs. III/IV), + Type-3 Wald χ^2 test

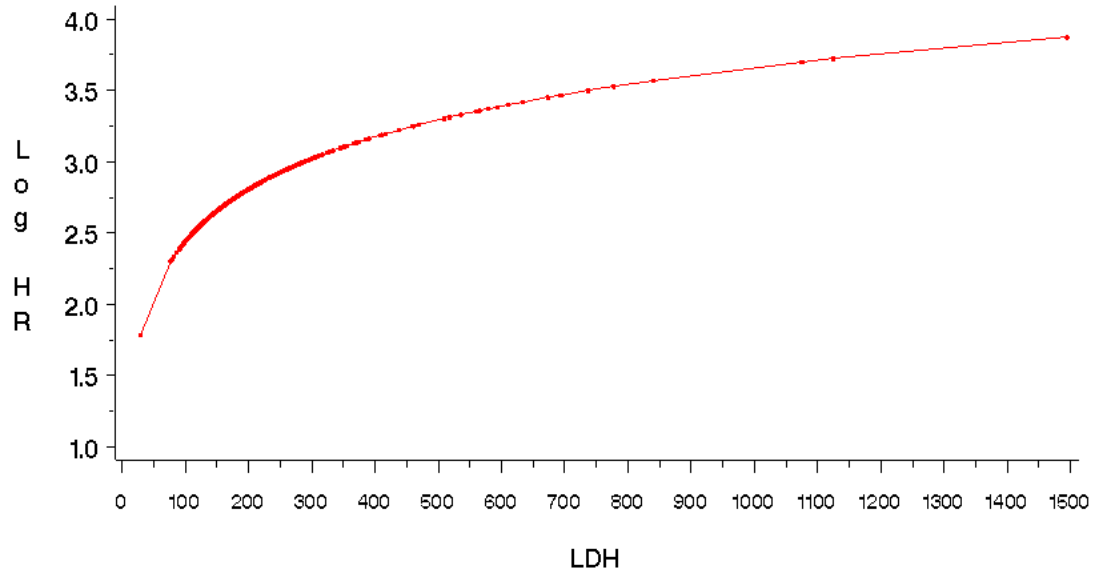
^ HR based on 5 unit increase in albumin g/L, WBC 10^9/L, BUN mmol/L; 25 unit increase in AST IU/L and 100 unit increase in alkaline phosphatase IU/L

Non-linear transformations were selected for LDH IU/L (log) and CA19-9 KU/l (second degree fractional polynomial).

The estimated function for LDH is shown graphically in Figure 6.1. The log function for LDH estimates increasing risk for increasing values of LDH:

Figure 6.1: Estimated Functional Form for LDH IU/L

(dots indicate actual data values)



The estimated log hazard ratio function for CA19-9 is shown graphically in Figure 6.2. The second degree fractional polynomial function for CA19-9 estimates increasing risk up to an approximate CA19-9 value of 14,000 KU/l and then decreases with increasing CA19-9:

Figure 6.2a: Estimated Functional Form for CA19-9 KU/l

(dots indicate actual data values)

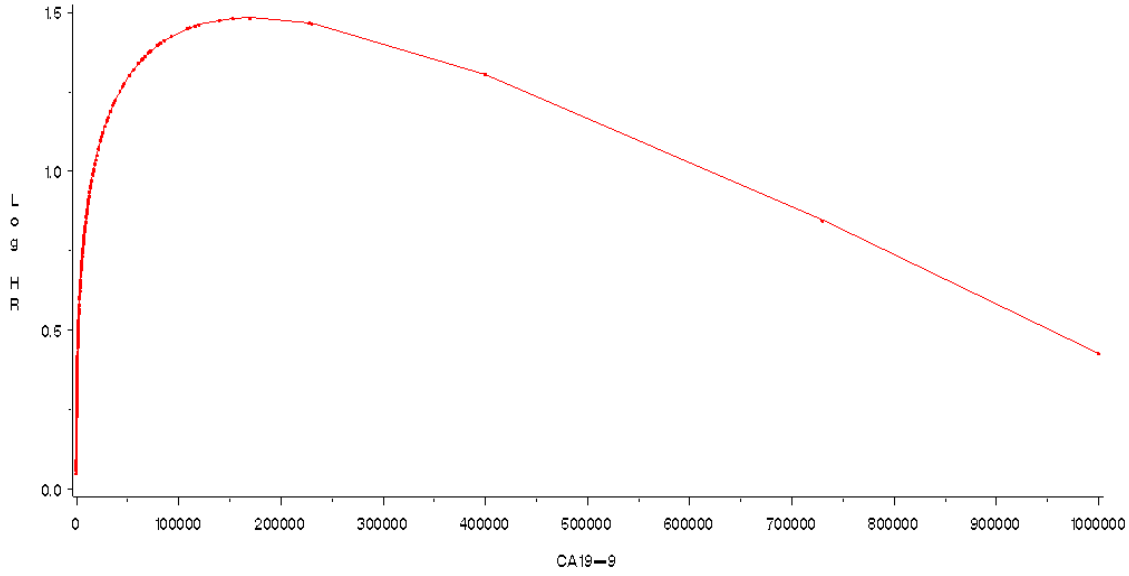
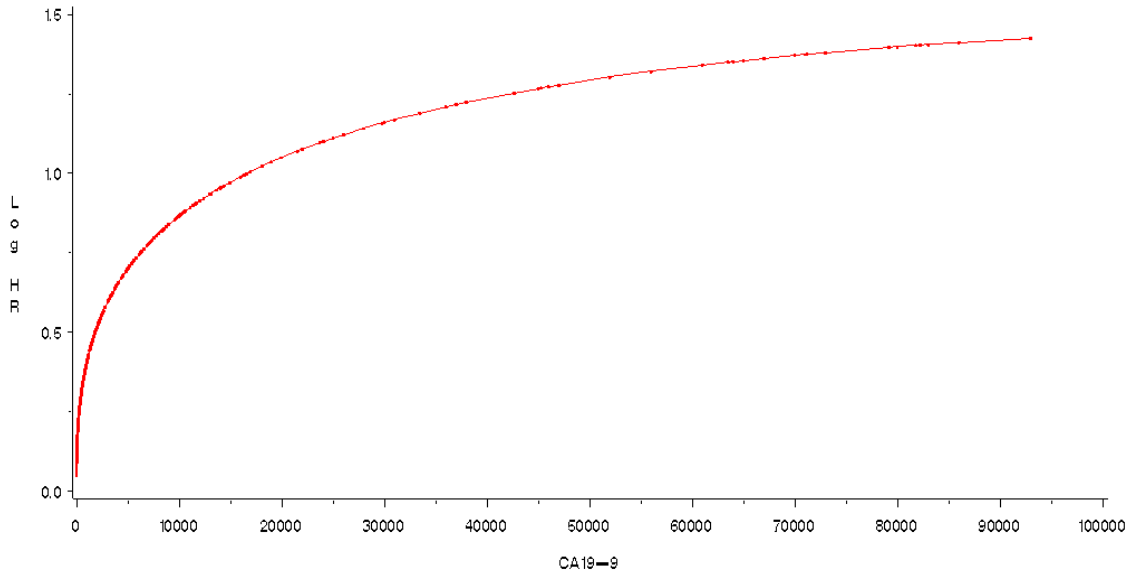


Figure 6.2b: Estimated Functional Form for CA19-9 KU/l

(dots indicate actual data values <100,000 KU/l)



Three final models have been presented under three differing assumptions for dealing with modelling continuous covariates: dichotomisation ('categorical' model, Table 5.4), assumption of linearity ('linear' model, Table 5.5) and non-linear fractional polynomial transformation ('FP' model, Table 6.2). The 'FP' model has the lowest AIC of 5510.7, substantially lower compared with an AIC of 5583.2 for the 'categorical' model and an AIC of 5557.1 for the 'linear' model. The 'FP' model selected seven important independent prognostic factors over and above the 'forced' variables, the 'categorical' model selected six and the 'linear' model selected five.

All three models considered albumin, LDH and WBC to be highly significant ($p<0.001$) independent prognostic factors. The strength of evidence as indicated by the Wald χ^2 statistic was similar for albumin in the 'linear' and 'FP' models but was underestimated by half in the 'categorical' model. The Wald χ^2 statistic was similar for LDH in the 'linear' and 'FP' models but was overestimated by double in the 'categorical' model. The strength of evidence was similar for WBC in all three models.

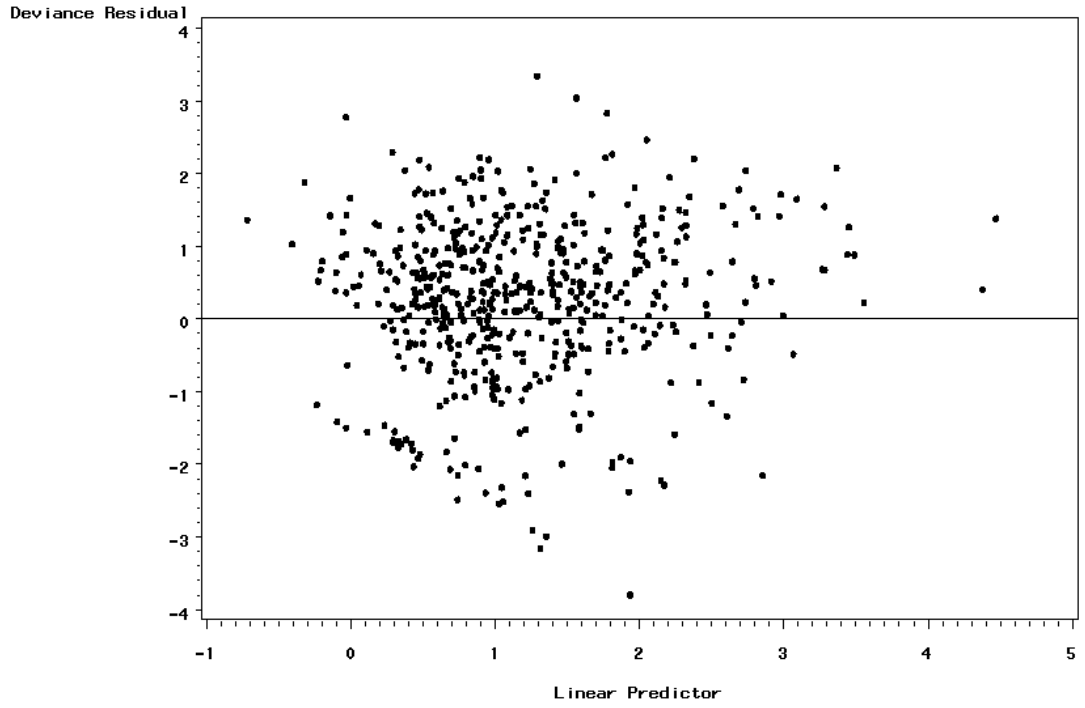
The parameter estimate and overall significance of metastases were reduced in the 'FP' model when continuous covariates were included in a more appropriate format. No non-linear transformation was required or applied to alkaline phosphatase IU/L which is why it was only selected as a highly significant and influential prognostic factor in both the 'linear' and 'FP' models. CA19-9 KU/l was also a highly significant and influential prognostic factor in both the 'categorical' and 'transformed' models. This variation is largely explained by the significant non-linear relation of CA19-9 to survival (Figure 6.2) and why it was considered important, albeit with massively reduced Wald χ^2 evidence, when dichotomised but not when assumed as linear. When considered as a transformed second degree fractional polynomial it's significance is much greater.

Bilirubin μmol/L was only selected as a highly significant factor in the 'categorical' model but was not included in either the 'linear' or 'transformed' models. AST IU/L and BUN mmol/L were only selected as borderline prognostic variables in the 'transformed' model when other factors had been modelled appropriately.

6.4. Model Performance

The AIC statistic was smallest for the 'FP' model (Table 6.2) and was substantially lower than either the 'categorical' or 'linear' models indicating a substantially better fit to the data. Deviance residuals for the 'FP' model were plotted against the linear predictor and were randomly scattered and centred symmetrically around a residual value of zero ranging between -3.80 and 3.34 which suggests the data have not been mis-modelled (Figure 6.3):

Figure 6.3: Deviance Residual Plot



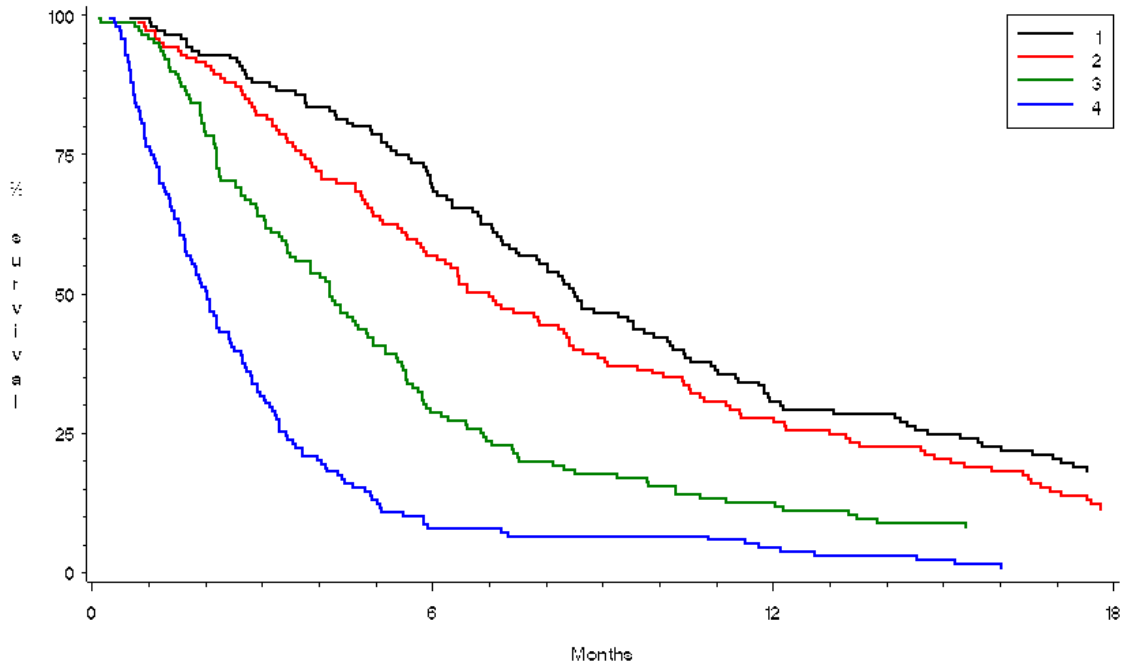
At least four risk groups of patients with advanced pancreatic cancer are expected (personal communication Professor Bass Hassan, 2004): i) very ill patients with survival $\leq$2 months not usually ideal candidates for therapy due to rapid disease progression or late presentation of large bulk disease; ii) fit patients with loco regional disease, low bulk metastatic disease and a median life expectancy of approximately 9 months; iii) and iv) two groups of patients with 'intermediate' prognosis between the two other clinical extremes, maybe patients with better prognosis but delayed clinical management. As such, patients were split into four equal sized groups based on quartiles of the distribution of linear predictor scores from the 'FP' model. Linear predictor scores (LP) were calculated as:

LP = $\eta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots\ldots + \beta_p X_{pi}$

LP = 0.37711*TRIAL + 0.17286*STAGEGP - 0.16115*SEX + 0.16934*TRTGP - 0.06802*ALB(g/L) + 0.02383*CA199FP1(KU/l) - 0.00173*CA199FP2(KU/l) + 0.04409*WBC(10^9/L) + 0.00119*ALKPHOS(IU/L) + 0.56220*LOGLDH(IU/L) - 0.00367*AST(IU/L) + 0.02772*BUN(mmol/L)

with the following code: TRIAL (1=BB128, 2=BB193), STAGEGP (1=I/II, 2=III/IV), SEX (1=MALE, 2=FEMALE), TRTGP (0=GEM, 1=GEMPLUS). Quartiles of the linear predictor scores (0.654140, 1.036423, 1.596797) created the four groups. Kaplan-Meier survival estimates (Figure 6.4) show four distinct prognostic groups with descending median survival estimates of 8.5 (95%CI: 7.4 to 10.1), 7.0 (95%CI: 5.8 to 8.4), 4.2 (95%CI: 3.5 to 5.0) and 2.0 (95%CI: 1.6 to 2.4) months and descending 12-month survival estimates of 30.5% (95%CI: 22.8% to 38.2%), 27.6% (95%CI: 20.1% to 35.1%), 12.5% (95%CI: 6.9% to 18.1%) and 4.3% (95%CI: 0.9% to 7.7%):

Figure 6.4: Survival by Prognostic Group



The hazard ratios for groups 2, 3 and 4 using prognostic group 1 as the baseline were: 1.26 (95%CI: 0.98 to 1.61), 2.02 (95%CI: 1.58 to 2.60) and 4.57 (95%CI: 3.55 to 5.90) respectively demonstrating distinct risk with almost non-overlapping confidence intervals of increasing risk of death.

A bootstrap re-sampling assessment of the extent of model optimism (over-fitting) in the final 'FP' model was carried out based on 200 bootstrap re-samples using the 'validate' statement in the Design Library of the programming language R (The R Foundation for Statistical Computing, 2008). Bootstrap samples are created by sampling one patient at a time with replacement from the original dataset up to the same size of the original set of data, repeated until 200 simulated bootstrap samples have been created. Bootstrap samples provide an unbiased estimate of the

stability of the estimates from the analysis of original data. Somers D_{xy} rank correlation is the measure of association between predicted log hazard and observed survival time on which extent of optimism is assessed (Harrell, 2001). The apparent index D_{xy} is calculated as -0.3989 but a better estimate of how well the model will discriminate prognoses in the future is the index corrected D_{xy} calculated as -0.3787. The estimate of model optimism is -0.0203 and as such is small (less than 5%).

The majority of variables had $\leq$5% missing values (Table 2.2). Tumour stage, CA19-9 and WBC were missing in less than 10% of patients and lymph node status was missing for 24% of patients. Metastases or lymph node status were considered in the analysis as classification variables using 'negative' as a reference level and including 'missing' as a classification level. Principal analysis was based on complete cases. Supportive analysis used multiple imputation to investigate the possible influence of variables with larger amounts of missing data and provided valid inferential alternative results. The MI procedure in SAS (SAS Institute Inc., 1999) was used to carry out multiple imputation based on five imputations using a Markov Chain Monte Carlo (MCMC) method (Schafer, 1997) that assumes multivariate normality to impute missing values. Each of the five complete data sets were analyzed using PHREG procedure in SAS (SAS Institute Inc., 1999) then the MIANALYZE procedure was used to generate valid statistical inferences about these parameters by combining results. Multiple imputation was a supportive analysis which allowed all 653 patients to be included in the modelling process (Table 6.3):

Table 6.3: Supportive Analysis of Imputed Dataset

(shaded highlight factors identified in all models)

FULL 'FP' MODEL BASED ON IMPUTED DATA (n=653 patients, 612 deaths)			
	Variable	**t**	**p-value**
Stratification factors	TRIAL	-4.1	<0.001
	*CANCER STAGE	-1.8	0.073
	SEX	-2.1	0.037
	TRT	2.1	0.033
Independent factors Linear	ALBUMIN g/L	-5.2	<0.001
1st degree FP	Log(LDH)IU/L	3.0	0.004
Linear	WBC 10^9/L	3.9	<0.001
Linear	ALKPHOS IU/L	3.5	<0.001
Linear	AST IU/L	-2.5	0.014
Linear	BUN mmol/L	2.1	0.034
2nd degree FP	$CA199^{0.5}$ KU/l	6.1	<0.001
2nd degree FP	$CA199^{0.5}$ KU/l x log(CA199)KU/l	-5.6	
Binary	METASTASES	3.2	0.001
Binary	RACE	-0.4	0.72
Binary	*T STAGE	0.1	0.89
Binary	NODES	2.7	0.016
Linear	AGE	1.4	0.15
Linear	HAEMOGLOBIN g/dL	-0.6	0.54
1st degree FP	Log(BILIRUBIN) μmol/L	1.8	0.067

HR=hazard ratio, FP= Fractional Polynomial

* Stage (I/II vs. III/IV)

The 'full' model including all covariates and all 653 patients (612 deaths) confirmed all the variables included in the 'FP' model but with increased significance for metastases (p=0.001) and the inclusion of nodal status (p=0.016) which had been excluded from all models prior to

imputation, suggesting a strong link to other variables already in the model requiring further investigation.

6.5. Centre Effect

A grouped frailty model (Therneau and Grambsch, 2000) is the term used to describe a survival model which incorporates a random effect (RE) element into hazard function to account for heterogeneity between patients. A random effect is a continuous variable that describes excess risk (frailty) for specific groups of patients. Those patients who are most 'frail' are expected to have shorter survival time. Known and expected frailties are associated with hospital and surgeon. The aim of a frailty model is to account for any 'centre effect' whilst adjusting for other prognostic factors. An important assumption is that frailty is independent of any censoring. The advanced pancreas clinical trials both stratified the randomisation procedure by centre and as such it seems sensible to investigate this further during the analysis. The frailty model is based on the proportional hazards model:

$$h_i(t) = e^{\beta X_i + \omega Z_i} h_0(t)$$

where X and β correspond to the fixed effects in the model (covariates such as age and sex), ω is a vector containing the unknown random effects (frailties) and Z is an indicator of the 'group' of dependent patients.

Frailty models are not easy to implement in SAS Version 9.1 (SAS Institute Inc., 1999) for the Cox proportional hazards model (although may be easier in the updated SAS v9.2) and as such data were exported into the programming language R (The R Foundation for Statistical Computing, 2008) where an approximate grouped frailty term is easier to include using a 'frailty()' statement in the 'Survival' Library based on a gamma distribution for the random effects. The frequent

assumption that frailty follows a gamma distribution is derived from the fact that the distribution for the frailty term must be positive since the hazard function cannot be negative.

The advanced pancreas dataset included patients from 60 randomising hospitals. A random effect associated with centre was non-significant (Wald χ^2=2.66, p=0.103) when forced into the existing fractional polynomial model.

6.6. Conclusions

The analysis of continuous variables in the advanced pancreatic cancer dataset has shown that non-linear transformations were appropriate for some covariates, but had not been previously investigated in this disease setting. The analysis including transformed variables resulted in a model which was a substantially better fit and included significant prognostic factors which were excluded in models under the assumptions of either step-functional (excluded alkaline phosphatase) or linear (excluded CA19-9) relationships. Through addressing the mis-modelling and non-linearity of continuous covariates, a more accurate prognostic model has been developed, displaying minimal overfitting. The model has the ability to be developed into a validated prognostic index based on robust statistical methodology, not previously presented in this disease site. Four distinct prognostic risk groups were identified based on information available in a routine NHS out-patient clinic. The transformation method used in the development of this model was based on fractional polynomial transformations but alternative transformation methods could also be used. Chapter 7 investigates an alternative more traditional method based on restricted cubic spline transformations and directly compares both transformation methods for the analysis of non-linear continuous variables in the advanced pancreatic cancer dataset. Chapter 8 continues the comparison of methods in an alternative disease with alternative survival

distribution based on a large dataset in cardiac surgery patients. Since an external dataset was anticipated, external validation of the prognostic model in advanced pancreatic cancer was possible and is the focus of Chapter 9.

CHAPTER 7: COMPARISON OF NON-LINEAR FUNCTIONS

- **SUMMARY**
- Fractional polynomials is one approach for analysing non-linear covariates, an alternative is the restricted cubic spline approach
- Restricted cubic splines are linear in estimated coefficients and restricted to be linear in the tails therefore allowing use of standard methods of inference
- Restricted cubic spline transformations can be implemented using SAS macros
- A bootstrap validation compares the model fit of the two polynomial based approaches

7.1. Introduction

The use of fractional polynomials is one of various approaches for analysing independent continuous covariates which do not comply with the assumption of linearity. An alternative and more conventional method is the restricted cubic spline approach which also utilises the full information when handling continuous variables with the aim of identifying important non-linear continuous prognostic factors.

There is a certain amount of inflexibility for manual model building in the multivariable setting using the fractional polynomial macros and there are constraints in fitting a full model since continuous variables may not be selected in the automated process. As such, comparison with the more conventional restricted cubic spline method was considered. An advantage of the restricted cubic spline method is that no particular functional form needs to be specified. A drawback is the arbitrary choice of the number and position of knots of the cubic spline function which must be pre-specified but in most instances are not known. Extrapolation beyond the outer knots has also been found to be spurious (Harrell, 2001). The difference between two spline based procedures

and the fractional polynomial methods was investigated in longitudinal cohort studies in lung cancer (Govindarajulu *et al.*, 2007). The area under three curves were compared (two spline based, one fractional polynomial) and concluded that the two spline based curves were more similar than either was to the fractional polynomial curve and that fractional polynomials were more global in nature in contrast to a more local fit of splines.

In this chapter these two polynomial based strategies are compared directly for determining the functional form of the non-linear relationship between a continuous prognostic covariate and survival. A bootstrap validation was carried out which directly compared the models based on the difference in AIC statistics for the two polynomial based approaches within each bootstrap resample.

7.2. Restricted Cubic Spline Functions

The name 'spline' came from a draftman's spline which is a flexible strip used to draw curves. Spline functions in statistics (Stone and Koo, 1985) are piecewise polynomials that are used for fitting curved relationships, that is polynomials connected across intervals of X and constrained to join at the interval endpoints or 'knots'. The simplest spline function is the linear spline function, that is the functions within each interval are linear and joined at the knot positions. The expression for a spline function *S(X)* of variable *X* (Harrell, 2001;Durrleman and Simon, 1989) is given as:

$$S(X) = \sum_{j=0}^{n} \beta_{0j} X^{j} + \sum_{i=1}^{k} \beta_{in} (X - t_i)^{n}$$

and can be substituted for $\beta_j X_{ji}$ in the proportional hazards model (Equation 1, Chapter 5) where n is the degree of the spline function, k is the number of knots, t is the position of the knots (t1<t2<…<tk) or more practically:

$S(X)=\beta_0 + \beta_1 X + \beta_2(X-a) + \beta_3(X-b) + \beta_4(X-c) + ...$

where *a, b* and *c* are the position of the knots. As such, a spline function of variable X in a multivariable model will introduce k+n new variables and regression coefficients determined to optimise fit usually based on the log-likelihood function. The spline is not constrained to be continuous or 'smooth' at the knots and as such does not fit highly curved functions well (Harrell, 2001). Cubic polynomials have the ability to fit sharp curves and can be constrained to be smooth at knot positions *a, b* and *c*:

$S(X)=\beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4(X-a)^3 + \beta_5(X-b)^3 + \beta_6(X-c)^3 + ...$

requiring estimation of *k+3* parameter estimates for each of *k* knots. Cubic splines have been found to behave poorly in the tails, before the first and after the last knot position (Stone and Koo, 1985). Restricted cubic splines are designed to control this unsatisfactory behaviour in the tails. Restricted cubic splines are linear in estimated coefficients and restricted to be linear in the tails therefore allowing use of standard methods of inference and computation requiring estimation of k-1 regression coefficients for each of *k* knots $t_1, ..., t_k$:

$S(X)=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_{k-1} X_{k-1}$

where: $X_1=X$

and for j=1, ..., k-2: $X_{j+1} = (X-t_j)^3 - (X-t_{k-1})^3(t_k-t_j) / (t_k-t_{k-1}) + (X-t_k)^3(t_{k-1}-t_j) / (t_k-t_{k-1})$

The advantages of the restricted cubic spline method are that no particular functional form needs to be specified and the method produces smooth curves. Although the basis for restricted cubic splines allows for extrapolation beyond the outer knots, when these occur in the tails of the data, it has been found to be spurious (Harrell, 2001). A drawback is the arbitrary choice of the number and position of knots of the cubic spline function which must be pre-specified and that each k-1 spline component spends an additional degree of freedom. Model fit has been shown to depend more on the number of knots rather than location (Stone, 1986) and 3 to 5 knots will usually

suffice (Durrleman and Simon, 1989;Heinzl *et al.*, 1996;Hess, 1994) set at fixed percentiles of the observed data distribution to guard against influence of outliers on knot placement (Durrleman and Simon, 1989).

Restricted cubic spline transformations were implemented using the %RCSPLINE macro (Harrell, 1991) in SAS (SAS Institute Inc., 1999) to compute the k-1 components of the cubic spline function restricted to be linear before the first and after the last knots for each continuous covariate. Models with 3, 4 and 5 knots were compared based on knot positions identified by the SAS %DASPLINE macro (Harrell, 1991). The %RCSPLINE macro generates formulas to create classification variables that allow fitting of splines constrained to be linear in the tails and to be fitted in multivariable models. Knot positions identified by %DASPLINE are based on quanitiles according to the number of knots specified but forced to be real data points:

Knots	Quantiles					
3	.05	.5	.95			
4	.05	.35	.65	.95		
5	.05	.275	.5	.725	.95	
6	.05	.23	.41	.59	.77	.95
7	.025	.18333	.34166	.5	.65833	.81666	.975

7.3. Univariate Comparison of Functions in Advanced Pancreatic Cancer

Principal analyses of the advanced pancreatic cancer dataset was based on the assessment of the functional form of each continuous covariate in the univariate setting selecting a ‘best’ fitting fractional polynomial and also a ‘best’ fitting restricted cubic spline (with varying number and position of knots) based on comparison of AIC. Smaller values of the AIC indicate a better model but the ‘best’ model may be based on a complex transformation with almost identical AIC statistic

to a less complex transformation. As such, the AIC had to be substantially lower to permit a complex transformation compared to a simple transformation. Substantial was indicated as a reduction in AIC of 4, based on an arbitrary decision rule taken from similar values as the chi-square distribution with 1 degree of freedom. As before, the strategy was that if a simple log transformation reduced the AIC compared to the untransformed variable, then more complex non-linear transformations were investigated further using restricted cubic splines (based on 3, 4 and 5 knot models) and second degree fractional polynomials.

Univariate analysis (Stocken *et al.*, 2008) of the ten continuous variables identified that non-linear transformations were appropriate for three variables in their relationship with survival: bilirubin, LDH and CA19-9 (Table 7.1):

Table 7.1: Univariate AIC statistics

(shaded is selected transformation with lowest AIC)

Variable	Linear	Log	RCS-3	RCS-4	RCS-5	FP
Age	6870.4	6870.8	na	na	na	na
Albumin	6488.4	6493.7	na	na	na	na
AlkPhos	6526.0	6532.4	na	na	na	na
AST	6499.1	6498.5	na	na	na	na
Bilirubin	6563.4	6555.6	6556.0	6557.9	6559.0	6554.5
BUN	6462.9	6465.8	na	na	na	na
CA19-9	6288.6	6233.6	6235.8	*	*	6223.8
Haemoglobin	6347.0	6347.8	na	na	na	na
LDH	6441.5	6425.4	6430.1	6429.4	6431.1	6425.4
WBC	6318.5	6322.0	na	na	na	na

RCS-k=Restricted Cubic Spline with k knots, FP=Fractional Polynomial
na=not applicable since log transformation not substantial improvement over linear assumption
* contains spline variables that are not included in the model, df=0 hence use 3-knot RCS

Restricted cubic spline transformations for bilirubin μmol/L, LDH IU/L and CA19-9 KU/l were investigated based on 3, 4 and 5 knot models. For bilirubin and LDH, the AIC was not reduced by any spline polynomial and was not substantially reduced by the more complex fractional polynomial function (bilirubin^-0.5 and log(LDH) respectively) transformation compared to a simple log transformation. For CA19-9, the 5 knot restricted cubic spline transformation appeared to be better fit compared to the simple log transformation and better than the 3 or 4 knot spline models. However models with >3 knots introduced spline covariates that were not included in the model with degrees of freedom equal to 0 and as such the 3 knot model was used with knot positions at CA19-9 KU/l values of 10, 710 and 70,000. An alternative approach could be based on rescaling or truncating. The second degree fractional polynomial transformation identified for CA19-9 KU/l was [CA199^(0.5) + (CA199^(0.5)*log(CA199))] and had substantially reduced AIC compared to the log or spline transformations (Table 7.1). The seven remaining continuous covariates were analysed assuming a linear relationship with log-hazard.

The estimated risk over increasing values of CA19-9 KU/l was plotted for each transformation method to enable graphical comparison of the estimated functional format (Figure 7.1). The plots of the estimated functional form appear similar for CA19-9 but with a steeper elevated increasing hazard in lower values of CA19-9 less than 30,000 KU/l using the fractional polynomial function. The linear predictor was scaled so the log(HR) was zero at the midpoint of each continuous covariate to enable plots on the same axes, altering the fractional polynomial plot from that in Chapter 6 (Figure 6.2).

Figure 7.1a: Estimated Functional Form for CA19-9 KU/l: Restricted Cubic Spline Transformation

(dots indicate actual data values)

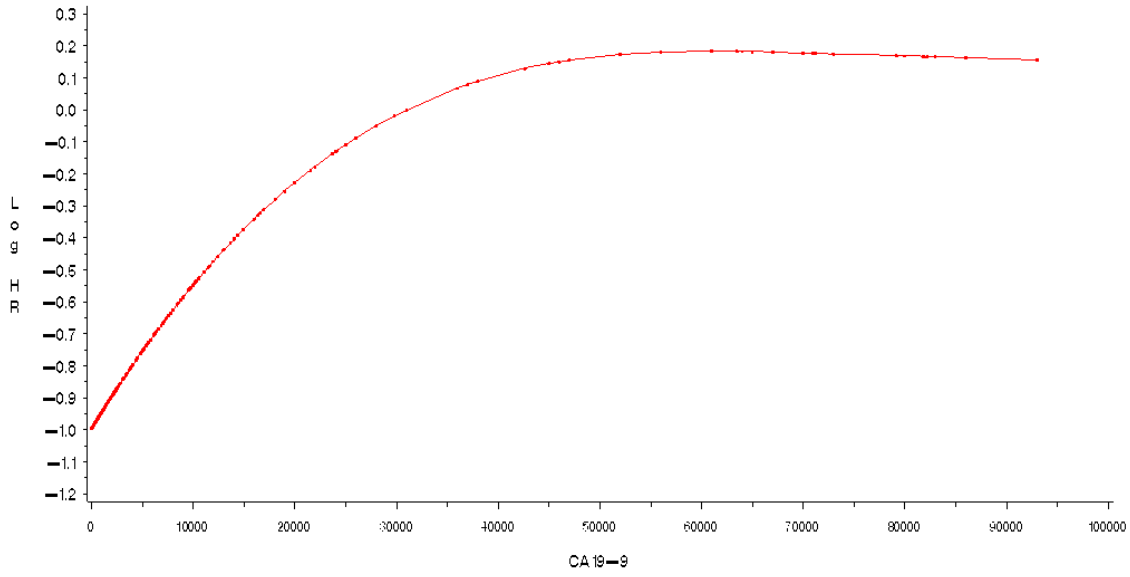
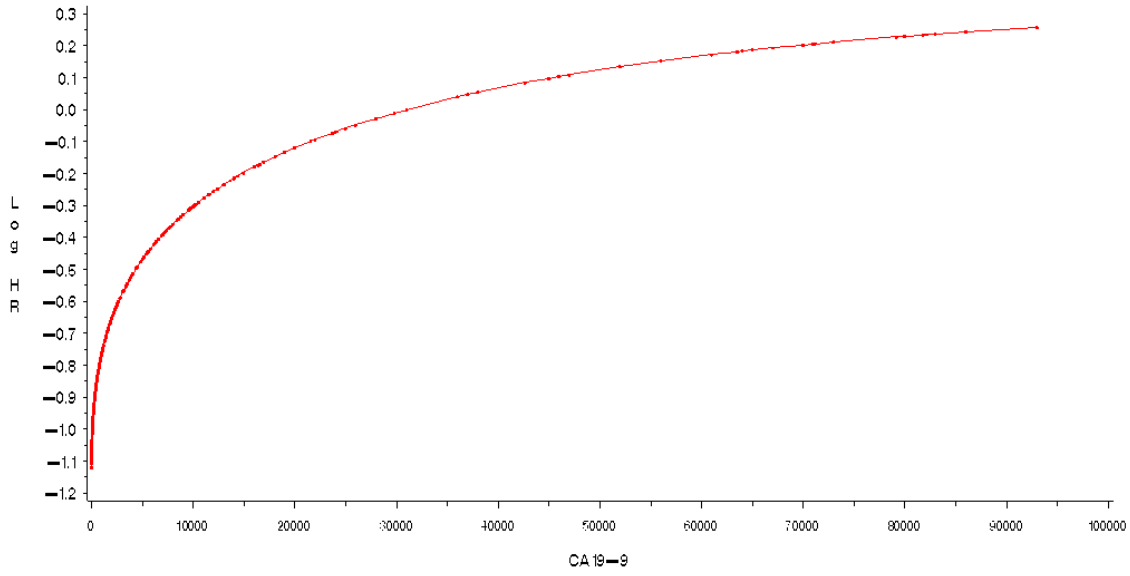


Figure 7.1b: Estimated Functional Form for CA19-9 KU/l: Fractional Polynomial Transformation

(dots indicate actual data values)



7.4. Multivariable Comparison of Functions in Advanced Pancreatic Cancer

Each continuous covariate was included in the multivariable analysis under either a linear, log or non-linear univariate transformation based on the 'best' fitting restricted cubic spline identified and the 'best' fitting fractional polynomial identified. As before, all variables were then considered in the multivariable setting using Cox proportional hazards regression based on a backward elimination selection method using a 5% significance level for exclusion of variables and including trial, sex, cancer stage (stratification factors at randomisation) and randomised treatment as 'forced' variables in each model. A final check was made to ensure that no term in the model could be omitted and that none of the omitted terms could be included without significantly affecting the model. All variables omitted from the model based on backward elimination were considered and re-instated if they significantly improved the model. Two final Cox proportional hazards regression models (Table 7.2) were developed based on i) restricted cubic spline and ii) fractional polynomial transformations, which could be directly compared based on AIC statistics:

Table 7.2: Cox Proportional Hazards Regression Models (n=556, deaths=520), shaded highlight factors identified by both models

Restricted Cubic Spline Transformations						
AIC=5526.2	**Transformation (knot position)**	**log HR (β)**	**se(β)**	**χ^2_W**	**p-value**	**HR (95% CI)^**
Stratification factor	TRIAL	-0.35	0.099	12.7	<0.001	0.70 (0.58 to 0.85)
Stratification factor	* STAGE	0.21	0.13	2.6	0.11	1.23 (0.96 to 1.59)
Stratification factor	SEX	-0.19	0.090	4.2	0.040	0.83 (0.70 to 0.99)
Stratification factor	TRT	0.18	0.098	3.5	0.063	1.20 (0.99 to 1.45)
Continuous factor Linear	ALB g/L	-0.070	0.010	45.4	<0.001	0.70 (0.64 to 0.78)
Continuous factor RCS3	# CA199 (10,710,70000)KU/l	0.000047	0.000007	43.3	<0.001	NA
	CA1991 KU/l	-0.0016	0.00025			NA
Continuous factor Linear	WBC 10^9/L	0.045	0.013	12.2	<0.001	1.26 (1.11 to 1.43)
Continuous factor Linear	ALKPHOS IU/L	0.00090	0.00028	10.5	0.0012	1.10 (1.04 to 1.16)
Continuous factor Log	Log(LDH)IU/L	0.39	0.14	8.4	0.0039	1.48 (1.14 to 1.93)
Fractional Polynomial Transformations						
AIC=5510.7	**Transformation**	**log HR (β)**	**se(β)**	**χ^2_W**	**p-value**	**HR (95% CI)^**
Stratification factor	TRIAL	-0.38	0.099	14.5	<0.001	0.69 (0.57 to 0.83)
Stratification factor	* STAGE	0.17	0.13	1.8	0.18	1.19 (0.92 to 1.53)
Stratification factor	SEX	-0.16	0.092	3.1	0.079	0.85 (0.71 to 1.02)
Stratification factor	TRT	0.17	0.097	3.0	0.082	1.19 (0.98 to 1.43)
Continuous factor Linear	ALB g/L	-0.068	0.010	42.3	<0.001	0.71 (0.64 to 0.79)
Continuous factor FP2	# CA199^(0.5)KU/l	0.024	0.0040	51.5	<0.001	NA
	CA199^(0.5)*log(CA199)KU/l	-0.0017	0.00031			NA
Continuous factor Linear	WBC 10^9/L	0.044	0.013	11.1	<0.001	1.25 (1.10 to 1.42)
Continuous factor Linear	ALKPHOS IU/L	0.0011	0.00032	14.2	<0.001	1.13 (1.06 to1.20)
Continuous factor FP1	Log(LDH)IU/L	0.56	0.15	14.5	<0.001	1.76 (1.31 to 2.34)
Continuous factor Linear	AST IU/L	-0.0037	0.0016	5.3	0.022	0.91 (0.84 to 0.99)
Continuous factor Linear	BUN mmol/L	0.028	0.013	4.9	0.027	1.15 (1.02 to 1.30)

HR=Hazard Ratio, RCS3=3 knot Restricted Cubic Spline, FP1/2=1st or 2nd degree Fractional Polynomial, NA=not appropriate

* Stage (I/II vs. III/IV); # CA199: RCS knots positioned at (10, 710, 70000), Type III Wald χ^2 test presented with 2 degrees of freedom

^ HR based on 5 unit increase in albumin g/L, WBC 10^9/L, BUN mmol/L; 25 unit increase in AST IU/L and 100 unit increase in alkaline phosphatase IU/L

The sign of each corresponding variable in the models were identical and the significance of each variable similar except for LDH which was reduced in the restricted cubic spline model. Parameter estimates only differed substantially for the two transformed variables CA199 and LDH. The AIC was smallest for the model under the fractional polynomial transformation (5510.7 compared with 5526.2). However the fractional polynomial model selected an additional two variables with borderline significance: AST (p=0.022) and BUN (p=0.027). The comparable AIC of the restricted cubic spline model including AST (p=0.050) and BUN (p=0.067) was 5522.7 showing the fractional polynomial model to remain a better fit to the data by a reduction of 12.0.

7.5. Bootstrap Validation Comparing Non-Linear Functions

Internal validation, to directly compare the fit of the restricted cubic spline and fractional polynomial strategies, was carried out by calculating the sampling distribution of the difference in the AIC statistics between the models in multiple bootstrap resamples of the data using an in-house developed SAS program (Appendix) based on nonparametric bootstrap analyses (Efron and Tibshirani, 1993). The bootstrap analysis fitted the final models, one based on the restricted cubic spline transformations and one based on the fractional polynomial transformations, to a series of 1000 bootstrapped resamples taken from and with the same size as the original data. Smaller values of the AIC indicate a better model but it is unclear whether AIC statistics from non-nested models will approximate the chi-square distribution for hypothesis testing. Using a fixed level of significance for the comparison of AIC is not appropriate in non-nested models with the same parameters. The difference in AIC statistics for the different models in multiple bootstrap resamples provides a distribution which can then be summarised appropriately with bootstrap percentile confidence intervals derived from the bootstrap distribution of the difference between the restricted cubic spline and fractional polynomial models.

The %BOOT macro (Sarle, 2000) in SAS (SAS Institute Inc., 1999) is very computationally intensive carrying out non-parametric bootstrap analyses for simple random samples to produce inferences such as approximate standard errors, bias-corrected estimates and confidence intervals without knowing the type of distribution from which a sample has been taken assuming a normal sampling distribution. The %BOOT macro was used to calculate 1000 resamples of the original dataset using identical univariate transformations, if any, as in the original dataset.

Modelling was carried out in 1000 bootstrap resamples forcing in trial, stage, sex, trt and CA19-9 components (restricted cubic spline or fractional polynomial) and considering all other variables for inclusion (including the borderline variables of AST and BUN) based on backward elimination variable selection method. Two models were created within each bootstrap sample, one based on restricted cubic spline transformations and one based on fractional polynomial transformations. The AIC for each model was output from the screen to a temporary SAS dataset and the difference (AIC[RCS]-AIC[FP]) between them calculated, a positive difference indicating a better fit by the fractional polynomial transformation model. This was repeated in each 1000 bootstrap resamples and the mean difference calculated with 95% confidence interval (Table 7.3). The AIC statistics had to be output from the screen rather than delivered to an output dataset within the modelling process, as the two AIC values differed. This was reported to SAS who replied “you have an excellent point” and confirmed they would investigate further and recommended outputting the AIC from screen using an ‘ods output’ statement instead.

Table 7.3: Bootstrap Comparison of AIC based on 1000 Resamples

	Bootstrap Mean AIC	Bootstrap Percentile 95% CI
Restricted Cubic Spline	5498.1	5441.7 to 5667.1
Fractional Polynomial	5487.1	5419.9 to 5648.5
Difference (RCS-FP)	11.0	9.0 to 31.4

The bootstrap mean AIC (Table 7.3) was smallest for the model containing fractional polynomial transformations and on average was a better fit than the model based on restricted cubic spline transformations by a bootstrap mean reduction of 11.0 (bootstrap percentile 95% CI: 9.0 to 31.4).

7.6. Conclusions

Fractional polynomial and restricted cubic spline functions are both polynomial functions particularly suitable for modelling smooth curved relationships between response and a predictor and both are easily implementable in SAS. Both models containing non-linear transformations in the advanced pancreatic cancer dataset gave a substantially better fit compared to the models which dichotomised or assumed linearity of continuous covariates. The fitted functions generated by restricted cubic splines and fractional polynomials were similar but the model AIC and bootstrap mean difference in AIC was smallest for the fractional polynomial model. The methods were generally different in the extremities, in the left-hand tails for CA19-9 (values less than 30,000 KU/l), where there is often a paucity of data. Due to the availability of an extremely large dataset in a different disease site, the reproducibility of these results could be investigated in data with an alternative event rate and survival distribution to investigate the stability of the conclusions. Analyses of data from 42802 cardiac surgery patients are reported in Chapter 8.

CHAPTER 8: CARDIAC SURGERY EXAMPLE

- **SUMMARY**
- Prospective data were collected on 44902 patients undergoing cardiac surgery
- 5563 (12.4%) patients had died and median follow-up was 5.2 years
- Further analyses possible due to the size of dataset with alternative event rate and survival distribution
- Aim to investigate reproducibility of results seen when comparing non-linear methods in advanced pancreatic cancer dataset
- In addition, a univariate unadjusted fractional polynomial transformation was recalculated within each of 200 bootstrap resamples and compared directly against a 5-knot restricted cubic spline
- The influence of the size of the bootstrap samples was investigated

8.1. Introduction

Cardiovascular disease is the most common cause of premature death in the Western world and is closely related to socio-economic deprivation. Cardiac surgery includes a number of operations known to carry significant prognostic benefit. Prospective data were collected on 44902 patients undergoing cardiac surgery, followed for an median of 5.2 years, to assess whether social deprivation based on post-codes using the Carstairs score influences survival following surgery (Pagano *et al.*, 2009). The event rate (5563 (12.4%) patients had died) and survival distribution (approximate 90% survival at 5-years) in this disease site is very different to advanced pancreatic cancer.

The aim of the analysis was to investigate the reproducibility of the results and conclusions drawn from the comparison of non-linear methods in the advanced pancreatic cancer dataset, where the

fractional polynomial transformation gave a better fit compared to the restricted cubic spline transformation. Analysis was carried out based on 42802 patients (5486 deaths) with complete data on the prognostic factors of interest. Seven baseline clinical and demographic variables were considered potential prognostic factors of survival (randomising centre, smoking status, diabetes, surgical procedure, body mass index (BMI), Carstairs score (CS) and EuroScore (ES)), three collected as continuous measurements (BMI, CS and ES). Due to the extreme size of this cardiac surgery dataset, further comparison of the two polynomial based strategies was able to be carried out. Principal analysis investigated each continuous covariate in turn but allowing the univariate fractional polynomial transformation to be recalculated within each of 200 bootstrap resamples taken from the original dataset and compared against a 5-knot restricted cubic spline. Supportive analysis investigated the influence of sample size by decreasing the bootstrap resample sizes compared to the size of the original dataset.

8.2. Univariate Comparison of Non-Linear Functions

Analyses were based on an identical strategy adopted in the analysis of the advanced pancreatic cancer dataset. Principal analysis was based on the assessment of the functional form of each continuous covariate in the univariate setting selecting a ‘best’ fitting fractional polynomial and also a ‘best’ fitting restricted cubic spline (with varying number and position of knots) based on comparison of AIC. Smaller values of the AIC indicate a better model but the ‘best’ model may be based on a complex transformation with almost identical AIC statistic to a less complex transformation. As such, the AIC had to be substantially lower to permit a complex transformation compared to a simple transformation. Substantial was indicated as a reduction in AIC of 4, based on an arbitrary decision rule taken from similar values as the chi-square distribution with 1 degree of freedom. As before, the strategy was that if a simple log transformation reduced the AIC

compared to the untransformed variable, then more complex non-linear transformations were investigated further using restricted cubic splines (based on 3, 4 and 5 knot models) and second degree fractional polynomials.

Univariate analysis of the three continuous variables (BMI, CS and ES) identified that non-linear transformations were appropriate in their relationship with survival (Table 8.1). However, there is a historically accepted transformation for EuroScore which was analysed throughout under a log transformation. Restricted cubic spline transformations for BMI and CS were investigated based on 3, 4 and 5 knot models.

Table 8.1: Univariate AIC Statistics

(shaded is selected transformation)

Variable	Linear	Log	RCS-3	RCS-4	RCS-5	FP
BMI	112643.0	112598.8	112497.3	112495.8	112497.8	112514.7
Carstairs score	114547.7	114538.6	114533.9	114533.1	114535.0	114532.9
***Euroscore**	111920.9	111769.8	111720.7	111722.6	111723.7	111712.6

RCS-k=Restricted Cubic Spline with k knots, FP=Fractional Polynomial

* historical transformation is log

AIC was not substantially reduced by restricted cubic spline transformations based on more than 3 knots. Knot positions were set at BMI values of 20.999, 27.110 and 35.339 and at CS of -3.507, -0.539 and 5.658. Second degree fractional polynomial transformations were identified for BMI as [BMI^0.5 + (BMI^0.5*log(BMI))] and CS as [(CS+5.2) + ((CS+5.2)^2)]. The fractional polynomial transformation had similar AIC to the restricted cubic spline for CS and was worse for BMI.

The estimated risk over increasing values of each continuous covariate were plotted for each transformation method to enable graphical comparison of the estimated functional format. The linear predictor was scaled so the log(HR) was zero at the midpoint of each continuous covariate to enable plots on the same axes. The plots of the estimated functional forms appear similar for BMI (Figure 8.1) with smoother transformation for peak values between 25 to 45 with fractional polynomial but with higher estimated risk in the tail for scores greater than 50 compared with the restricted cubic spline function, which was a substantially better fit at univariate analysis:

Figure 8.1a: Estimated Functional Form for BMI: Restricted Cubic Spline Transformation

(dots indicate actual data values)

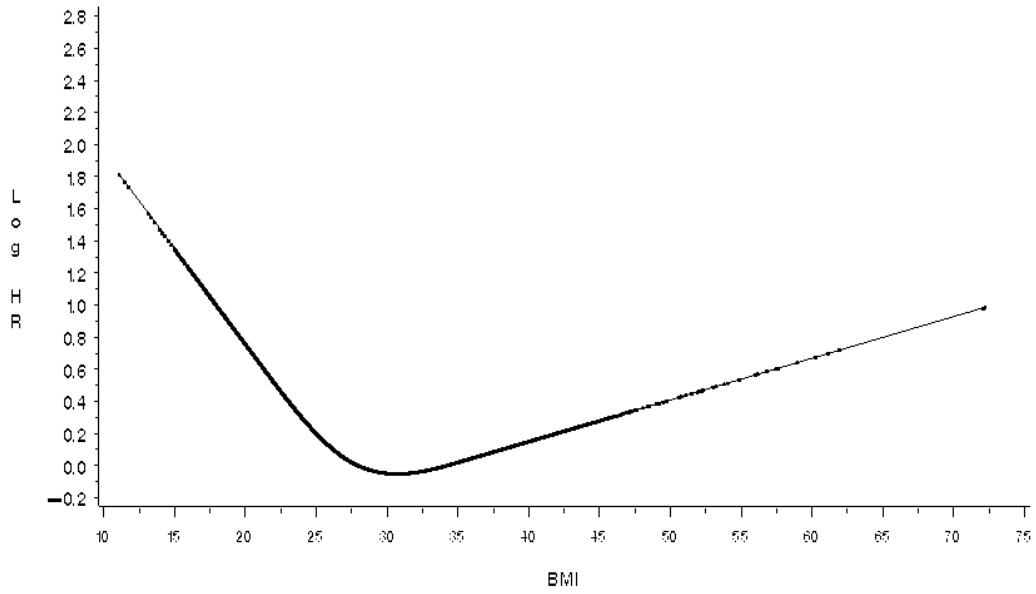
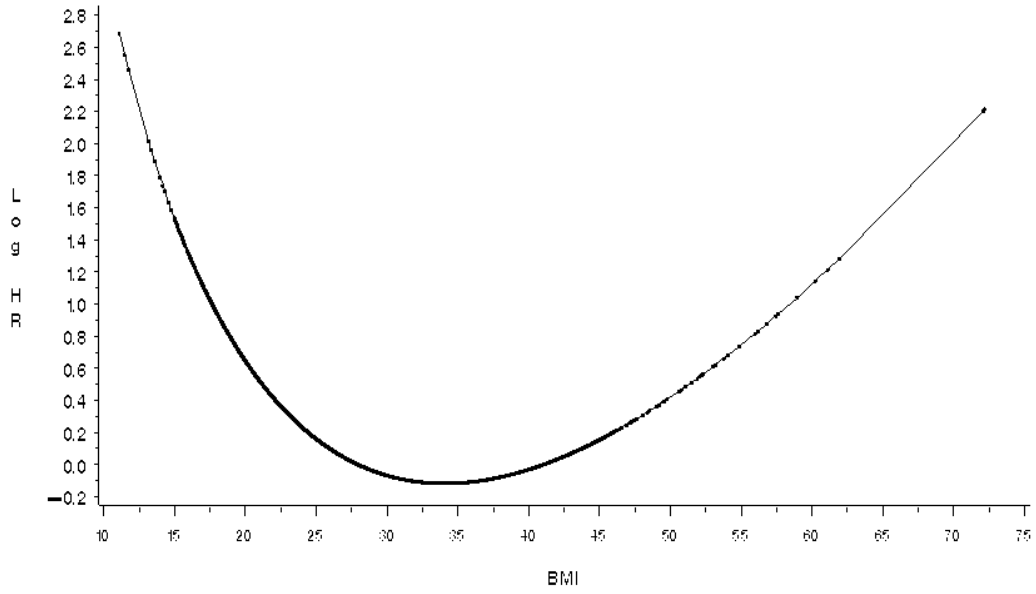


Figure 8.1b: Estimated Functional Form for BMI: Fractional Polynomial Transformation

(dots indicate actual data values)



Similarly, for Carstairs score, the plots of the estimated functional forms (Figure 8.2) appear similar but with lower estimated risk in the tail for scores greater than 10 compared with the restricted cubic spline function, both having a similar fit to the data at univariate analysis:

Figure 8.2a: Estimated Functional Form for Carstairs Score: Restricted Cubic Spline

(dots indicate actual data values)

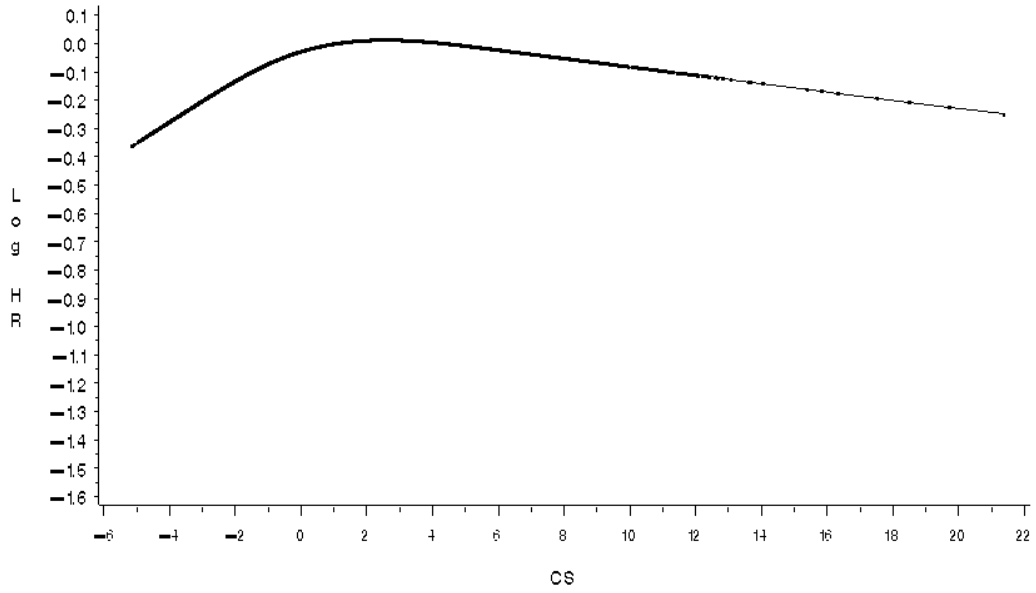
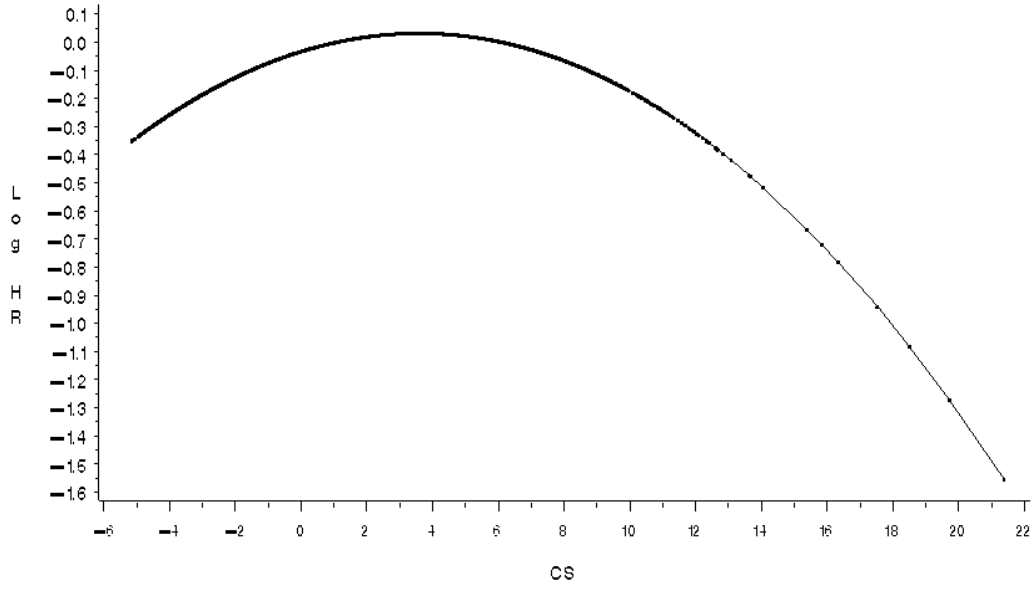


Figure 8.2b: Estimated Functional Form for Carstairs Score: Fractional Polynomial

(dots indicate actual data values)



The methods are generally different in the extremities where there is often a paucity of data. In this example the difference between transformations occurred in the right hand tails for outlying BMI and Carstairs scores (greater than 50 and10 respectively).

8.3. Multivariable Comparison of Non-Linear Functions

As in the advanced pancreatic cancer dataset, all variables were then considered in the multivariable setting using Cox proportional hazards regression based on a backward elimination selection method using a 5% significance level for exclusion of variables and including centre (stratification factor at randomisation) as a 'forced' variable in each model. A final check was made to ensure that no term in the model could be omitted and that none of the omitted terms could be included without significantly affecting the model. All variables omitted from the model based on backward elimination were considered and re-instated if they significantly improved the model. Two final Cox proportional hazards regression models (Table 8.2) were developed based on i) restricted cubic spline and ii) fractional polynomial transformations, which could be directly compared based on AIC statistics:

Table 8.2: Cox Proportional Hazards Regression Models (n=42802, deaths=5486)

a)

AIC=109223.4	Restricted Cubic Spline Transformations Transformation (knot positions)	log HR (β)	se(β)	χ^2_w	p-value	HR (95% CI)
Stratification factor	* CENTRE-1	0	-	527.4	<0.001	1.0
	CENTRE-2	-0.65	0.047			0.52 (0.48 to 0.57)
	CENTRE-3	0.20	0.044			1.22 (1.12 to 1.33)
	CENTRE-4	0.17	0.044			1.19 (1.09 to 1.30)
	CENTRE-5	0.23	0.051			1.26 (1.14 to 1.39)
Continuous factor	^ log(ES)	1.25	0.029	1804.5	<0.001	3.49 (3.29 to 3.70)
Continuous factor RCS3	# BMI (21.0, 27.1, 35.3)	-0.064	0.0066	97.1	<0.001	NA
	BMI1	0.069	0.0088			NA
Categorical factor	! SMOKING-0	0	-	58.0	<0.001	1.0
	SMOKING-1	0.22	0.034			1.25 (1.17 to 1.34)
	SMOKING-2	0.25	0.042			1.28 (1.18 to 1.39)
	SMOKING-NK	0.83	0.45			2.31 (0.96 to 5.56)
Categorical factor	DIABETES	0.26	0.036	53.2	<0.001	1.30 (1.21 to 1.39)
Categorical factor	$ CABG ONLY	0	-	43.8	<0.001	1.0
	CABG+OTHER	-0.02	0.095			0.98 (0.82 to 1.19)
	CABG+VALVE	0.23	0.041			1.26 (1.17 to 1.37)
	CABG+VALVE+OTHER	0.22	0.16			1.25 (0.91 to 1.72)
	OTHER	0.014	0.095			1.01 (0.84 to 1.22)
	VALVE ONLY	-0.038	0.039			0.96 (0.89 to 1.04)
	VALVE+OTHER	-0.027	0.12			0.97 (0.77 to 1.23)
Continuous factor RCS3	# CS (-3.5, -0.5, 5.7)	0.047	0.014	16.7	<0.001	NA
	CS1	-0.056	0.023			NA

Continued...

b)

Fractional Polynomial Transformations						
AIC=109223.2	**Transformation**	**log HR (β)**	**se(β)**	**χ^2_W**	**p-value**	**HR (95% CI)**
Stratification factor	* CENTRE-1	0	-	528.6	<0.001	1.0
	CENTRE-2	-0.65	0.047			0.52 (0.48 to 0.57)
	CENTRE-3	0.19	0.044			1.22 (1.12 to 1.32)
	CENTRE-4	0.17	0.044			1.19 (1.09 to 1.30)
	CENTRE-5	0.23	0.050			1.26 (1.14 to 1.39)
Continuous factor	^ log(ES)	1.25	0.029	1806.4	<0.001	3.49 (3.29 to 3.70)
Continuous factor FP2	# BMI^0.5	-8.31	1.078	100.0	<0.001	NA
	BMI^0.5*log(BMI)	1.53	0.20			NA
Categorical factor	! SMOKING-0	0	-	56.9	<0.001	1.0
	SMOKING-1	0.22	0.034			1.25 (1.17 to 1.33)
	SMOKING-2	0.25	0.042			1.28 (1.18 to 1.39)
	SMOKING-NK	0.84	0.45			2.32 (0.96 to 5.60)
Categorical factor	DIABETES	0.26	0.036	54.5	<0.001	1.30 (1.21 to 1.39)
Categorical factor	$ CABG ONLY	0	-	44.1	<0.001	1.0
	CABG+OTHER	-0.18	0.095			0.98 (0.82 to 1.18)
	CABG+VALVE	0.23	0.041			1.26 (1.17 to 1.37)
	CABG+VALVE+OTHER	0.22	0.16			1.25 (0.91 to 1.72)
	OTHER	-0.017	0.095			1.02 (0.84 to 1.23)
	VALVE ONLY	-0.038	0.038			0.96 (0.89 to 1.04)
	VALVE+OTHER	-0.024	0.12			0.98 (0.77 to 1.24)
Continuous factor FP2	# CS+5.2	0.062	0.017	19.3	<0.001	NA
	(CS+5.2)^2	-0.0034	0.0012			NA

HR=Hazard Ratio, RCS3=3 knot Restricted Cubic Spline, FP2=2nd degree Fractional Polynomial; NA=not appropriate; ^ Euroscore based on historical transformation

* Centre: 'classification' variable using 'Centre-1' as the reference level, Type III test presented with 4 degrees of freedom

! Smoking status: missing data included in analysis as a separate 'classification' variable using 'Non-smoker as the reference level, Type III Wald χ^2 test presented with 3 degrees of freedom

$ Cardiac procedures: 'classification' variable using 'CABG Only' as the reference level, Type III Wald χ^2 test presented with 6 degrees of freedom;

Type III Wald χ^2 test presented with 2 degrees of freedom

The models were identical in the variables selected as significantly prognostic. All variables were selected as highly significant in both models. The sign of each corresponding variable in the models were identical and the significance of each variable similar. Parameter estimates only differed significantly for the two transformed variables, BMI and CS. The AIC were almost identical for each model.

8.4. Bootstrap Validation Comparing Non-Linear Functions

Internal validation, to directly compare the fit of the restricted cubic spline and fractional polynomial strategies, was carried out by calculating the sampling distribution of the difference in the AIC statistics between the models in multiple bootstrap resamples. This was programmed in SAS (appendix) based on nonparametric bootstrap analyses, using an identical strategy as in the analysis of the advanced pancreatic cancer dataset. The bootstrap analysis fitted the final models, one based on the restricted cubic spline transformations and one based on the fractional polynomial transformations, to a series of 1000 bootstrapped resamples taken from and with the same size as the original data. Smaller values of the AIC indicate a better model but it is unclear whether AIC statistics from non-nested models will approximate the chi-square distribution. The difference in AIC statistics for the different models in multiple bootstrap resamples provides a distribution which can then be summarised appropriately with bootstrap percentile confidence intervals derived from the bootstrap distribution of the difference between the restricted cubic spline and fractional polynomial models.

The %BOOT macro was used to calculate 1000 resamples of the original dataset using identical univariate transformations, if any, as in the original dataset. Modelling was carried out in 1000 bootstrap resamples forcing in centre, BMI and CS components (restricted cubic spline or

fractional polynomial) and considering all other variables for inclusion based on a backward elimination variable selection method. Two models were created within each bootstrap resample, one based on restricted cubic spline transformations and one based on fractional polynomial transformations. Again, the AIC for each model was output from the screen to a temporary SAS dataset (rather than delivered to an output dataset due to different AIC values being output) and the difference (AIC[RCS]-AIC[FP]) between them calculated, a positive difference indicating a better fit by the fractional polynomial transformation model. This was repeated in each 1000 bootstrap resamples and the mean difference calculated with 95% confidence interval (Table 8.3):

Table 8.3: Bootstrap Analyses of final models based on 1000 Resamples

	Bootstrap Mean AIC	**Bootstrap Percentile 95% CI**
Restricted Cubic Spline	109273.2	106443.5 to 111903.8
Fractional Polynomial	109273.1	106443.2 to 111903.6
Difference (RCS-FP)	0.12	-8.7 to 9.2

The bootstrap mean AIC (Table 8.3) were almost identical across the two models with a bootstrap mean reduction of 0.12 with the fractional polynomial transformation (bootstrap percentile 95% CI: -8.7 to 9.2) compared to the restricted cubic spline.

The extreme sample size enabled further nonparametric bootstrap analyses to be carried out which enabled the univariate fractional polynomial transformation for each continuous covariate to be recalculated within each of 200 bootstrap resamples taken from the original dataset and comparing against a 5-knot restricted cubic spline. The %RCSPLINE and %MFP8 macros had to be adapted to enable them to run within the %BOOT macro in SAS. The macros automatically

delete temporary SAS datasets but had to be adapted to retain the temporary datasets to enable output of the -2log-likelihood from each model in each bootstrap resample. The %MFP8 macro calculates the -2log-likehood for each model from which the AIC can be calculated if the number of parameters in each model are known. The number of parameters for each continuous covariate can be output using the FFORM variable created by the %MFP8 macro. However, this is not updated by the number of other variables selected in a multivariable model. The full models could not be compared since the %MFP8 does not allow non-significant continuous covariates to be forced into the model. As such, an unadjusted univariate fractional polynomial transformation was calculated for each continuous covariate in each bootstrap resample and compared against a 5-knot restricted cubic spline transformation. The -2log-likelihood was output from the screen to a temporary SAS dataset (rather than delivered to an output dataset due to different AIC values being output), the AIC was then calculated for each model and the difference (AIC[RCS]-AIC[FP]) between them calculated, a positive difference indicating a better fit by the fractional polynomial transformation model. This was repeated in each 1000 bootstrap resamples and the mean difference calculated together with the range of differences seen (Table 8.4). The %MFP8 macro is an automated procedure which provides a fractional polynomial transformation for variables only if they have a significant independent effect. As such, the 'best' fitting fractional polynomial transformation may exclude the variable of interest and as such the proportion of samples where the variable of interest is excluded was also calculated. Also, the influence of the size of the bootstrap resamples was investigated by decreasing the bootstrap resample sizes from the size of the original dataset to 20000, 10000, 5000 and 1000 observations.

Table 8.4: Univariate Bootstrap Resampling with Varied Size Resamples

200 Bootstrap Resamples comparing 5-knot RCS^ and FP functions			
Variable	**Bootstrap size**	**Bootstrap Mean Difference in AIC * (Minimum to maximum)**	**% excluded (FP only) #**
BMI	**42802**	-16.32 (-45.09 to 10.09)	0%
	20 000	-5.44 (-24.14 to 11.83)	0%
	10 000	-1.12 (-18.78 to 16.05)	0%
	5000	0.32 (-16.64 to 16.19)	0%
	1000	0.89 (-16.64 to 10.53)	46%
CS	**42802**	-0.42 (-11.36 to 7.92)	0%
	20 000	-0.01 (-12.52 to 7.96)	5%
	10 000	0.31 (-8.60 to 7.01)	45%
	5000	1.63 (-10.91 to 8.68)	72%
	1000	3.80 (-7.73 to 12.20)	95%

BMI=body mass index, CS=Carstairs score

RCS=restricted cubic spline, FP=fractional polynomial

^ 5 fixed knot positions based on original full data set

* Difference based on AIC[RCS]-AIC[FP] within each bootstrap sample

automated FP macro excludes non-significant variables

The univariate bootstrap mean AIC was lower for BMI with the restricted cubic spline function with a bootstrap mean reduction of -16.32 (the bootstrap mean difference ranging: -45.09 to 10.09), based on a sample size of 42802 as in the original dataset and as in Table 8.1. Similarly, the bootstrap mean AIC were almost identical for CS across the two functions (bootstrap mean reduction of 0.42, bootstrap mean difference ranging: -11.3 to 7.92), based on a sample size of 42802 as in the original dataset and as seen in Table 8.1.

When investigating BMI in bootstrap resamples of reduced sizes, the size of the bootstrap mean difference seen in the full dataset was reduced in line with a reduction in samples size. In the smallest resample size of 1000, BMI was excluded in 46% of 200 fractional polynomial models yet still gave similar bootstrap mean AIC to the restricted cubic spline. BMI was not excluded in any of the fractional polynomial models when using a bootstrap resample size of 5000 or more. When investigating CS in bootstrap resamples of reduced sizes, the bootstrap mean difference was similar in sample sizes of 10,000 observations and above. With decreasing sample sizes of less than 10,000 observations, the bootstrap mean difference increases, being largest in the smallest resample size of 1000. CS was excluded from 95% of the models when using the smallest sample size of 1000 but was progressively selected being excluded in 72% of models of size 5000, 45% of models of size 10,000 and 5% of models of size 20,000.

8.5. Conclusions

The aim of the analysis was to investigate the reproducibility of the results and conclusions drawn from the comparison of non-linear methods in the advanced pancreatic cancer dataset, where the fractional polynomial transformation gave a better fit compared to the restricted cubic spline transformation. The main advantage of the cardiac surgery dataset is its extreme size, analysis being based on 42,802 patients. The cardiac surgery data has an alternative event and survival rate. Direct comparison of the non-linear methods was based on the same model building strategy employed when investigating the advanced pancreas dataset. The results presented for the cardiac surgery dataset differ slightly from the published paper (Pagano 2009) where the analysis plan was based on a frailty model adjusting by other trial factors such as surgeon and using a default 5-knot spline for all continuous data other than EuroScore. In the analysis presented here, BMI and Carstairs score were transformed using 3-knot restricted cubic splines

and second degree fractional polynomial transformations. The fitted functions generated by splines and fractional polynomials were similar resulting in comparable models. The AIC were almost identical when comparing the final multivariable models. Similarly, there was almost no difference in the bootstrap mean difference when comparing AIC based on 1000 bootstrap resamples. Larger differences were seen between the two methods when resample sizes were reduced due to the reduction in available data to detect small effects and also to detect non-linearity. When reducing the resample size of the bootstrap resample, consistency in the conclusions made from analysis of the full dataset were not seen until a resample size of 20,000 for both BMI and Carstairs score, based on both the observed bootstrap mean difference and the inclusion/ exclusion of covariates by the fractional polynomial function. In analysis of the Carstairs score, the bootstrap mean difference in AIC is small despite a high proportion of exclusions by the fractional polynomial function showing this automated method to provide a simpler model with similar model fit. In summary, the fitted functions generated by splines and fractional polynomials were similar resulting in comparable models. The fractional polynomial transformation is derived through a fully automated procedure whereas separate procedures need to be carried out to determine the knots required for the 'best' fitting restricted cubic spline transformation. The fractional polynomial method is selected as the transformation of preference due to i) the simplicity of the fractional polynomial procedure, ii) the comparability of the results seen in the cardiac surgery dataset and iii) the better fit of the fractional polynomial model in the advanced pancreatic surgery dataset. External validation of a derived prognostic model is required to determine if is transportable and will accurately predict outcome in new patients. The focus of Chapter 9 is to externally validate the prognostic model, based on the fractional polynomial non-linear transformation method, derived for patients with advanced pancreatic cancer.

CHAPTER 9: EXTERNAL VALIDATION

- **SUMMARY**
- Internal validation methods are valuable in the absence of an external dataset
- External validation is the most stringent validation process in a new group of patients
- The prognostic model is fitted to an external dataset to assess predictive value
- The prognostic model can be updated in light of new external data using an identical modelling strategy as that used in the development of the original model
- Model performance of the updated model should be assessed
- A prognostic index can be derived from the updated prognostic model
- A statistically validated prognostic index for advanced pancreatic cancer was derived
- Treatment effects within identified prognostic factors can be investigated

9.1. Introduction

There are two main validation methods, internal and external validation. Internal validation methods, described and presented earlier (Chapter 4), are very valuable methods in the absence of an external dataset but are of less value than validation on a completely independent dataset. An external dataset in advanced pancreatic cancer was expected to become available during the project and as such it was anticipated that the final derived prognostic model could undergo external validation. As such, data-splitting the original dataset was avoided thus maximising the available data in order to detect significant prognostic factors and using internal validation methods until the independent, external dataset became available.

External validation is the most stringent validation process and assessment of accuracy of a prognostic model (Harrell, 2001) based on evaluating the model in a new group of patients, despite being subject to differences in geographical location of patients and calendar differences.

It provides an important opportunity to review, strengthen and update a prognostic model with new data (Royston and Sauerbrei, 2004). The primary role of model validation is to ascertain if a derived prognostic model has external validity, that is that the model may be transportable and will accurately predict outcome in other patients not used in the model building process, and thus by inference in future patients. There are three major causes of failure of a model to validate: i) overfitting/ model optimism, ii) changes in the measurement of variables and iii) major changes in the inclusion criteria of patients (Harrell, 2001).

The aim of this chapter is to describe and present an external validation of the prognostic model derived in advanced pancreatic cancer. A validated prognostic index of clinical and laboratory criteria in advanced pancreatic cancer will enable stratification of patients into risk-sets to enable different treatment approaches to be targeted to different subgroups of patients. Routine availability and use of a validated prognostic index could ensure patients receive the most appropriate and targeted treatment modalities.

9.2. External Validation Dataset in Advanced Pancreatic Cancer

An improved, statistically validated prognostic index for advanced pancreatic cancer was calculated based on independent, external data from 533 patients recruited in a prospective randomised controlled trial in advanced pancreatic cancer (Cunningham *et al.*, 2009). The trial was designed to determine if there was any survival advantage in patients receiving gemcitabine in combination with capecitabine compared to current standard treatment of gemcitabine alone. The primary outcome measure was overall survival. Combination treatment was associated with a non-significant improvement in survival (Hazard Ratio=0.86; 95% confidence interval: 0.72 to 1.02; p=0.08) compared to gemcitabine alone, consistent across prognostic subgroups of cancer

stage and performance status (stratification factors at randomisation). The trial showed nominally significantly improved objective response rates (19.1% vs. 12.4%; p=0.034) and progression free survival (Hazard Ratio=0.78; 95% confidence interval: 0.66 to 0.93; p=0.004) in favour of combination treatment (pre-specified secondary outcome measures).

In an external validation, datasets should be investigated (Machin *et al.*, 2006) for differences in terms of:

1. Patient eligibility and disease characteristics
2. Differences in data items and measurements including endpoints and follow-up
3. Possible bias through missing data
4. Size of treatment effects and timing of treatment from diagnosis.

Both the development and validation datasets had similar patient eligibility criteria: histologically or cytologically unresectable pancreatic cancer, within 8 weeks of diagnosis or disease recurrence and adequate performance status. Previous therapy for metastatic or locally advanced disease was an exclusion criteria. Randomisation of patients in the validation dataset was stratified by cancer stage (locally advanced (stage III/IVA) or metastatic (stage IVB)) and performance status (0,1 vs. 2). The primary outcome measure in all trials was survival time calculated from the date of randomisation to date of death from any cause.

Sample size of the validation set should provide sufficient evidence to detect prognostic factors of specific magnitude. The number of observed events for validation of a model ($O_{validation}$) with k regression coefficients has been suggested as: $10 \times k_{validation} \leq O_{validation} \leq 20 \times k_{validation}$ (Machin *et al.*, 2006). A total of 529 patients in the validation dataset had died and as such indicates sufficient

evidence to investigate the 12 regression coefficients identified in the developed prognostic model (Table 6.2).

A total of 1186 patients were available for analysis across the two datasets (Table 9.1). Factors were generally balanced across the two datasets with the average age being 62 years (range 26 to 89), 681 (57%) male, 799 (67%) presenting with metastases and 450 (38%) with known lymph node involvement. All patients in the validation dataset had cancer stage III or IV compared to 84% in the development dataset. Similarly 67% of patients in the validation dataset had tumour stage III or IV compared with 44% in the development dataset. Five common biochemical variables were collected across both studies: bilirubin, alkaline phosphatase, albumin, LDH and CA19-9.

Overall, the majority of patients (1141, 96%) had died by the time of analysis with a median follow-up time of 19.9 months (Range: 0.9 to 44.0) for the 45 patients still alive. The median survival estimate for the group is 5.6 months (95%CI: 5.3 to 5.9) with 12-month survival estimate of 20.0% (95%CI: 17.7% to 22.3%).

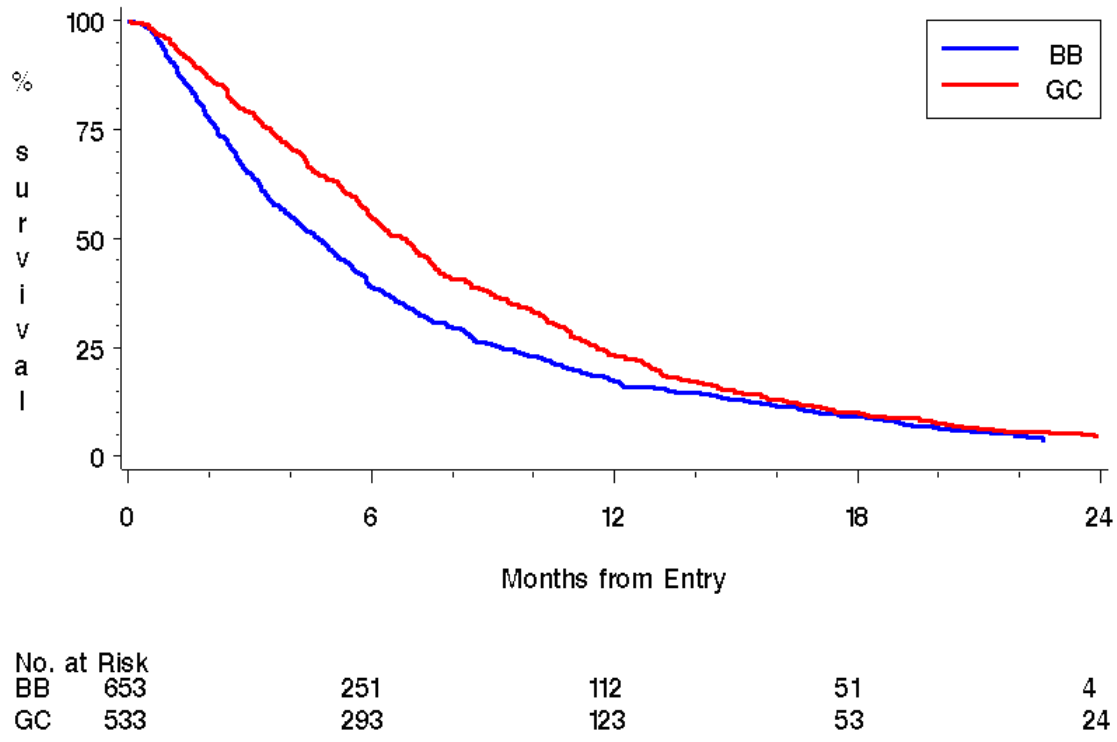
Table 9.1: Common Variables across Datasets

Variable		Development N=653 (55%)	Validation N=533 (45%)	TOTAL N=1186 (100%)
DEMOGRAPHICS				
* Age at entry (yrs)	Median	63	62	62
	Range	29 to 89	26 to 83	26 to 89
Ethnic race	White	590 (90%)	510 (97%)	1100 (93%)
	Other	62 (10%)	17 (3%)	79 (7%)
	Missing	1	6	7
Sex	Male	368 (56%)	313 (59%)	681 (57%)
	Female	285 (44%)	220 (41%)	505 (43%)
Treatment	Gem alone	222 (34%)	266 (50%)	488 (41%)
	Gem+	431 (66%)	267 (50%)	698 (59%)
TUMOUR INFORMATION				
Cancer stage	I	32 (5%)	0	32 (3%)
	II	72 (11%)	0	72 (6%)
	III	104 (16%)	156 (29%)	260 (22%)
	IV	439 (68%)	377 (71%)	816 (69%)
	Missing	6	0	6
Distant Metastases	M0	194 (30%)	155 (29%)	349 (30%)
	M1	436 (67%)	363 (68%)	799 (67%)
	Missing	23 (3%)	15 (3%)	38 (3%)
Regional lymph nodes	N0	243 (37%)	167 (31%)	410 (35%)
	N1	251 (39%)	199 (38%)	450 (38%)
	Missing	159 (24%)	167 (31%)	326 (27%)
Primary tumour T stage	T0	8 (1%)	0	8 (1%)
	T1	158 (24%)	13 (2%)	170 (14%)
	T2	144 (22%)	71 (14%)	215 (18%)
	T3	280 (43%)	91 (17%)	371 (31%)
	T4	7 (1%)	266 (50%)	273 (23%)
	Missing	56 (9%)	93 (17%)	149 (13%)
SERUM CHEMISTRY and HAEMATOLOGY				
Laboratory Variables		Median (Range), Missing	Median (Range), Missing	Median (Range), Missing
* AST (SGOT) IU/L		25 (6 to 538), 29	NA	-
* BILIRUBIN μmol/L		13.7 (3 to 277), 24	11 (2 to 35), 3	12 (2 to 277), 27
* ALK PHOSPHATASE IU/L		140 (35 to 2064), 24	173 (39 to 2188), 4	153 (35 to 2188), 28
* ALBUMIN g/L		38 (22 to 47), 25	38 (18 to 52), 3	38 (18 to 52), 28
* LDH IU/L		164 (29 to 1495), 32	331 (21 to 1616), 66	211 (21 to 1616), 98
* BUN mmol/L		9.3 (2.9 to 34.3), 33	NA	-
* CA19-9 KU/l		710 (5 to 1,000,000), 47	820.5 (0.6 to 14,000,000), 15	745.9 (0.6 to 14,000,000), 62
* HAEMOGLOBIN g/dL		12.4 (5.5 to 19.1), 41	NA	-
* WBC 10^9/L		7.9 (2.3 to 31.6), 41	NA	-
OUTCOME				
Event indicator	Alive	41 (6%)	4 (1%)	45 (4%)
	Dead	612 (94%)	529 (99%)	1141 (96%)
* Follow-up alive (mo)	Median	20.7	30.2	19.9
	Range	0.9 to 24.6	8.6 to 44.0	0.9 to 44.0

NA=not available, * = continuous measurements

Median survival was estimated as 4.7 months (95%CI: 4.2 to 5.1) with 12-month survival estimate of 17.4% (95%CI: 14.5% to 20.3%) in the development dataset compared with a median survival of 6.8 months (95%CI: 6.1 to 7.4) and 12-month survival estimate of 23.2% (95%CI: 19.6% to 26.8%) in the validation dataset (Figure 9.1):

Figure 9.1: Survival by Dataset



BB=British Biotech Development dataset; GC=GEMCAP Validation dataset

9.3. Model Fit and Discrimination

The fit of the developed prognostic model (Table 6.2), is investigated by forcing the regression coefficients from the derived prognostic model onto the validation dataset and assessing its predictive value by comparison of predicted and observed survival estimates and residuals.

Prognostic separation is the term used to explain statistical discrimination between prognostic scores when categorising patients into different prognostic risk-sets. Discrimination between subgroups of patients is based on the comparison of predicted and observed survival. Predicted prognostic information can be quantified in the original dataset used for model building using Predicted Separation (PSEP) methods (Altman and Royston, 2000) and it's equivalent calculated in the validation dataset termed Observed Separation (OSEP). PSEP is an index of separation of predicted prognostic information defined as:

PSEP = $p_{worst} - p_{best}$

where p_{worst} = predicted probability of dying for a patient in the group with the *worst* prognosis and p_{best} = predicted probability of dying for a patient in the group with the *best* prognosis. Close agreement of PSEP and OSEP values would indicate usefulness of the prognostic model. The PSEP method has been shown to have some drawbacks as a validation tool (Royston and Sauerbrei, 2004;Royston *et al.*, 2004) namely that the separation of a model may be quantified by the amount of variability in the linear predictor scores. An improved measure related to PSEP, is the D statistic which was suggested as an alternative measure of the discrimination and prognostic ability of a survival model. The generality of a prognostic model across k datasets can be evaluated by internal-external cross-validation on D by omitting each study in turn, estimating model parameters using remaining studies and evaluating D in the omitted study. This procedure has limited use when k=2 and as such discrimination of the prognostic model in advanced pancreas cancer is based on the PSEP method.

The proportional hazards written on the log-hazard scale is:

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

Linear predictor scores (LP) from the derived prognostic model (Table 6.2), including fractional polynomial transformations for non-linear covariates, were calculated as:

$LP = \eta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots.. + \beta_p X_{pi}$

LP = -0.37711*TRIAL + 0.17286*STAGEGP - 0.16115*SEX + 0.16934*TRTGP +

0.02383*CA199FP1(KU/l) - 0.00173*CA199FP2(KU/l) - 0.06802*ALB(g/L) - 0.00367*AST(IU/L) +

0.00119*ALKPHOS(IU/L) + 0.02772*BUN(mmol/L) + 0.04409*WBC(10^9/L) + 0.56220*LOGLDH(IU/L)

with the following code: TRIAL (1=BB128, 2=BB193), STAGEGP (1=I/II, 2=III/IV), SEX (1=MALE, 2=FEMALE), TRTGP (0=GEM, 1=GEMPLUS).

One of the three major causes of failure of a model to validate was 'changes in the measurement of variables'. The validation dataset did not include measurements for WBC, AST and BUN. Also the validation dataset is not influenced by the TRIAL variable. One option is to remove these variables from the derived prognostic model and re-estimate the regression coefficients for the remaining variables, available across both datasets. However, it is preferable to keep the regression coefficients adjusted by all known factors as estimated in the original prognostic model but remove the factors which are not available in the validation dataset. This is equivalent to assuming that omitted variables take the value zero which will directly affect the linear predictor score but will not bias groupings into risk sets which is based on available and validated factors. In this way the size of the effect of each identified prognostic factor as estimated in the original prognostic model is validated directly for variables available across both datasets.

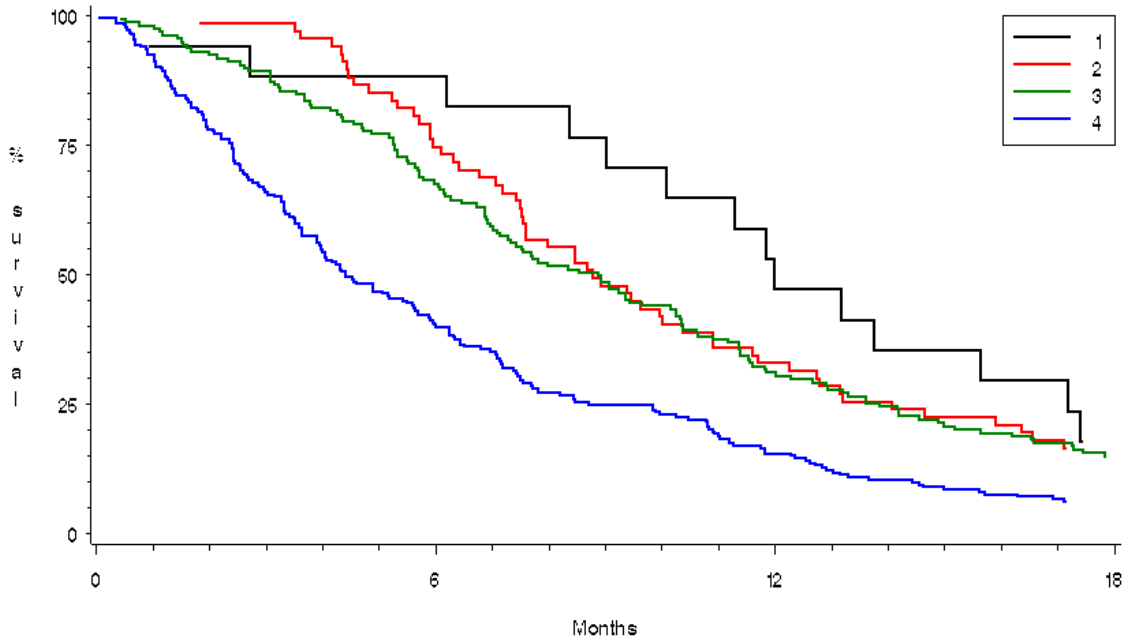
As such, individual scores were calculated for patients in the validation dataset according to linear predictor above but removing terms for WBC, AST, BUN and TRIAL:

LP = 0.17286*STAGEGP - 0.16115*SEX + 0.16934*TRTGP + 0.02383*CA199FP1(KU/l) - 0.00173*CA199FP2(KU/l) - 0.06802*ALB(g/L) + 0.00119*ALKPHOS(IU/L) + 0.56220*LOGLDH(IU/L)

with the following code: STAGEGP (1=I/II, 2=III/IV), SEX (1=MALE, 2=FEMALE), TRTGP (0=GEM, 1=GEMPLUS).

Model fit was assessed comparing differences in the direction of the parameter estimates and comparing observed and predicted survival estimates using PSEP and OSEP statistics. Patients were split into four groups based on the quartiles of the distribution of linear predictor scores used in the original model development (0.654140, 1.036423, 1.596797). A total of 455 of the 533 patients had complete data for the identified factors. Only 17 patients were allocated to prognostic group 1, 67 patients to group 2, 157 patients to group 3 and the remaining 214 patients to prognostic group 4. Kaplan-Meier survival estimates (Figure 9.2) show the four prognostic groups with descending median survival estimates of 12.0 (95%CI: 9.0 to 17.2), 8.8 (95%CI: 7.5 to 10.9), 8.9 (95%CI: 7.1 to 10.3) and 4.4 (95%CI: 3.9 to 5.7) months and descending 12-month survival estimates of 47.1% (95%CI: 23.4% to 70.8%), 32.8% (95%CI: 21.5% to 44.1%), 31.0% (95%CI: 23.7% to 38.3%) and 15.4% (95%CI: 10.6% to 20.2%):

Figure 9.2: Survival by Prognostic Group



The plot indicates some similarity in prognostic groups 2 and 3 from 9-months survival and beyond. Prognostic groups 2 and 3 have a similar survival distribution to prognostic group 1 in the development dataset (median survival estimate 8.5 (95%CI: 7.4 to 10.1) months, 12-month survival estimate 30.5% (95%CI: 22.8% to 38.2%)) re-iterating the better survival seen in patients in the validation dataset. The hazard ratios for groups 2, 3 and 4 using prognostic group 1 as the baseline were: 1.09 (95%CI: 0.64 to 1.86), 1.23 (95%CI: 0.75 to 2.04) and 2.07 (95%CI: 1.26 to 3.39) respectively.

PSEP was calculated in the development dataset based on 12-month survival rates of the four prognostic groups (Figure 6.4) as p_{best} = 1-0.305=0.695, p_{worst} = 1-0.043=0.957 and thus PSEP = 0.262. OSEP was calculated in the validation dataset based on 12-month survival rates of the four prognostic groups (Figure 9.2) as p_{best} = 1-0.471=0.529, p_{worst} = 1-0.154=0.846 and thus OSEP =

0.317. The observed separation is increased by an absolute difference of 5.5% and this is typical of some extent of overfitting in the original analysis. Since PSEP and OSEP were similar then the conclusion is that the prognostic information appears reproducible.

9.4. Updating the Prognostic Model

Once the prognostic model has been evaluated on an independent dataset, it may be updated in light of the new dataset. Updating the prognostic model was carried out by pooling the data from both datasets. Since some prognostic factors have already been identified from the original dataset then it is does not seem sensible to re-start the model building process, rather fit these identified factors to the pooled date to re-estimate the regression coefficients (Machin *et al.*, 2006). To account for any significant survival differences between the datasets and treatment, then additional terms for DATASET and TREATMENT were included to obtain adjusted estimates of the regression coefficients. Final models were developed and compared based on 1036 patients (993 deaths) with complete data on the prognostic factors of interest. The same modelling strategy used to develop the original prognostic model was used based on common variables across the two datasets (Table 9.1).

Univariate analysis of the six continuous variables identified that non-linear transformations were appropriate for three variables in their relationship with survival based on comparison of AIC: bilirubin μmol and alkaline phosphatase IU/L both as log transformations and CA19-9 KU/l as a second degree fractional polynomial transformation [CA19-9 $^{-0.5}$ + log(CA19-9)], updated from that in the original prognostic model. The three remaining continuous covariates were analysed assuming a linear relationship with log-hazard. Cancer stage was included in the development dataset as a binary variable coded on a clinical basis as ‘I/II’ vs. ‘III/IV’. In the validation dataset

only stage III and IV patients were recruited and it did not seem sensible to combine these. Investigation of the AIC of models containing only cancer stage showed a better model with reduced AIC when using four individual levels of stage ($AIC_{2levels}$=13948.2, $AIC_{4levels}$=13910.1). As such, cancer stage was considered as an ordered categorical variable with four levels and incorporated into the multivariable model using classification variables.

As in the development of the original prognostic model, multivariable modelling was carried out using Cox proportional hazards regression modelling with variable reduction based on a manual backward elimination method using a nominal significance level of α=0.05. Primarily models included prognostic variables already identified and common to both datasets (Cancer Stage, Sex, CA19-9, Albumin, Alkaline Phosphatase, LDH) using updated functional formats as specified. Validation group and treatment were also included in each model. Six additional variables common across both datasets were considered using classification variables for tumour stage, metastases and nodes (missing included as a classification level in both metastases and nodes). The updated multivariable model containing prognostic factors identified previously, but with updated regression coefficients:

Table 9.2: Updated Prognostic Model

	Variable	log HR (β)	se(β)	χ^2_W#	p-value	HR (95% CI)^
Stratification factor	DATASET	-0.49	0.078	39.8	<0.001	0.61 (0.53 to 0.71)
Stratification factor	TRT	0.054	0.066	0.67	0.41	1.06 (0.93 to 1.20)
Classification variable	* STAGE-2	-0.34	0.24	25.8	<0.001	0.71 (0.44 to 1.15)
	STAGE-3	-0.077	0.22			0.93 (0.61 to 1.41)
	STAGE-4	0.23	0.21			1.26 (0.84 to 1.88)
Binary variable	SEX	-0.16	0.065	6.2	0.013	0.85 (0.75 to 0.97)
Continuous factor Linear	ALB g/L	-0.058	0.0072	66.6	<0.001	0.75 (0.70 to 0.80)
Continuous factor FP2	# CA199^(-0.5)KU/l	0.77	0.37	48.0	<0.001	NA
	log(CA199)KU/l	0.11	0.018			NA
Continuous factor FP1	Log(ALKPHOS)IU/L	0.16	0.049	10.7	0.0010	1.17 (1.07 to 1.29)
Continuous factor Linear	LDH IU/L	0.00040	0.00021	3.8	0.053	1.04 (1.00 to 1.09)

HR=Hazard Ratio, FP1/2=1st or 2nd degree Fractional Polynomial, NA=not appropriate

* Stage: four level classification variable using level I as the reference level

Type III Wald χ^2 test

^ HR based on 5 unit increase in albumin g/L and 100 unit increase in LDH IU/L

The updated prognostic model, considering all common variables across the pooled datasets, selected identical factors as in the original model, except for those factors not collected in the validation dataset. No other variables common to both datasets were selected as additional prognostic variables. Differences between the two models included the format of the variables specifically; stage of disease which was expanded from a binary to a four-level ordinal variable, alkaline phosphatase IU/L modelled under a log transformation (Figure 9.3) as opposed to being linear previously, CA19-9 KU/l modelled with an alternative fractional polynomial transformation (Figure 9.4) and LDH IU/L modelled under the assumption of linearity, dropping the log transformation imposed previously:

Figure 9.3: Estimated Functional Form for Alkaline Phosphatase IU/L

(dots indicate actual values)

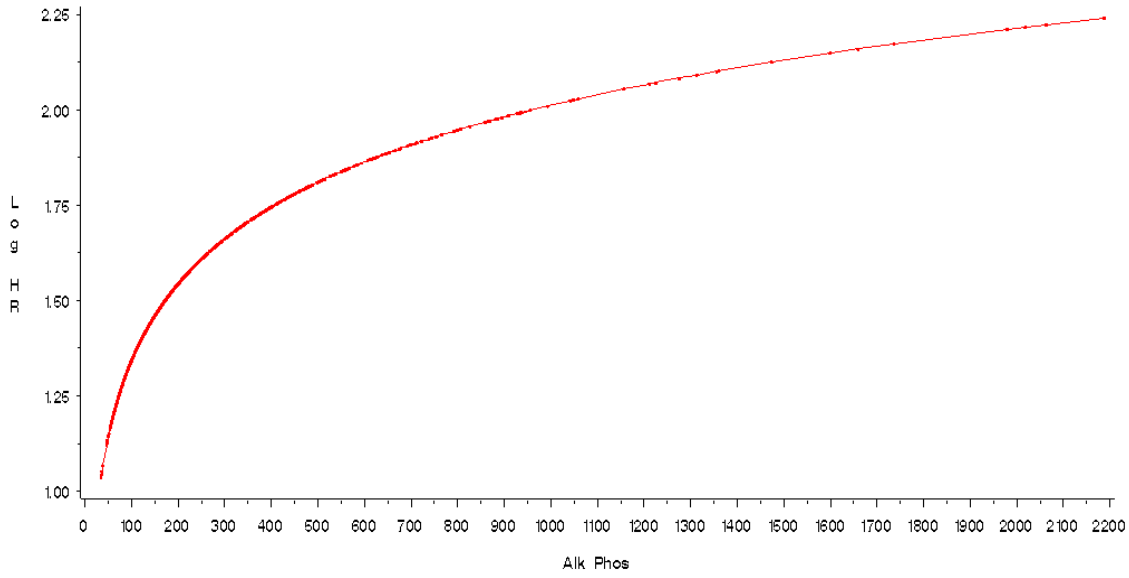
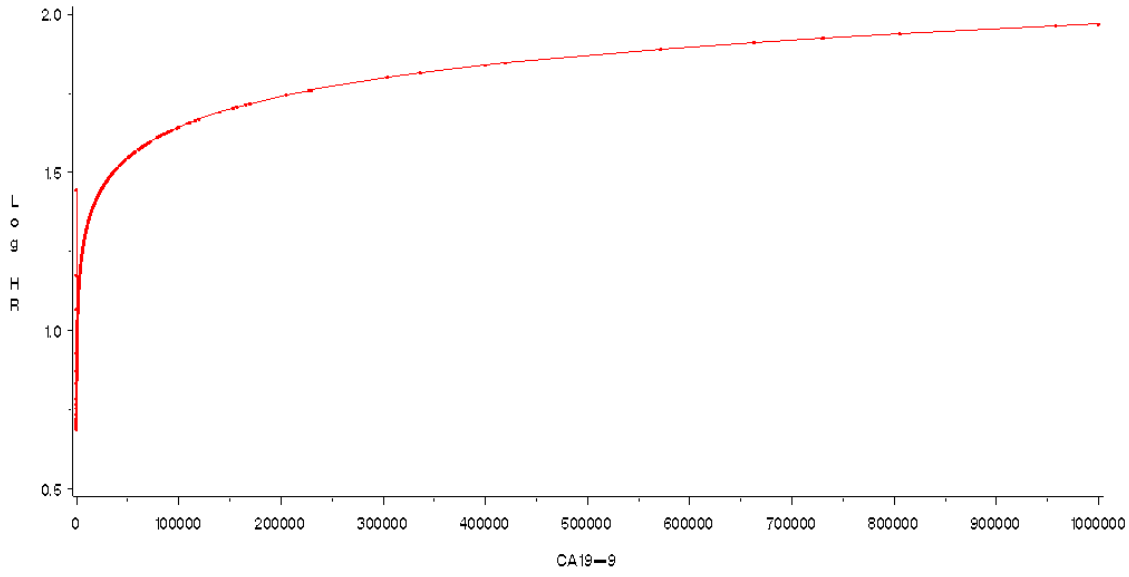


Figure 9.4: Estimated Functional Form for CA19-9 KU/l

(dots indicate actual values)

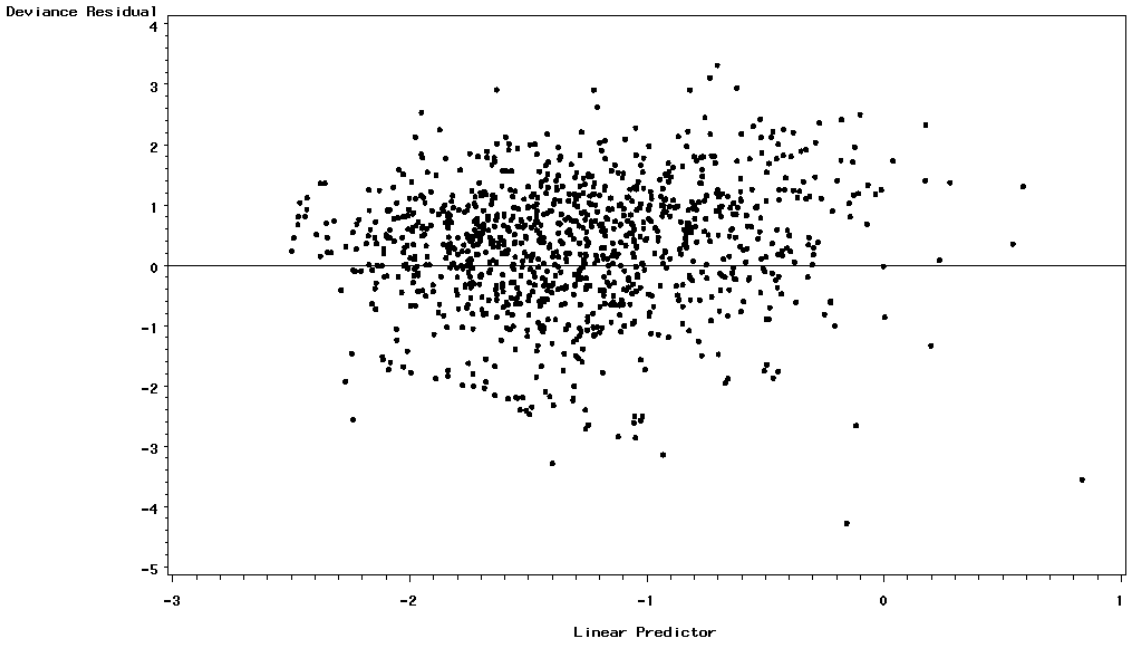


This updated model provides updated estimates of the regression coefficients of identified factors re-estimated in the pooled data and adjusted by dataset and treatment. Parameter estimates for those variables under an alternative format differed. Those variables not reformatted (SEX, TREATMENT, ALBUMIN) showed similar parameter estimates with smaller standard errors, as expected due to the increased number of events, although interpretation of the TREATMENT variable should be avoided due to varying treatments across the datasets and trials, its main purpose being to adjust other parameter estimates.

9.5. Model Performance

Deviance residuals for this updated model were plotted against the linear predictor and were randomly scattered and centred symmetrically around a residual value of zero ranging between -4.28 and 3.32 which suggests the data have not been mis-modelled (Figure 9.5):

Figure 9.5: Deviance Residual Plot



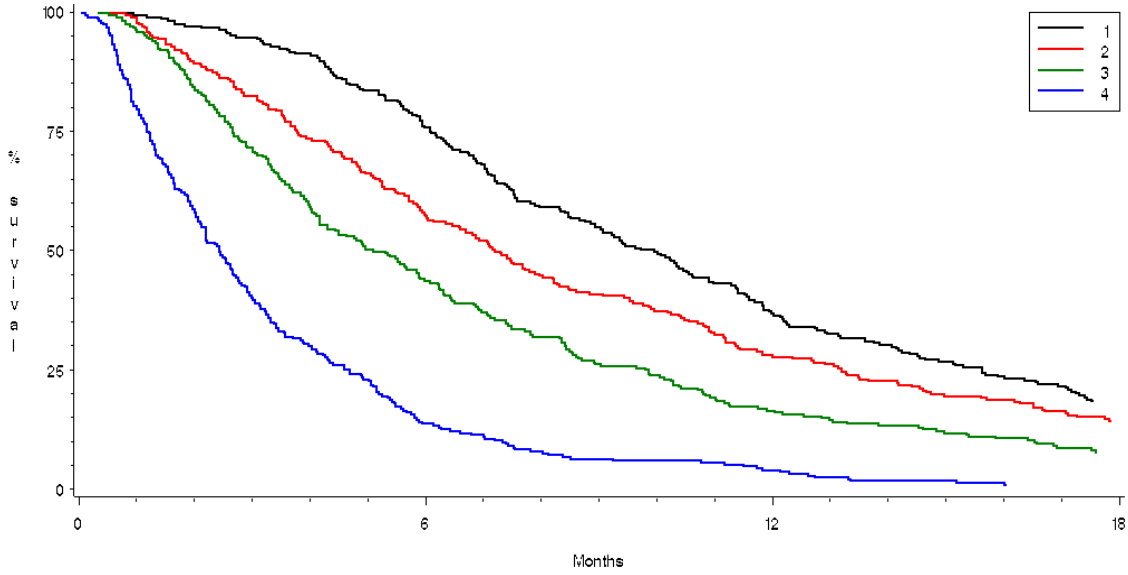
As previously, patients were split into four equal sized groups based on quartiles of the distribution of linear predictor scores from the updated prognostic model. Linear predictor scores (LP) were calculated as:

LP = -0.49006*DATASET + 0.05398*TRTGP - 0.33869*STAGE2 - 0.07656*STAGE3 + 0.22828*STAGE4 - 0.16126*SEX + 0.76745*CA199FP1(KU/l) + 0.10686*CA199FP2(KU/l) - 0.05836*ALB(g/L) + 0.15941*LOGALKPHOS(IU/L) + 0.0004020*LDH(IU/L)

substituting 1 for the classification variable for STAGE component as appropriate, else 0, and using the following code: DATASET (1=BB, 2=GC), TRTGP (0=GEM, 1=GEM+), SEX (1=MALE, 2=FEMALE).

Quartiles of the linear predictor scores were (-1.68213386, -1.32708021, -0.93425151) which created the four groups. Kaplan-Meier survival estimates (Figure 9.6) show four distinct prognostic groups with updated, descending median survival estimates, with non-overlapping confidence internals, of 9.7 (95%CI: 8.8 to 10.9), 7.2 (95%CI: 6.1 to 8.2), 5.1 (95%CI: 4.2 to 5.9) and 2.4 (95%CI: 2.1 to 2.8) months and updated, descending 12-month survival estimates of 36.7% (95%CI: 30.8% to 42.6%), 28.0% (95%CI: 22.5% to 33.4%), 16.3% (95%CI: 11.8% to 20.8%)and 3.9% (95%CI: 1.5% to 6.3%):

Figure 9.6: Survival by Prognostic Group



The updated hazard ratios for groups 2, 3 and 4 using prognostic group 1 as the baseline were: 1.26 (95%CI: 1.06 to 1.51), 1.79 (95%CI: 1.50 to 2.14) and 4.39 (95%CI: 3.65 to 5.27) respectively showing non-overlapping and tighter intervals in this updated model.

9.6. Defining the Prognostic Index

Since a prognostic index is to be used on future patients when treatment option remains to be decided and could be different to that in the model, then the DATASET and TREATMENT terms were removed from the updated prognostic model. This is equivalent to assuming that omitted variables take the value zero which will directly affect the linear predictor score but will not bias groupings into risk sets which is based on available and validated factors. Regression coefficients for remaining covariates were retained providing a new and updated prognostic index (PI) for use in prospective patients based on only those patient characteristics known at the time of diagnosis (Table 9.3):

Table 9.3: Prognostic Index

```
PI = -0.33869*STAGE2 - 0.07656*STAGE3 + 0.22828*STAGE4 - 0.16126*SEX +
0.76745*CA199FP1(KU/l) + 0.10686*CA199FP2(KU/l) - 0.05836*ALB(g/L) +
0.15941*LOGALKPHOS(IU/L) + 0.0004020*LDH(IU/L)
```

substituting 1 for the classification variable for STAGE component as appropriate, else 0, and using the following code: SEX (1=MALE, 2=FEMALE)

and classifying patients into prognostic risk sets with the associated estimates of survival and hazard ratios (Table 9.4):

Table 9.4: Survival Estimates and Estimated Risk of Death for Prognostic Groups

Risk Score	Prognostic Group	Estimate of Median Survival (95% CI)	Estimate of 12-month Survival (95% CI)	HR (95% CI)
Score < -0.995	1	9.5 (8.2 to 10.4)	36.0% (30.1% to 41.9%)	1.00
-0.995 ≤ Score < -0.661	2	7.1 (6.3 to 8.0)	26.4% (21.0% to 31.8%)	1.36 (1.13 to 1.62)
-0.661 ≤ Score < -0.293	3	4.9 (4.1 to 5.8)	16.3% (11.8% to 20.8%)	1.81 (1.51 to 2.17)
Score ≥ -0.293	4	2.7 (2.4 to 3.3)	6.2% (3.3% to 9.1%)	3.29 (2.74 to 3.95)

Two of the three major causes of failure of a model to validate (Harrell, 2001) have been addressed whilst deriving this prognostic index: i) the estimate of overfitting in model development was minimal (<1%) with minimal prognostic separation (5.5%) when validated and ii) the validation dataset was based on a similar patient population with similar inclusion criteria. The third cause of

failure was due to changes in the measurement of variables since the validation dataset did not record data for three variables (WBC, AST and BUN) selected for inclusion in the original model (Table 6.2). WBC was considered highly significant in the original model whereas the influence of AST and BUN were more borderline. These variables will require further evaluation of their prognostic importance. In response to the suggested post-validation queries (Altman and Royston, 2000):

1. Are the same variables still important? *Yes, but 3 were not evaluable at validation.*
2. Is the functional form of the prognostic model correct? *Yes, similar but updated.*
3. Are the estimated regression coefficients compatible? *Yes, where formats remain.*
4. How well does the model fit the new data? *Minimal overfitting identified.*
5. Is the ordering of the prognostic groups preserved? *Yes, and more distinct.*
6. Are the event rates between the prognostic groups significantly different? *No.*

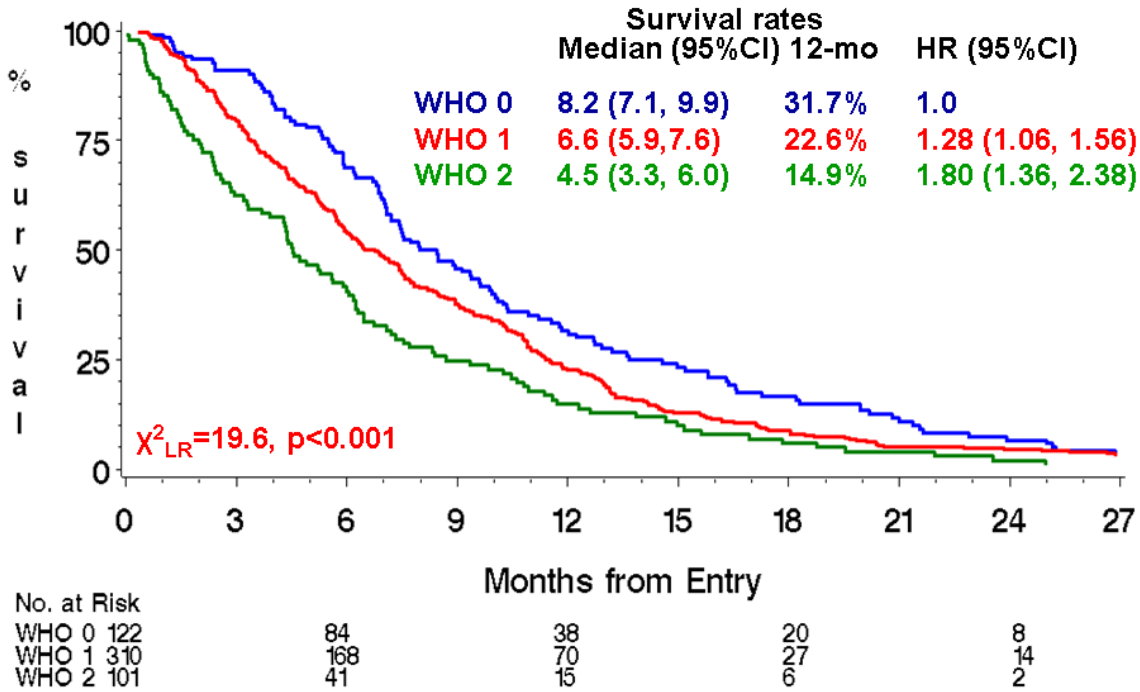
The updated, validated prognostic index in advanced pancreatic cancer is based on clinically relevant information routinely available at the time of diagnosis: cancer stage, sex, albumin, CA19-9, alkaline phosphatase and LDH. The index identified four distinct and clinically relevant risk-groups of patients which will enable routine stratification of patients and allow appropriate treatments to be targeted to groups of patients. It is intended that this prognostic calculator will be available to clinicians through the world wide web, attached to pancreas cancer specific web-sites.

9.7. Additional Information Collected In the Validation Dataset

The validation dataset trial collected additional factors to development dataset (diabetes, pancreatitis, tumour grade and site, ALT and creatinine), the most important being performance status. WHO performance status has been shown elsewhere to be an important prognostic factor

(Chapter 4). It was selected as a stratification factor at randomisation and was identified as a potentially important prognostic factor based on univariate analysis (Figure 9.7):

Figure 9.7: Survival by Performance Status



Additional exploratory analysis of the validation dataset was carried out independent of the development dataset but based on an identical modelling strategy. A Cox proportional hazards regression model was developed using 388 patients (384 deaths) with complete data on the clinical factors of interest and included stage of disease (p=0.0035) and performance status (p<0.001), stratification factors at randomisation, as well as randomised treatment group. The model identified five additional prognostic factors: albumin, CA19-9, size of primary tumour (all p<0.001), alkaline phosphatase (p=0.014) and LDH (p=0.034). Non-linear transformations were

appropriate for albumin g/L, CA19-9 KU/l, size of primary tumour and alkaline phosphatase IU/L indicating strong non-linear effects on log-hazard. A model including all 533 patients (529 deaths) based on multiple imputation methods to control for missing covariate data, selected grade of disease as an additional prognostic variable (p=0.034) which had been excluded from all models prior to imputation suggesting a link to other variables. A Cox model, excluding consideration of LDH and size of primary tumour due to missing data associated with these variables based on an 498 patients (494 deaths), included presence of a stent (p=0.0086) again suggesting a link to other variables, notably size of primary tumour. Clinical guidance has indicated concern over the accuracy of the measurement of the primary tumour size in patients with advanced disease with possibly multiple metastases, as such the clinical significance of size of primary tumour is limited. Clinical guidance has also highlighted presence of a stent to be directly related to other laboratory data, specifically bilirubin, alkaline phosphatase and AST for measuring jaundice in patients.

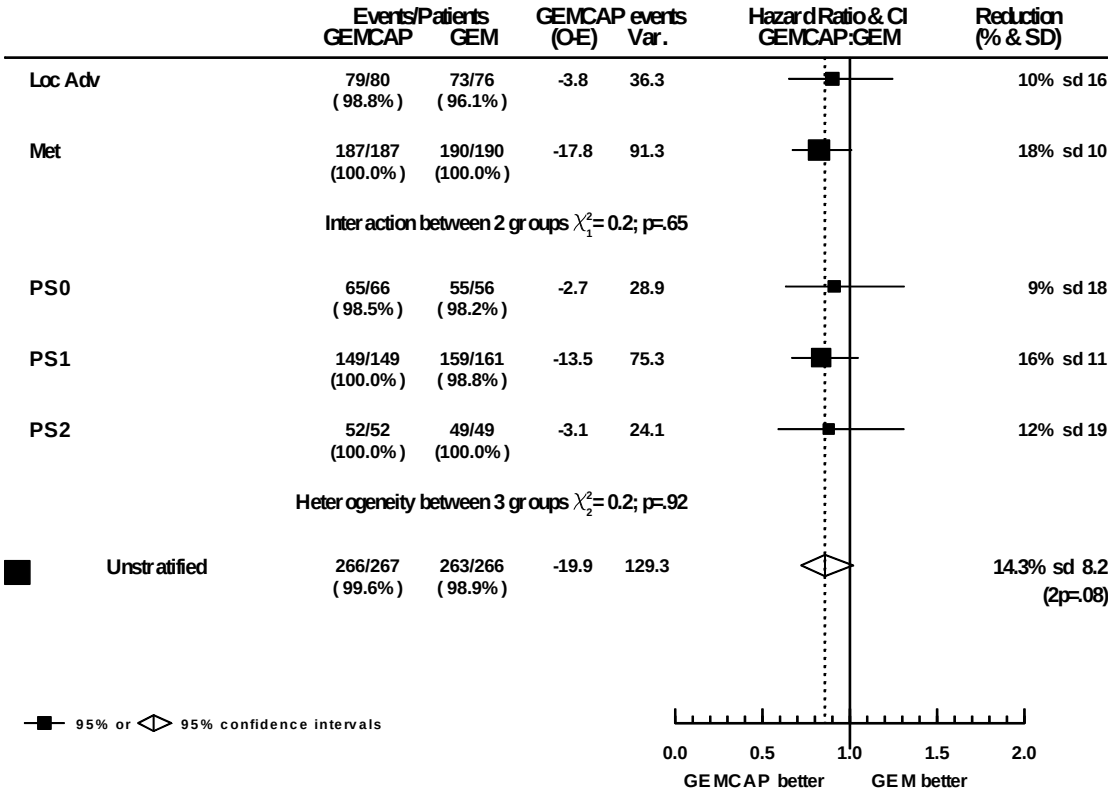
As such, the independent prognostic influence of primary tumour size and presence of a stent do not require further investigation. However, the independent prognostic influence of performance status and grade of disease and the relationships of these with other prognostic variables should be explored in further studies, highlighting the importance of collecting this data in future to allow the prognostic index to be updated with new relevant information when it becomes available.

9.8. Treatment Effects across Prognostic Subgroups

Interpretation of the treatment variable in the updated prognostic model was avoided due to the varying treatments across the datasets, its main purpose being to adjust other parameter estimates. Treatment effect across prognostic groups were investigated graphically, without significance testing, in the validation dataset using Forest plots (Early Breast Cancer Trialists'

Collaborative Group., 1990) to assess consistency of the hazard ratio of the treatment effect across categorical prognostic covariates. The treatment effect was consistent across prognostic subgroups of cancer stage and performance status (stratification factors at randomisation) (Cunningham *et al.*, 2009) (Figure 9.8, adapted):

Figure 9.8: Treatment Effect within Prognostic Subgroups



Subpopulation treatment effect pattern plots (STEPP) (Bonetti and Gelber, 2000) were used to assess graphically treatment-covariate interactions across identified continuous prognostic covariates. Plots were based on creating approximately 20 subgroups of 50 consecutive patients with increasing values of CA19-9 KU/l, with 30 overlapping patients between subgroups. The hazard ratio of the treatment effect was scattered around the value of one (no difference in treatments) with no apparent trends for increasing values of CA19-9 KU/l (Figure 9.9), albumin g/L

(Figure 9.10), and alkaline phosphatase IU/L (Figure 9.11) and with confidence intervals spanning unity:

Figure 9.9a: STEPP plot of Hazard Ratio of Treatment Effect for increasing CA19-9 to 10,000 KU/l

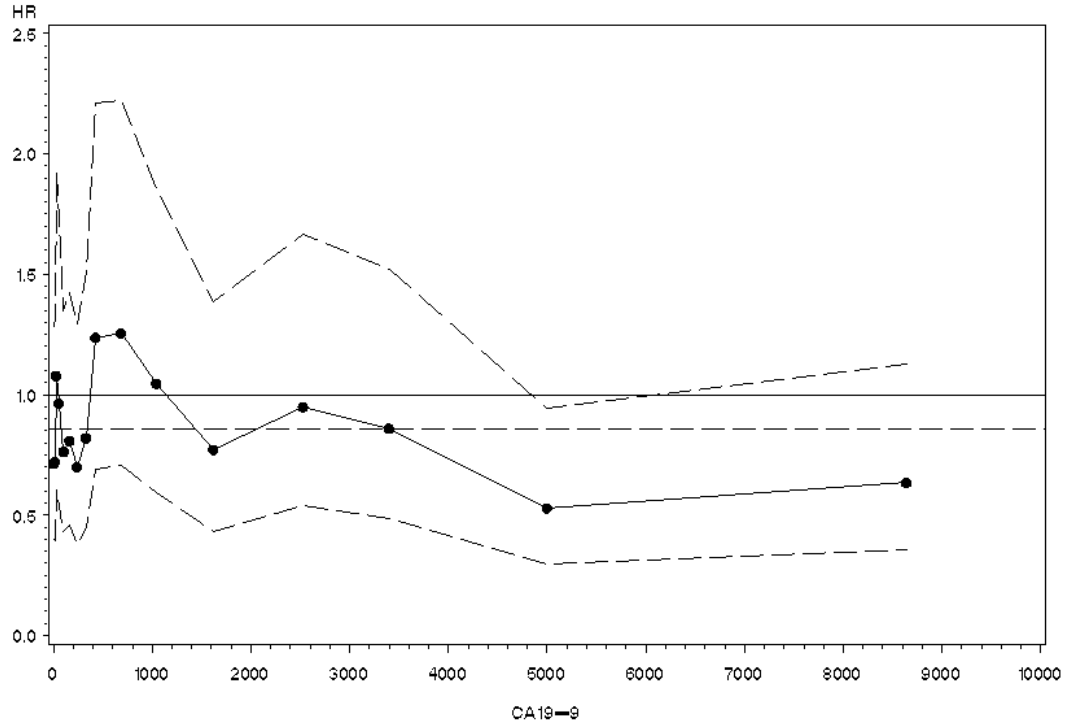


Figure 9.9b: STEPP plot of Hazard Ratio of Treatment Effect for increasing CA19-9 to 1,000 KU/l

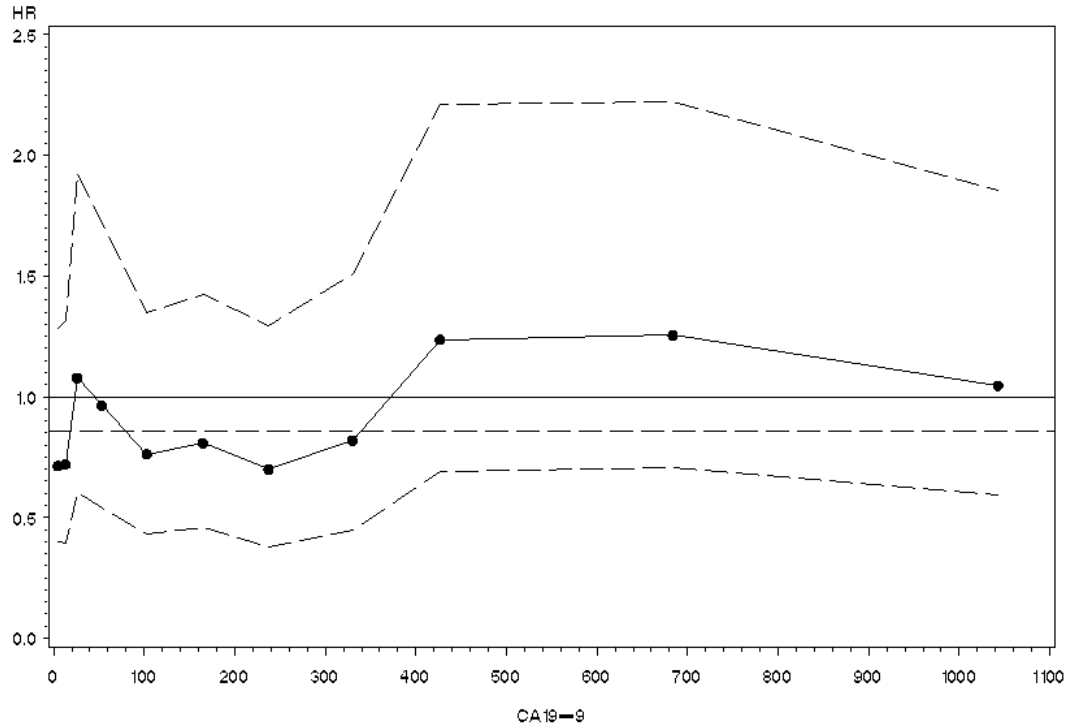


Figure 9.10: STEPP plot of Hazard Ratio of Treatment Effect for increasing Albumin g/L

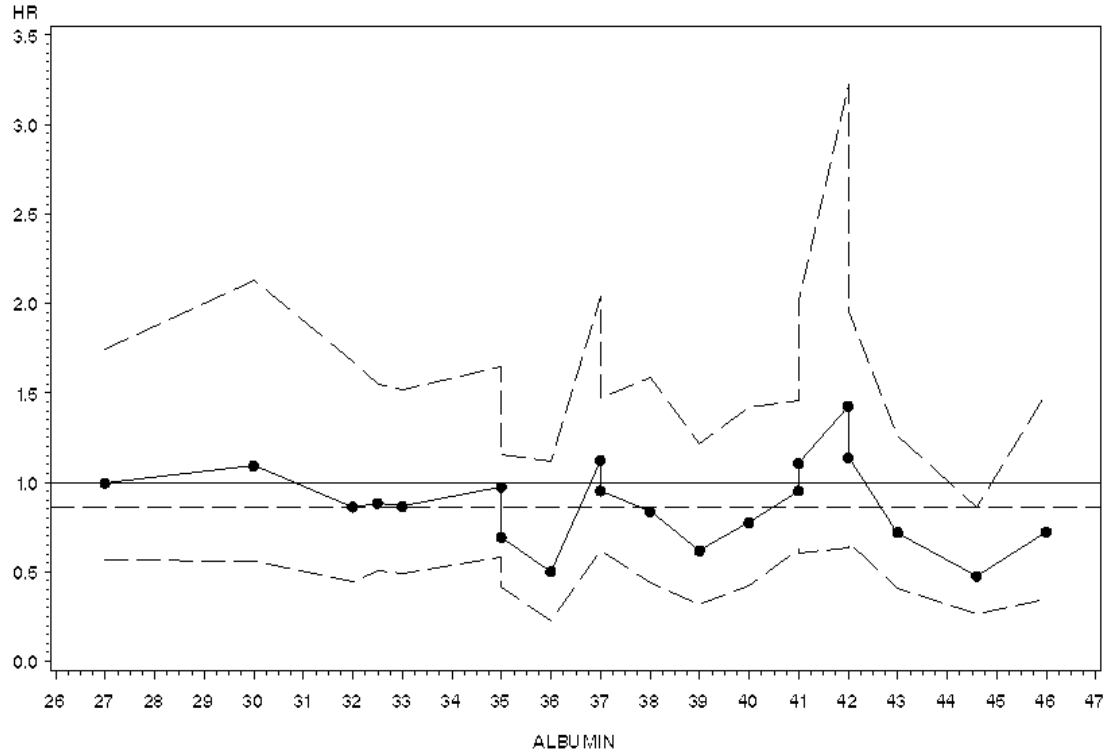
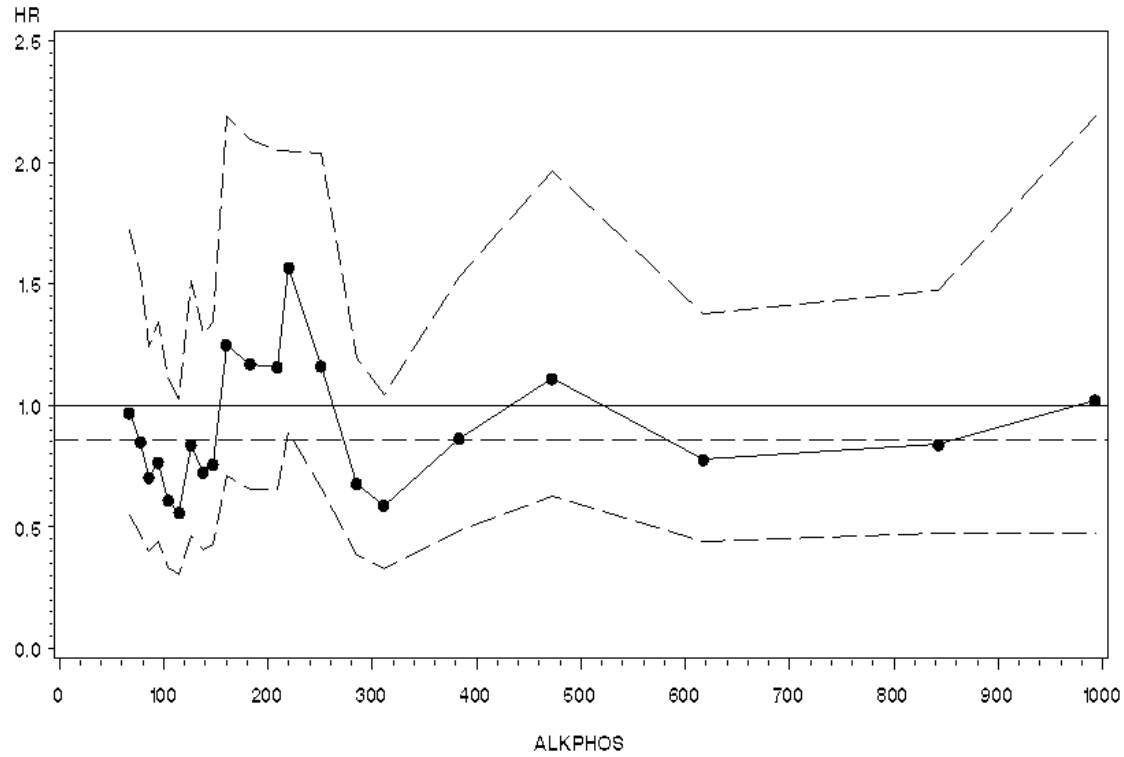


Figure 9.11: STEPP plot of Hazard Ratio of Treatment Effect for increasing Alkaline Phosphatase IU/L



9.9. Conclusions

None of the previously published prognostic factors studies in advanced pancreatic cancer had been validated. The prognostic model developed in Chapter 6 has now been internally and externally validated. External validation is the most stringent form of validation and allows a prognostic model to be reviewed and updated with new data. The model was externally validated using new data from 455 advanced pancreatic cancer patients. Prognostic separation techniques showed the prognostic information in the original prognostic model to be reproducible in new patients. The advantage of external validation is that the prognostic model can be updated based on the new data. An updated model revealed that no other variables common in both datasets were selected for inclusion in the model. The functional format and regression coefficients were updated based on the pooled data producing more accurate estimates. Four distinct clinically relevant prognostic risk groups were identified with updated, non-overlapping and tighter confidence intervals for the survival and hazard ratio estimates. A validated prognostic index was derived from the updated prognostic model based on only clinically relevant information routinely available at the time of diagnosis. This prognostic index is the first to be based on robust and appropriate statistical methodology and is the first to have been statistically validated. The validation process has also identified possible prognostic factors which require further investigation and validation and as such should are recommended as required data collection in future studies.

CHAPTER 10: DISCUSSION AND RECOMMENDATIONS

10.1. Summary

This thesis provides the first definitive, validated, prognostic tool developed using a large international representative sample of patients with advanced ductal adenocarcinoma of the pancreas based on appropriate and robust statistical methodology and intended to be accessible to clinicians through convenient media, including the world-wide-web. This research has enabled recommendations for the design and analysis of future randomised controlled trials, specifically identification of possible important stratification factors and the appropriateness of different statistical modelling approaches including dealing with different data types. Cancer Research UK supported the development of this prognostic tool through a Population and Behavioural Sciences project grant (reference: C19491/A6150, Lead investigator: DD Stocken).

10.2. Clinical Relevance of Research

The clinical focus of this research is pancreatic cancer which is an important disease to research as it is a common cause of cancer death and is difficult to diagnose and treat. Prognosis is important to help clinicians in their decision making when identifying patients who may be suitable for further treatment and are also relevant for the appropriate design and analysis of future clinical trials.

Pancreatic cancer is the eighth most common cause of cancer death in the world and fifth in the western world (Jemal *et al.*, 2003;Parkin *et al.*, 2005). It is particularly difficult to treat because of its remote location, late presentation and resistance to conventional chemotherapy. Prognosis is primarily based on resectability with improved survival for the 10 to 15% of patients suitable for

'curative resection'. As such, long-term survival remains poor with a reported 5-year survival of <5% (Bramhall *et al.*, 1995;Carr *et al.*, 1999;Jemal *et al.*, 2003) increasing to 25% in resected patients (Bramhall *et al.*, 1995;Carr *et al.*, 1999;Neoptolemos *et al.*, 2009b;Stocken *et al.*, 2005;Yeo, 1998). Survival benefits are anticipated by identifying risk groups of patients in whom treatment strategies can be targeted. The overall aim of this research was to identify the factors at the time of diagnosis which predict the risk group to which a patient best belongs, based on robust and appropriate analyses. There is currently no prognostic tool in routine use for prediction of survival from advanced pancreatic cancer. Such a tool would help clinicians identify subgroups of patients and help in their decision making regarding appropriate treatment strategies.

Randomised controlled trials researching treatments for pancreatic cancer have been limited by the aggressive nature of this disease with it's resistance to therapy and it's associated short survival. To provide sufficient evidence to confirm a small expected survival advantage with treatment requires large studies with random allocation of patients to treatment groups to minimise bias. Sample sizes in pancreas cancer trials have tended to be small and as such raises the problem of comparing across studies due to differences in the samples of patients. Generally patients are not clinically separated into prognostic groups for consideration of treatment except surgically (Wagner *et al.*, 2004). Cancer Research UK funded the first large (n=289 patients) randomised controlled trial of adjuvant treatment in patients undergoing curative resection (Neoptolemos *et al.*, 2004). This trial had a direct impact on patient care changing the standard of care from surgery alone to surgery plus chemotherapy and has led the way for further follow-on adjuvant studies (Neoptolemos *et al.*, 2009a). Gemcitabine became the standard of care for patients with non-resectable (advanced) disease based on a smaller study (n=126 patients) using clinical benefit response (a composite measure of pain, performance status and weight) as the

primary outcome (Burris *et al.*, 1997) but with the recommendation (NICE, 2010) to gather further evidence for the role of gemcitabine. A recently reported large Cancer Research UK randomised controlled trial (n=533 patients) in the advanced disease setting concluded a benefit for combination gemcitabine plus capecitabine compared against gemcitabine alone (Cunningham *et al.*, 2009). Historically, randomised controlled trials have stratified by country and resection margin status (adjuvant trials) and centre, stage of disease and performance status (advanced trials). Identification of important prognostic factors will not only enable identification of risk groups for the appropriate targeting of treatments, but also enable accurate stratification of patients into future clinical trials.

A review of prognostic factors conducted in 2004 (Stocken *et al.*, 2008) identified 36 prognostic factor studies reporting a total of 34 possible prognostic factors for advanced pancreatic cancer patients including demographic, clinical (including performance status, weight loss, treatment), surgical (including palliative procedures, site and stage of disease) and laboratory (including CA19-9, LDH, alkaline phosphatase and albumin). The majority of studies were questionable in terms of sample size and statistical methods and investigated different possible factors in differing formats across studies. Six previously reported prognostic indices included one large study of 1020 patients but with only three common variables to investigate. The remaining five studies were based on $\leq$166 patients, four were single centre studies, three were retrospective, all excluded consideration of centre effect and treatment effect and dichotomised continuous covariates when constructing the prognostic index. Statistical models derived on independent datasets ideally need validating on external data to ensure external validity and generalisability across new datasets. None of the six studies included any validation. There is currently no statistically robust, validated prognostic tool in routine clinical use which could help clinicians

predict survival and target treatments to advanced pancreatic cancer patients more appropriately and have an impact on future trial design and analysis.

10.3. Methodology and Findings

This research was based on robust data from three large, phase III randomised controlled trials representative of patients with advanced pancreatic cancer with a high event rate (96%), long follow-up (average 20 months) and an overall one-year survival rate of 17% (Bramhall *et al.*, 2002;Bramhall *et al.*, 2001;Cunningham *et al.*, 2009) similar to the 13% one-year survival rate reported for England and Wales (Cancer Research UK, 2010). Two datasets were combined for model building (n=653) and one was used as an external validation dataset (n=533). The high number of events enabled stable estimation of the prognostic ability of eighteen possible clinical, histological, biochemical and demographic prognostic factors.

Valid statistical analyses are essential to make best use of data and optimise clinical application. Cox proportional hazards regression model (Cox, 1972) investigates the relationship between survival and one or more possible factors detecting the extent to which potential factors affect the hazard function. A multivariable approach, based on Cox proportional hazards modelling using backward elimination variable selection, was used to investigate multiple possible prognostic factors with the ability to include categorical and continuous covariates. Continuous variables are often simplified by assuming a linear relationship between predictor and log-hazard i.e. the log risk increases or decreases linearly as the value of the factor increases, which may not be appropriate and could be misleading. Dichotomisation of continuous data is common but problematic. Since the variability in outcome within groups is ignored by categorisation then the variability between groups may be significantly underestimated as patients close to the cut point are analysed as

being very different rather than being very similar, resulting in a serious reduction of statistical power to detect relationships between predictors and outcome, residual confounding and serious bias (Altman and Royston, 2006;Royston *et al.*, 2006a). It is important to account for non-linear relationships with outcome in the prognostic factor setting since it has been shown that a treatment effect can be underestimated unless a strongly prognostic non-linear covariate was modelled correctly using a suitable non-linear function (Royston *et al.*, 2006a). The functional forms of the relationships between continuous covariates and log-hazard were not reported in previous studies in advanced pancreas cancer and were usually dichotomised. Non-linear relationships were addressed for the first time in this disease site and were investigated for each of the ten continuous variables in the advanced pancreatic cancer dataset used for model building.

Fractional polynomial (Royston and Altman, 1994) and restricted cubic spline (Stone and Koo, 1985) functions are both polynomial functions particularly suitable for modelling smooth curved relationships between response and a predictor. Fractional polynomial functions are an extended family of curves defining functions with non-integer and negative power terms, first-degree transformations providing monotonic curves and second-degree transformations having a maximum of one turning point. Restricted cubic spline functions are piecewise polynomial functions constrained to join and be smooth at the endpoints of each interval (knot position) and restricted to be linear in the tails (before first and after last knots). Both restricted cubic splines and fractional polynomial transformations for non-linear continuous covariates are easy to implement within standard generalised modelling methods providing the number of knots in restricted cubic splines is small, else parameter estimation must be penalised to stabilise functions. There are convenient fractional polynomial (Meier-Hirmer *et al.*, 2003) and restricted cubic spline macros

(Harrell, 1991) available in SAS (SAS Institute Inc., 1999) and R (The R Foundation for Statistical Computing, 2008). Regression using fractional polynomials have shown that important additional prognostic information could be extracted where less sophisticated approaches miss (Sauerbrei *et al.*, 1999;Royston *et al.*, 2006b). Spline functions with multiple knot positions in smaller samples using many degrees of freedom can generate uninterpretable 'wiggles' and instability and can become interpretable (Sauerbrei *et al.*, 2007). One study comparing fractional polynomial and 4-knot restricted cubic spline models (Hollander and Schumacher, 2006) concluded that the fractional polynomial was the preferred approach in achieving a simple model and secondly, if a variable was linear in its relationship with log-hazard, the fractional polynomial method identified as linear more often. The fractional polynomial macro has the advantage that it can simultaneously select appropriate transformations within the multivariable model building process. No spline-based procedure for simultaneously selecting variables and functional forms has found wide acceptance (Sauerbrei *et al.*, 2007). Fractional polynomials also have the added potential ability to be able to extrapolate beyond the limits of the data and the ability to fit functions with longitudinal and time dependent non-linear covariates (Lambert *et al.*, 2005;Lehr and Schemper, 2007) unlike spline transformations.

The analysis of continuous variables in the advanced pancreatic cancer dataset has shown that non-linear transformations were appropriate for some covariates. Models containing non-linear transformations in the advanced pancreatic cancer dataset gave a substantially better fit compared to the models which dichotomised or assumed linearity of continuous covariates and included significant prognostic factors which were excluded in models under the assumptions of either step-functional (excluded alkaline phosphatase IU/L) or linear (excluded CA19-9 KU/l) relationships. The fractional polynomial and restricted cubic spline models confirmed previously

reported prognostic factors: albumin, CA19-9, alkaline phosphatase and LDH. Both models also identified WBC as a prognostic factor which had not been previously reported at that time, but has been reported subsequently (Siddiqui *et al.*, 2007). The fractional polynomial model also identified two additional possible prognostic factors with borderline significance, not previously reported: AST and BUN but these were not selected as prognostic in the restricted cubic spline model, requiring confirmation in further studies. Most importantly, the effect of CA19-9 was not apparent in the model under the log-linear assumption, the effect of alkaline phosphatase was not apparent in the model which dichotomised variables indicating how the significant effect of these variables may go unrecognised due to simplistic assumptions made in statistical modelling.

The fitted functions generated by splines and fractional polynomials were similar, resulting in comparable models but the AIC was smallest for the fractional polynomial model indicating a better fit to the data. Different models are directly compared using Akaike's Information Criteria (Akaike, 1974). Smaller values of the AIC indicate a better model but it is unclear whether AIC statistics from non-nested models will approximate the chi-square distribution. The AIC can be seen as a method of ranking models and as such the method forces a 'best' model which may be more complex and not as parsimonious as the 'next best' model but with almost identical AIC statistics. To avoid this, comparison of the restricted cubic spline and fractional polynomial strategies was carried out by calculating the sampling distribution of the difference in the AIC statistics between the models in multiple bootstrap resamples of the data. This was executed using an in-house developed SAS program. The distribution could then be summarised appropriately avoiding the need for distributional assumptions. The AIC statistics had to be output from the screen rather than the preferred option of delivering to an output dataset, as recommended by SAS (SAS Institute Inc., 1999), after reporting that the AIC values delivered to

screen and the output dataset differed, interestingly. The bootstrap mean AIC was smallest for the model containing fractional polynomial transformations and on average was a better fit than the model based on restricted cubic spline transformations by a bootstrap mean reduction of 11.0 (bootstrap percentile 95% CI: 9.0 to 31.4) in 1000 bootstrap resamples. The methods were generally different in the extremities where there is often a paucity of data and it could be argued that the restriction imposed in the tails in restricted cubic splines leads to greater stability. One possible solution could be to use the mid 90% of the range of 'noisy' continuous data with outlying extremes.

None of the six reported prognostic indices investigated centre effect. A grouped frailty model is a survival model which incorporates a random effect (RE) element into hazard function to account for heterogeneity between patients. An investigation of any centre effect within in this dataset using frailty modelling revealed no evidence of systematic differences in outcome across centres possibly due to the overall poor prognosis of patients with this disease and absence of effective treatments. Multiple imputation methods (Rubin, 1987) were used as a supportive analysis to the principal complete case analysis, to assess the impact of missing data which was minimal in the advanced pancreatic cancer dataset but again had not been addressed previously in the literature.

Model validation was addressed for the first time in this disease site and is the assessment of the accuracy of statistical estimates and of the stability of a set of prognostic factors. Internal validation was carried out to assess the degree of model optimism (over-fitting) which is a measure of poor external validity leading to inflated estimates of model fit and is a potentially important source of bias in prognostic models. Internal validation methods based on statistical resampling simulation techniques (Efron and Tibshirani, 1993) suggested minimal optimism and

since the degree of overfitting estimated for the model is minimal, rescaling the model estimates appears neither helpful nor necessary. External validation is the most stringent form of validation and allows a prognostic model to be reviewed and updated with new data. The regression coefficients from the derived prognostic model were fitted to external data from 455 advanced pancreatic cancer patients and the predictive value of the model assessed by comparison of predicted and observed survival estimates. Prognostic separation techniques (Altman and Royston, 2000) showed the prognostic information in the original prognostic model to be reproducible in new patients. The advantage of external validation is that the prognostic model can be updated based on the new data. An updated model revealed no other variables common in both datasets, were selected for inclusion in the model. The functional format and regression coefficients were updated based on the pooled data producing more accurate estimates. Four distinct clinically relevant prognostic risk groups were identified with non-overlapping and tighter confidence intervals for the survival and hazard ratio estimates. This is the first prognostic model developed in advanced pancreatic cancer to have undergone both internal and external validation.

Prognostic indices are derived from the validated regression model and prognostic scores grouped to form risk sets. A statistically validated prognostic index was derived from the updated prognostic model based on only clinically relevant information routinely available at the time of diagnosis. Four distinct clinically relevant risk-sets of patients were identified which could help clinicians to target treatments to prognostic groups of patients more appropriately. This prognostic index is the first to be based on robust and appropriate statistical methodology and is the first to have been statistically validated. The prognostic index should have the ability to aid decision making by identifying patients with better prognosis, possibly for inclusion in clinical trials of more toxic therapies, or identifying those with worse prognosis, possibly more suitable for more

palliative treatments off-study, especially since a greater number of palliative and toxic combinations of treatments are becoming available and being trialled in this disease. The validation process has also identified possible prognostic factors which require further investigation and validation and as such should are recommended as required data collection in future studies. It is intended that a prognostic calculator of the final validated index will be available through a pancreatic/ trials unit specific web-pages, based on an access front end for ease of use, being directly available to the clinical community and with the ability to be updated with future data from randomised controlled trials.

10.4. Derived Prognostic Model

Linear Predictor = -0.37711*TRIAL + 0.17286*STAGEGP - 0.16115*SEX + 0.16934*TRTGP + 0.02383*CA199FP1(KU/l) - 0.00173*CA199FP2(KU/l) - 0.06802*ALB(g/L) - 0.00367*AST(IU/L) + 0.00119*ALKPHOS(IU/L) + 0.02772*BUN(mmol/L) + 0.04409*WBC(10^9/L) + 0.56220*LOGLDH(IU/L)

substituting the following code: TRIAL (1=BB128, 2=BB193), STAGEGP (1=I/II, 2=III/IV), SEX (1=MALE, 2=FEMALE), TRTGP (0=GEM, 1=GEMPLUS).

10.5. Validated Prognostic Index

Prognostic Index = -0.33869*STAGE2 - 0.07656*STAGE3 + 0.22828*STAGE4 - 0.16126*SEX +

0.76745*CA199FP1(KU/l) + 0.10686*CA199FP2(KU/l) - 0.05836*ALB(g/L) +

0.15941*LOGALKPHOS(IU/L) + 0.0004020*LDH(IU/L)

substituting 1 for the classification variable for STAGE component as appropriate, else 0, and using the following code: SEX (1=MALE, 2=FEMALE) and classifying patients into prognostic risk sets with the associated estimates of survival and hazard ratios (Table 10.1):

Table 10.1: Survival Estimates and Estimated Risk of Death for Prognostic Groups

Risk Score	Prognostic Group	Estimate of Median Survival (95% CI)	Estimate of 12-month Survival (95% CI)	HR (95% CI)
Score < -0.995	1	9.5 (8.2 to 10.4)	36.0% (30.1% to 41.9%)	1.00
-0.995 ≤ Score < -0.661	2	7.1 (6.3 to 8.0)	26.4% (21.0% to 31.8%)	1.36 (1.13 to 1.62)
-0.661 ≤ Score < -0.293	3	4.9 (4.1 to 5.8)	16.3% (11.8% to 20.8%)	1.81 (1.51 to 2.17)
Score ≥ -0.293	4	2.7 (2.4 to 3.3)	6.2% (3.3% to 9.1%)	3.29 (2.74 to 3.95)

10.6. Generalisability

It may be argued that prognostic models derived using randomised controlled trial data are restricted to the patient population governed by the strict eligibility criteria of the trial and as such may lack generalisability (Moons *et al.*, 2009). Within pancreatic cancer trials, the eligibility criteria are usually mindful of the need to recruit on a national or international basis and are generally not overly restrictive. However, all patients do have to be fit enough to undergo any of the trial treatments and this does introduce the possibility of there being an additional worse prognostic

group of patients that are too ill to undergo any further treatment, hence excluded from clinical trials and not available for inclusion in the prognostic modelling. Data from randomised controlled trials are usually the largest datasets available in this disease and as such provide the most available evidence by which to investigate potentials prognostic factors.

Since trials of pancreatic cancer are being conducted by academia and pharmaceutical companies on an international level then it would seem sensible to derive a common outcomes dataset for this disease to ensure all trials subscribe to collection of a minimally important dataset with specific agreed outcome measurements. This could be addressed through the MRC Hubs for Trials Methodology Research (http://www.mrc.ac.uk/Newspublications/News/MRC005063) which are already investigating this in different disease sites and will ensure that potentially important information is routinely collected across all research projects undertaken within specific disease areas.

10.7. Application in Cardiac Surgery

The aim of the analysis was to investigate the reproducibility of the results and conclusions drawn from the comparison of non-linear methods in the advanced pancreatic cancer dataset, where the fractional polynomial transformation gave a better fit compared to the restricted cubic spline transformation. Data from 42802 patients with cardiovascular disease (Pagano *et al.*, 2009) were analysed as an additional example to investigate the stability of the conclusions in an alternative dataset with alternative event rate and survival distribution. Non-linear transformations were identified for all three continuous factors: historical log transformation for EuroScore and either 3 knot restricted cubic spline or second degree fractional polynomial transformations for both BMI and Carstiars Score.

The fractional polynomial and restricted cubic spline models were similar in that they selected identical variables for inclusion with high significance, similar magnitude and almost identical AIC values. Comparison of the restricted cubic spline and fractional polynomial strategies was carried out by calculating the sampling distribution of the difference in the AIC statistics between the models in multiple bootstrap resamples of the data. This was executed using an in-house developed SAS program. The distribution could then be summarised appropriately avoiding the need for distributional assumptions. The bootstrap mean AIC were also similar across the fractional polynomial and restricted cubic spline models with a bootstrap mean reduction with the fractional polynomial transformation of 0.12 (bootstrap percentile 95% CI: -8.7 to 9.2) in 1000 bootstrap resamples.

One of the main advantages of this additional dataset is the large sample size allowing further bootstrap analyses to be carried out which enabled an unadjusted univariate fractional polynomial transformation for each continuous covariate to be recalculated within each of 200 bootstrap resamples taken from the original dataset and comparing against a 5-knot restricted cubic spline using in-house developed SAS program. The %RCSPLINE and %MFP8 macros also had to be adapted to enable them to run within the %BOOT macro in SAS since they automatically delete temporary SAS datasets which were required to enable output of the -2log-likelihood, hence further calculation of the AIC, from each model in each bootstrap resample. The comparison of models was interesting since a final selected fractional polynomial model may be selected as the null model (i.e. no prognostic significance under any transformation including the linear model) and as such it was important to address the number of times each bootstrap sample compared a null model against the 5-knot restricted cubic spline. The large size of this additional dataset allowed this process to be repeated reducing the bootstrap resample sizes from that of the size of

the original dataset. When reducing the bootstrap size, consistency with conclusions made from analysis of the full dataset were not seen until a sample size of 20,000 resamples based on both the observed bootstrap mean difference and the inclusion/ exclusion of covariates by the fractional polynomial function. The fitted functions generated by restricted cubic splines and fractional polynomials were similar with similar model fit but larger differences were seen when the resample sizes were reduced in favour of the fractional polynomial functions albeit a reduction in available evidence to detect non-linearity.

10.8. Recommendations

- Stage of disease, sex, CA19-9, Albumin, Alkaline Phosphatase and LDH have been identified as statistically validated prognostic factors
- WBC, AST and BUN were prognostic in the development dataset but require external validation
- Performance status and grade of disease were prognostic in the validation dataset but require external validation
- All of these data should be routinely collected in future studies, minimising missing data, so the prognostic index can be continually updated and validated
- Multivariable analysis should be standard, automated selection procedures used with care
- Analyses should include dataset(s), stratification factor(s) and treatment group(s) as 'forced' covariates to adjust parameter estimation of effect size of possible prognostic factors
- Functional form should always be assessed specifically the underlying assumption of linearity between continuous covariates and log-hazard
- Dichotomisation is not a solution for non-linearity, rather non-linear transformations should be considered
- Fractional polynomial transformations are an alternative approach to restricted cubic spline transformations for multivariable model building of continuous covariates with non-linear relationships with log-hazard
- Frailty models should be considered to investigate clustering in response, specifically due to centre or country effect and particularly where the number of events per cluster is small
- Internal validation should be carried out, estimates adjusted for overfitting if necessary
- Predicted separation methods to quantify prognostic ability of the model should be reported
- The prognostic index should be based on characteristics known at the time of diagnosis, excluding dataset and treatment group, but retaining parameter estimates adjusted by these factors
- The prognostic index can be used to classify patients into one of four distinct clinically relevant risk groups
- Estimates of predicted survival and hazard ratios are available for each prognostic group
- A prognostic calculator could be made available on a pancreatic/ trials specific web page
- The prognostic index could aid patient management and treatment decision making
- Stratification for randomisation of patients in future trials could be based on the most significant of the validated prognostic factors specifically, Stage of disease, CA19-9 and Albumin
- Stratification could be based on categorising laboratory variables according to their normal laboratory range but retaining the continuous nature of each covariate at analysis
- Stratification factors could be updated based on other identified factors, once validated

10.9. Further Research

- Investigate the inclusion of interactions of non-linear continuous covariates in the prognostic model and consider the amount of evidence available to be able to detect (the need for ~four times as many patients to detect an interaction)
- Investigate the functional form of longitudinal non-linear continuous covariates, such as CA19-9 measurements during the treatment phase
- Investigate the prognostic value of quality of life, longitudinal data usually collected 3-monthly from the date of randomisation in prospective randomised controlled trials
- Further investigate 'centre' effect and more specifically 'country' effect in the international trials of adjuvant treatment for resected pancreatic cancer considering interactions with prognostic factors
- Apply the adopted strategy for the development of prognostic models in the adjuvant setting with full statistical validation in an international group of patients undergoing 'curative' surgery
- Investigate alternative parametric models which assume specific distributions for the data but if appropriate can result in a more efficient estimation procedure yielding more precise parameter estimates, smaller standard errors, more clinically meaningful results and more accurate prediction of survival

10.10. Final Conclusions

This thesis has further advanced the knowledge of prognostic factors in pancreatic cancer and has developed a validated prognostic index intended for use by clinicians to help in treatment decision-making and inform clinical trial design. This thesis has applied advanced statistical techniques in the development and validation of the prognostic index, has evaluated empirically the application of different polynomial based methods for addressing non-linearity in continuous data and further endorses the use of these methods. This research has provided recommendations for statistical analysis of pancreatic cancer datasets, recommendations for clinical use of the prognostic index and future clinical trial design and recommendations for further research in this area.

APPENDIX

```
/*****************************************/
/**** PROGNOSTIC FACTORS IN ADV PANC CA ****/
/****  FP VS RCS BOOTSTRAP RESAMPLING   ****/
/*****************************************/

LIBNAME BB 'H:\Debs\pfadvpancca';
%INCLUDE 'H:\Debs\pfadvpancca\FORMAT.SAS';

proc format;
  value trialfmt 1='BB128' 2='BB193';
  value racefmt 1='WHITE' 2='BLACK' 3='ORIENTAL' 4='OTHER';
  value sexfmt 1='MALE' 2='FEMALE';
  value deadfmt 0='ALIVE' 1='DEAD';
  value STATFMT 0='NEG' 1='POS' 2='MISS';
  VALUE TRTFMT 0='GEM+PLACEBO' 328='GEM+MARIM' 901='GEM' 1328='MARIM10'
               1329='MARIM25' 1337='MARIM5' ;
  VALUE TRT2FMT 0='GEM' 1='MARIM';
  VALUE AGEFMT 1='<=63' 2='>63';
  VALUE NORMFMT 0='NORMAL' 1='ABNORMAL';
  VALUE STAGEFMT 1='I/II' 2='III/IV';
RUN;

/**** SETUP COMPLETE CASE FP AND RCS DATA
/**** n=556 complete case, 520 events
/**** univariate transformations

/* %INCLUDE 'H:\Debs\phd\RCS macro\rcspline.sas';
RUN;

DATA BB.CC;
  SET BB.BBMERGE;
  LOGBILI=LOG(BILI);
  LOGLDH=LOG(LDH);
  CA199FP1=CA199**0.5;
  CA199FP2=CA199**0.5*LOG(CA199);
  %RCSPLINE(CA199, 10, 710, 70000);
  IF STAGEGP NE . AND ALB NE . AND CA199 NE . AND LDH NE . AND WBC NE . AND ALKPHOS NE .
  AND AST NE . AND BUN NE . ;
RUN;
*/

/**** BOOTSTRAP ANALYSIS: COMPARE FP vs. RCS using AIC */

FILENAME JACKBOOT 'H:\Debs\phd\jackboot.sas';
%INCLUDE JACKBOOT;

%MACRO ANALYZE (DATA=, OUT= );
```

```
/* FP MODEL USING AUTOMATED BACKWARD SELECTION */

ods output FitStatistics=AICFP;
PROC TPHREG DATA=&data;
  CLASS M (MISSING REF='0') N (MISSING REF='0');
  MODEL SURVDOE*DEAD(0)= TRIAL STAGEGP SEX TRTGP CA199FP1 CA199FP2
                          M N ALB AGE AST ALKPHOS BUN HAEM WBC LOGBILI LOGLDH
                         / SELECTION=BACKWARD INCLUDE=6 RL;
run;

DATA FP;
  SET AICFP;
  where criterion='AIC     ';
  AICFP=WithCovariates;
  keep criterion AICFP;
RUN;

DATA FP2;
  SET FP;  BY criterion;
  IF LAST.criterion;
RUN;


/* RCS MODEL USING AUTOMATED BACKWARD SELECTION */

ods output FitStatistics=AICRCS;
PROC TPHREG DATA=&data;
  CLASS M (MISSING REF='0') N (MISSING REF='0');
  MODEL SURVDOE*DEAD(0)= TRIAL STAGEGP SEX TRTGP CA199 CA1991
                         M N ALB AGE AST ALKPHOS BUN HAEM WBC LOGBILI LOGLDH
                         / SELECTION=BACKWARD INCLUDE=6 RL;
run;

DATA RCS;
  SET AICRCS;
  where criterion='AIC     ';
  AICRCS=WithCovariates;
  keep criterion AICRCS;
RUN;

DATA RCS2;
  SET RCS;  BY criterion;
  IF LAST.criterion;
RUN;


/* MERGE RESULTS */

DATA AIC;
  SET FP2;
  SET RCS2;
  AICDIFF=AICRCS-AICFP;
RUN;
```

```
/* SUMMARY STATISTICS ACROSS SAMPLES */
/* IF AICDIFF= + then AICFP<AICRCS indicating FP is better fit */

DATA &OUT;
  SET AIC;
  RUN;
%MEND;

title1 'BB: BOOTSTRAP ANALYSIS COMPARING AIC DIFFERENCE IN FP VS RCS MODELS';
title2 '1,000 bootstrap samples';
%BOOT(DATA=BB.CC, ALPHA=.05, SAMPLES=1000, RANDOM=123);
run;
quit;
run;


/************ END ************/
```

```
/*****************************************/
/****  PROGNOSTIC FACTORS IN HEART DATA  ****/
/**** BOOTSTRAP COMPARISON OF FP AND RCS ****/
/*****************************************/

/**** RERUN with: BOOTSTRAP SIZE=1000, 5000, 10 000, 20 000, SAMPLES=200 */

LIBNAME NF 'H:\Debs\phd\HEART';
RUN;

PROC FORMAT;
  VALUE op     1 = "CABG Only"        2 = "Valve(s) Only"    3 = "CABG+Valve(s)"
               4 = "CABG+Other"       5 = "Valve(s)+Other"   6 = "CABG+Valve(s)+Other"
               7 = "Other" ;
RUN;

%INCLUDE 'H:\Debs\phd\RCS macro\rcspline.sas';
RUN;

%INCLUDE 'H:\Debs\phd\FP macro\boxtid.sas';
%INCLUDE 'H:\Debs\phd\FP macro\brename.sas';
%INCLUDE 'H:\Debs\phd\FP macro\datasave.sas';
%INCLUDE 'H:\Debs\phd\FP macro\exinc.sas';
%INCLUDE 'H:\Debs\phd\FP macro\exlabb.sas';
%INCLUDE 'H:\Debs\phd\FP macro\fpmodels.sas';
%INCLUDE 'H:\Debs\phd\FP macro\funcfm.sas';
%INCLUDE 'H:\Debs\phd\FP macro\labs.sas';
%INCLUDE 'H:\Debs\phd\FP macro\MFP8.SAS';
%INCLUDE 'H:\Debs\phd\FP macro\xtop.sas';
%INCLUDE 'H:\Debs\phd\FP macro\xvars.sas';
RUN;

FILENAME JACKBOOT 'H:\Debs\phd\jackboot.sas';
%INCLUDE JACKBOOT;
RUN;

/* CREATE COMPLETE CASE DATASET FOR ANALYSIS AND COMPARISON OF MODELS
/* COMPARING 5-KNOT RCS BE MODEL WITH MV FP MODEL: SET UP 5K TRANS */

/*
DATA NF.BOOT2;
  SET NF.HEART2;
  IF FU_DAYS NE . AND BMI NE . AND SMOKING2 NE .;
  %RCSPLINE(BMI, 20.998698976, 24.84, 27.11, 29.714285714, 35.338593031);
  %RCSPLINE(CS, -3.5072, -2.0447, -0.539, 1.9408, 5.6578);
  %RCSPLINE(ES, 1, 3, 5, 7, 11);
RUN;
PROC CONTENTS DATA=NF.BOOT2;
run;
*/

/****  BOOTSTRAP ANALYSIS ****/

%MACRO ANALYZE (DATA=, OUT= );
```

```
/* FP MODEL USING AUTOMATED MPF8 MACRO */

DATA NF.TEMPNF;
SET &data;
RUN;

%mfp8 (DSNAME= NF.TEMPNF,
       YNAME=FU_DAYS,
       XNAME=BMI ,
       MODEL=S,
       PW=-2 -1 -0.5 0 0.5 1 2 3,
       M=2,
       CENSVAR=DEATH_AT_FU,
       CENSVAL=0,
       TIES=BRESLOW,
       ALPHA=0.05,
       MSELECT=RA2,
       MACPATH=H:\Debs\phd\FP macro,
       DSOUT=FPOUT,
       SHOWRES=r /* n */
       );
RUN;

DATA NFFP;
  SET FPOUT1;
  OUT=1;
RUN;

/* FFORM: -1="Omitted", 0="Linear", 1="First Degree", 2="Second Degree", 3="Third Degree" */

PROC SORT DATA=NFFP; BY OUT;
DATA NFFP2;
  SET NFFP;  BY OUT;
  IF LAST.OUT;
  DEVFP=DEVIANCE;
  IF FFORM=-1 THEN Q=0;
  IF FFORM IN (0 1) THEN Q=1;
  IF FFORM=2 THEN Q=2;
  IF FFORM=3 THEN Q=3;
  AICFP=DEVIANCE+(2*Q);
  KEEP DEVFP AICFP ;
RUN;


/* RCS MODEL USING TPHREG */

ods output FitStatistics=NFRCS;
PROC TPHREG DATA= &data; /*no summary;*/
       MODEL FU_DAYS*DEATH_AT_FU(0) = BMI BMI1 BMI2 BMI3 / RL;
RUN;

DATA NFRCS2;
  SET NFRCS;
  where criterion='-2 LOG L';
  DEVRCS=WithCovariates;
  AICRCS=DEVRCS+(2*4);        /* BETA ESTIMATE FOR RCS(5) */
  keep criterion DEVRCS AICRCS;
RUN;
```

```
DATA NFRCS3;
  SET NFRCS2;  BY criterion;
  IF LAST.criterion;
RUN;


/* MERGE RESULTS */

DATA DEV;
  SET NFFP2;
  SET NFRCS3;
  DEVDIFF=DEVRCS-DEVFP;
  AICDIFF=AICRCS-AICFP;
RUN;


/* COMPARISON OF AIC WITHIN BOOTSTRAP RESAMPLES: IF DEVDIFF= + then FP is better fit */

DATA &OUT;
  SET DEV;
  RUN;
%MEND;

title1 'NF HEART DATA: BOOTSTRAP ANALYSIS COMPARING AIC IN FP VS RCS MODELS';
title2 '200 bootstrap samples, SAMPLE SIZE =20,000';
%BOOT(DATA=NF.BOOT2, ALPHA=.05, SAMPLES=200, SIZE=20000, RANDOM=2001);
run;
quit;
run;


/******************* END *******************/
```

REFERENCES

Reference List

Akaike H (1974) New Look at Statistical-Model Identification. *Ieee Transactions on Automatic Control* AC19 (6): 716-723

Alexakis N, Halloran C, Raraty M, Ghaneh P, Sutton R, Neoptolemos JP (2004) Current standards of surgery for pancreatic cancer. *British Journal of Surgery* 91 (11): 1410-1427

Altman DG (2001) Systematic reviews in health care - Systematic reviews of evaluations of prognostic variables. *British Medical Journal* 323 (7306): 224-228

Altman DG, Anderson PK (1989) Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine* 8 771-783

Altman DG, Lyman GH (1998) Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Research and Treatment* 52 (1-3): 289-303

Altman DG, Royston P (2000) What do we mean by validating a prognostic model? *Statistics in Medicine* 19 (4): 453-473

Altman DG, Royston P (2006) Statistics notes - The cost of dichotomising continuous variables. *British Medical Journal* 332 (7549): 1080

Ambler G, Brady AR, Royston P (2002) Simplifying a prognostic model: a simulation study based on clinical data. *Statistics in Medicine* 21 (24): 3803-3822

Ambler G, Royston P (2001) Fractional polynomial model selection procedures: Investigation of type I error rate. *Journal of Statistical Computation and Simulation* 69 (1): 89-108

Bachmann MO, Alderson D, Peters TJ, Bedford C, Edwards D, Wotton S, Harvey IM (2003) Influence of specialization on the management and outcome of patients with pancreatic cancer. *British Journal of Surgery* 90 (2): 171-177

Bassi C, Stocken DD, Olah A, Friess H, Buckels J, Hickey H, Dervenis C, Dunn JA, Deakin M, Carter R, Ghaneh P, Neoptolemos JP, Buchler MW (2005) Influence of surgical resection and post-operative complications on survival following adjuvant treatment for pancreatic cancer in the ESPAC-1 randomized controlled trial. *Digestive Surgery* 22 (5): 353-363

Beger HG (1995) Prognostic Factors in Carcinoma of the Pancreas. *Langenbecks Archiv fur Chirurgie* 380 (3): 131-132

Berlin JD, Catalano P, Thomas JP, Kugler JW, Haller DG, Benson AB (2002) Phase III study of gemcitabine in combination with fluorouracil versus gemcitabine alone in patients with advanced pancreatic carcinoma: Eastern Cooperative Oncology Group Trial E2297. *Journal of Clinical Oncology* 20 (15): 3270-3275

Birkmeyer JD, Warshaw AL, Finlayson SRG, Grove MR, Tosteson ANA (1999) Relationship between hospital volume and late survival after pancreaticoduodenectomy. *Surgery* 126 (2): 178-183

Bonetti M, Gelber RD (2000) A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. *Statistics in Medicine* 19 (19): 2595-2609

Box GEP, Tidwell PW (1962) Transformation of Independent Variables. *Technometrics* 4 (4): 531

Bramhall SR, Allum WH, Jones AG, Allwood A, Cummins C, Neoptolemos JP (1995) Treatment and Survival in 13560 Patients with Pancreatic-Cancer, and Incidence of the Disease, in the West Midlands - An Epidemiologic-Study. *British Journal of Surgery* 82 (1): 111-115

Bramhall SR, Rosemurgy A, Brown PD, Bowry C, Buckels JAC (2001) Marimastat as first-line therapy for patients with unresectable pancreatic cancer: A randomized trial. *Journal of Clinical Oncology* 19 (15): 3447-3455

Bramhall SR, Schulz J, Nemunaitis J, Brown PD, Baillet M, Buckels JAC (2002) A double-blind placebo-controlled, randomised study comparing gemcitabine and marimastat with gemcitabine and placebo as first line therapy in patients with advanced pancreatic cancer. *British Journal of Cancer* 87 (2): 161-167

Burris HA, Moore MJ, Andersen J, Green MR, Rothenberg ML, Madiano MR, Cripps MC, Portenoy RK, Storniolo AM, Tarassoff P, Nelson R, Dorr FA, Stephens CD, VanHoff DD (1997) Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: A randomized trial. *Journal of Clinical Oncology* 15 (6): 2403-2413

Burton A, Altman DG (2004) Missing covariate data within cancer prognostic studes: a review of current reporting and proposed guidelines. *British Journal of Cancer* 91 4-8

Butturini G, Stocken DD, Wente MN, Jeekel H, Klinkenbijl JHG, Bakkevold KE, Takada T, Amano H, Dervenis C, Bassi C, Buchler MW, Neoptolemos JP (2008) Influence of resection margins and treatment on survival in patients with pancreatic cancer - Meta-analysis of randomized controlled trials. *Archives of Surgery* 143 (1): 75-83

Cameron JL, Pitt HA, Yeo CJ, Lillemoe KD, Kaufman HS, Coleman J, Herrington JL, Mason GR, Bradley EL, Jordan GL, Gadacz TR, Vanheerden JA, Watkins GH, Copeland EH (1993) 100 And 45 Consecutive Pancreaticoduodenectomies Without Mortality. *Annals of Surgery* 217 (5): 430-438

Cancer Research UK. Cancer Research UK Pancreatic Cancer Statistics - Key Facts. 2010. Ref Type: Online Source

Carr JA, Ajlouni M, Wollner I, Wong D, Velanovich V (1999) Adenocarcinoma of the head of the pancreas: Effects of surgical and nonsurgical therapy on survival - A ten-year experience. *American Surgeon* 65 (12): 1143-1149

Carter R, Stocken DD, Ghaneh P, Bramhall SR, Olah A, Kelemen D, Bassi C, Friess H, Dervenis C, Spry N, Buchler MW, Neoptolemos JP (2009) Longitudinal quality of life data can provide

insights on the impact of adjuvant treatment for pancreatic cancer-Subset analysis of the ESPAC-1 data. *International Journal of Cancer* 124 (12): 2960-2965

Collett D (1994) *Modelling Survival Data in Medical Research*. Chapman and Hall

Cox DR (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 34 (2): 187-220

Cubiella J, Castells A, Fondevila C, Sans M, Sabater L, Navarro S, Fernandez-Cruz L (1999) Prognostic factors in nonresectable pancreatic adenocarcinoma: A rationale to design therapeutic trials. *American Journal of Gastroenterology* 94 (5): 1271-1278

Cunningham D, Chau I, Stocken DD, Valle JW, Smith D, Steward W, Harper PG, Dunn J, Tudur-Smith C, West J, Falk S, Crellin A, Adab F, Thompson J, Leonard P, Ostrowski J, Eatock M, Scheithauer W, Herrmann R, Neoptolemos JP (2009) Phase III Randomized Comparison of Gemcitabine Versus Gemcitabine Plus Capecitabine in Patients With Advanced Pancreatic Cancer. *Journal of Clinical Oncology* 27 (33): 5513-5518

Doll R, Peto R, Wheatley K, Gray R, Sutherland I (1994) Mortality in Relation to Smoking - 40 Years Observations on Male British Doctors. *British Medical Journal* 309 (6959): 901-911

Ducreux M, Rougier P, Pignon JP, Douillard JY, Seitz JF, Bugat R, Bosset JF, Merouche Y, Raoul JL, Ychou M, Adenis A, Berthault-Cvitkovic F, Luboinski M (2002) A randomised trial comparing 5-FU with 5-FU plus cisplatin in advanced pancreatic carcinoma. *Annals of Oncology* 13 (8): 1185-1191

Durrleman S, Simon R (1989) Flexible Regression-Models with Cubic-Splines. *Statistics in Medicine* 8 (5): 551-561

Early Breast Cancer Trialists' Collaborative Group. (1990) *Treatment of Early Breast Cancer. Worldwide Evidence 1985-1990.* Oxford University Press

Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*. Chapman and Hall

Engelken FJF, Bettschart V, Rahman MQ, Parks RW, Garden OJ (2003) Prognostic factors in the palliation of pancreatic cancer. *European Journal of Surgical Oncology* 29 (4): 368-373

Everhart J, Wright D (1995) Diabetes-Mellitus As A Risk Factor for Pancreatic-Cancer - A Metaanalysis. *Jama-Journal of the American Medical Association* 273 (20): 1605-1609

Falconer JS, Fearon KCH, Ross JA, Elton R, Wigmore SJ, Garden OJ, Carter DC (1995) Acute-Phase Protein Response and Survival Duration of Patients with Pancreatic-Cancer. *Cancer* 75 (8): 2077-2082

Friedman GD, Vandeneeden SK (1993) Risk-Factors for Pancreatic-Cancer - An Exploratory-Study. *International Journal of Epidemiology* 22 (1): 30-37

Fujino Y, Suzuki Y, Ajiki T, Tanioka Y, Ku YS, Kuroda Y (2003) Predicting factors for survival of patients with unresectable pancreatic cancer: A management guideline. *Hepato-Gastroenterology* 50 (49): 250-253

Gordon TA, Burleyson GP, Tielsch JM, Cameron JL (1995) The Effects of Regionalization on Cost and Outcome for One General High-Risk Surgical-Procedure. *Annals of Surgery* 221 (1): 43-49

Gospodarowicz M, Henson D, Hutter R, O'Sullivan B, Sobin L, Wittekind C (2001) *International Union Against Cancer (UICC). Prognostic Factors in Cancer*.

Govindarajulu US, Spiegelman D, Thurston SW, Ganguli B, Eisen EA (2007) Comparing smoothing techniques in Cox models for exposure-response relationships. *Statistics in Medicine* 26 (20): 3735-3752

Gupta D, Lis CG, Dahlk SL, Vashi PG, Grutsch JF, Lammersfeld CA (2004) Bioelectrical impedance phase angle as a prognostic indicator in advanced pancreatic cancer. *British Journal of Nutrition* 92 (6): 957-962

Halm U, Schumann T, Schiefke I, Witzigmann H, Mossner J, Keim V (2000) Decrease of CA 19-9 during chemotherapy with gemcitabine predicts survival time in patients with advanced pancreatic cancer. *British Journal of Cancer* 82 (5): 1013-1016

Harrell, F. E. Harrell, F.E. SAS macros: %RCSPLINE %DASPLINE. 1991.
Ref Type: Computer Program

Harrell FE (2001) *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer

Harrell FE, Lee KL, Mark DB (1996) Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15 (4): 361-387

Heinzl H, Kaider A, Zlabinger G (1996) Assessing interactions of binary time-dependent covariates with time in cox proportional hazards regression models using cubic spline functions. *Statistics in Medicine* 15 (23): 2589-2601

Hess KR (1994) Assessing Time-By-Covariate Interactions in Proportional Hazards Regression-Models Using Cubic Spline Functions. *Statistics in Medicine* 13 (10): 1045-1062

Hollander N, Schumacher M (2006) Estimating the functional form of a continuous covariate's effect on survival time. *Computational Statistics & Data Analysis* 50 (4): 1131-1151

Howes N, Greenhalf W, Stocken DD, Neoptolemos JP (2004) Cationic trypsinogen mutations and pancreatitis. *Gastroenterology Clinics of North America* 33 (4): 767-

Huxley R, Ansary-Moghaddam A, de Gonzalez AB, Barzi F, Woodward M (2005) Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *British Journal of Cancer* 92 (11): 2076-2083

Ikeda W, Okada S, Tokuuye K, Ueno H, Okusaka T (2001) Prognostic factors in patients with locally advanced pancreatic carcinoma receiving chemoradiotherapy. *Cancer* 91 (3): 490-495

Ishii H, Okada S, Nose H, Yoshimori M, Aoki K, Okusaka T (1996) Prognostic factors in patients with advanced pancreatic cancer treated with systemic chemotherapy. *Pancreas* 12 (3): 267-271

Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ (2003) Cancer statistics, 2003. *Ca-A Cancer Journal for Clinicians* 53 (1): 5-26

Johnson CD, Puntis M, Davidson N, Todd S, Bryce R (2001) Randomized, dose-finding phase III study of lithium gamolenate in patients with advanced pancreatic adenocarcinoma. *British Journal of Surgery* 88 (5): 662-668

Kalbfleisch JD, Prentice RL (1980) *The Statistical Analysis of Failure Time Data*. Wiley, New York

Kalser MH, Ellenberg SS (1985) Pancreatic-Cancer - Adjuvant Combined Radiation and Chemotherapy Following Curative Resection. *Archives of Surgery* 120 (8): 899-903

Kaplan EL, Meier P (1958) Nonparametric-Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53 (282): 457-481

Karayiannakis AJ, Bolanaki H, Syrigos KN, Asimakopoulos B, Polychronidis A, Anagnostoulis S, Simopoulos C (2003) Serum vascular endothelial growth factor levels in pancreatic cancer patients correlate with advanced and metastatic disease and poor prognosis. *Cancer Letters* 194 (1): 119-124

Kuhlmann KFCD, de Castro SMM, Wesseling JG, ten Kate FJW, Offerhaus GJA, Busch ORC, van Gulik TM, Obertop H, Gouma DJ (2004) Surgical treatment of pancreatic adenocarcinoma: actual survival and prognostic factors in 343 patients. *European Journal of Cancer* 40 (4): 549-558

Lambert PC, Smith LK, Jones DR, Botha JL (2005) Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine* 24 (24): 3871-3885

Lehr S, Schemper M (2007) Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in Medicine* 26 (13): 2686-2698

Lin YS, Tamakoshi A, Kawamura T, Inaba Y, Kikuchi S, Motohashi Y, Kurosawa M, Ohno Y (2002) A prospective cohort study of cigarette smoking and pancreatic cancer in Japan. *Cancer Causes & Control* 13 (3): 249-254

Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK, Perrault J, Whitcomb DC, Aranha G, Banks P, Burton FR, CarrLocke D, Dyck WP, Gish RG, Goodale RL, Lehman G, Martin SP, Potts J, Sherman S, Ulrich CD, Yakshe P, Yeaton P, Hamanaka Y, Koizumi M, Tomioka T, Tsunoda T, Yamadera K, Delmont JP, Beger HG, Holstege A, Keim V, Layer P, Triantafillidis J, Boyle P, Cavallini G, Gullo L, Pedrazzoli S, Uomo G, Castano DGL, Ihse I, Buchler M, Elias E (1997) Hereditary pancreatitis and the risk of pancreatic cancer. *Journal of the National Cancer Institute* 89 (6): 442-446

Lundin J, Roberts PJ, Kuusela P, Haglund C (1995) Prognostic significance of serum CA 242 in pancreatic cancer. A comparison with CA 19-9. *Anticancer Research* 15 (5B): 2181-2186

Machin D, Cheung YB, Parmar MKB (2006) *Survival Analysis. A Practical Approach*. Wiley

Maisey N, Chau I, Cunningham D, Norman A, Seymour M, Hickish T, Iveson T, O'Brien M, Tebbutt N, Harrington A, Hill M (2002) Multicenter randomized phase III trial comparing protracted

venous infusion (PVI) fluorouracil (5-FU) with PVI5-FU plus mitomycin in inoperable pancreatic cancer. *Journal of Clinical Oncology* 20 (14): 3130-3136

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). *British Journal of Cancer* 93 (4): 387-391

Meier-Hirmer C, Ortseifen C, Sauerbrei W (2003) Multivariable Fractional Polynomials in SAS. *Institute of Medical Biometry, Freiburg*

Micke O, Bruns F, Kurowski R, Horst E, deVries AF, Hausler JW, Willich N, Schafer U (2003) Predictive value of carbohydrate antigen 19-9 in pancreatic cancer treated with radiochemotherapy. *International Journal of Radiation Oncology Biology Physics* 57 (1): 90-97

Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG (2009) Prognosis and prognostic research: what, why, and how? *British Medical Journal* 338 1317-1320

Neoptolemos, J. P., Buchler, M., Stocken, D. D., Ghaneh, P., Smith, D., Bassi, C., Moore, M., Cunningham, D., Dervenis, C., and Goldstein, D. (2009a) ESPAC-3(v2): A multicenter, international, open-label, randomized, controlled phase III trial of adjuvant 5-fluorouracil/folinic acid (5-FU/FA) versus gemcitabine (GEM) in patients with resected pancreatic ductal adenocarcinoma. Journal of Clinical Oncology 27, 18.
Ref Type: Abstract

Neoptolemos JP, Dunn JA, Stocken DD, Almond J, Link K, Beger H, Bassi C, Falconi M, Pederzoli P, Dervenis C, Fernandez-Cruz L, Lacaine F, Pap A, Spooner D, Kerr DJ, Friess H, Buchler MW (2001a) Adjuvant chemoradiotherapy and chemotherapy in resectable pancreatic cancer: a randomised controlled trial. *Lancet* 358 (9293): 1576-1585

Neoptolemos JP, Russell RCG, Bramhall S, Theis B (1997) Low mortality following resection for pancreatic and periampullary tumours in 1026 patients: UK survey of specialist pancreatic units. *British Journal of Surgery* 84 (10): 1370-1376

Neoptolemos JP, Stocken DD, Dunn JA, Almond J, Beger HG, Pederzoli P, Bassi C, Dervenis C, Fernandez-Cruz L, Lacaine F, Buckels J, Deakin M, Adab FA, Sutton R, Imrie C, Ihse I, Tihanyi T, Olah A, Pedrazzoli S, Spooner D, Kerr DJ, Friess H, Buchler MW (2001b) Influence of resection margins on survival for patients with pancreatic cancer treated by adjuvant chemoradiation and/or chemotherapy in the ESPAC-1 randomized controlled trial. *Annals of Surgery* 234 (6): 758-768

Neoptolemos JP, Stocken DD, Friess H, Bassi C, Dunn JA, Hickey H, Beger H, Fernandez-Cruz L, Dervenis C, Lacaine F, Falconi M, Pederzoli P, Pap A, Spooner D, Kerr DJ, Buchler MW (2004) A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. *New England Journal of Medicine* 350 (12): 1200-1210

Neoptolemos JP, Stocken DD, Smith CT, Bassi C, Ghaneh P, Owen E, Moore M, Padbury R, Doi R, Smith D, Buchler MW (2009b) Adjuvant 5-fluorouracil and folinic acid vs observation for pancreatic cancer: composite data from the ESPAC-1 and-3(v1) trials. *British Journal of Cancer* 100 (2): 246-250

Ni XG, Bai XF, Mao YL, Shao YF, Wu JX, Shan Y, Wang CF, Wang J, Tian YT, Liu Q, Xu DK, Zhao P (2005) The clinical value of serum CEA, CA19-9, and CA242 in the diagnosis and prognosis of pancreatic cancer. *Ejso* 31 (2): 164-169

NICE. National Institute for Clinical Excellence. 2010.
Ref Type: Online Source

Oettle H, Post S, Neuhaus P, Gellert K, Langrehr J, Ridwelski K, Schramm H, Fahlke J, Zuelke C, Burkart C, Gutberlet K, Kettner E, Schmalenberg H, Weigang-Koehler K, Bechstein WO, Niedergethmann M, Schmidt-Wolf I, Roll L, Doerken B, Riess H (2007) Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer - A Randomized controlled trial. *Jama-Journal of the American Medical Association* 297 (3): 267-277

Ohigashi H, Ishikawa O, Yokayama S, Sasaki Y, Yamada T, Imaoka S, Nakaizumi A, Uehara H (2003) Intra-arterial infusion chemotherapy with angiotensin-II for locally advanced and nonresectable pancreatic adenocarcinoma: Further evaluation and prognostic implications. *Annals of Surgical Oncology* 10 (8): 927-934

Pagano D, Freemantle N, Bridgewater B, Howell N, Ray D, Jackson M, Fabri BM, Au J, Keenan D, Kirkup B, Keogh BE (2009) Social deprivation and prognostic benefits of cardiac surgery: observational study of 44 902 patients from five hospitals over 10 years. *British Medical Journal* 338 B902

Paillaud E, Bories PN, Aita SL, Scherman E, Jeanfaivre V, Lejonc JL, Campillo B (2003) Prognostic value of dietary intake and inflammation on survival in patients with advanced cancer: Relationship with performance status, pain, and digestive disorders. *Nutrition and Cancer-An International Journal* 45 (1): 30-35

Palmer DH, Stocken DD, Hewitt H, Markham CE, Hassan AB, Johnson PJ, Buckels JAC, Bramhall SR (2007) A randomized phase 2 trial of neoadjuvant chemotherapy in resectable pancreatic cancer: Gemcitabine alone versus gemcitabine combined with cisplatin. *Annals of Surgical Oncology* 14 (7): 2088-2096

Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. *Ca-A Cancer Journal for Clinicians* 55 (2): 74-108

Peduzzi P, Concato J, Feinstein AR, Holford TR (1995) Importance of events per independent variable in proportional hazards regression analysis .2. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology* 48 (12): 1503-1510

Ridwelski K, Meyer F, Ebert M, Malfertheiner P, Lippert H (2001) Prognostic parameters determining survival in pancreatic carcinoma and, in particular, after palliative treatment. *Digestive Diseases* 19 (1): 85-92

Riley RD, Sauerbrei W, Altman DG (2009) Prognostic markers in cancer: the evolution of evidence from single studies to meta-analyis, and beyond. *British Journal of Cancer* 100 1219-1229

Roder J, Ott K (2001) Cancer of the Pancreas. In *UICC Prognostic Factors in Cancer*, Gospodarowicz M, Henson D, Hutter R, O'Sullivan B, Sobin L, Wittekind C (eds) pp 333-347. Wiley-Liss

Rothenberg ML, Moore MJ, Cripps MC, Andersen JS, Portenoy RK, Burris HA, Green MR, Tarassoff PG, Brown TD, Casper ES, Storniolo AM, VonHoff DD, Alberts DS, Wolff R, Kelsen DP, Kocha W, Haller DG, ModianoRevah M, Moore M, Oken MM, Schilsky RL, Shah A (1996) A phase II trial of gemcitabine in patients with 5-FU-refractory pancreas cancer. *Annals of Oncology* 7 (4): 347-353

Royston P, Altman DG (1994) Regression Using Fractional Polynomials of Continuous Covariates - Parsimonious Parametric Modeling. *Applied Statistics-Journal of the Royal Statistical Society Series C* 43 (3): 429-467

Royston P, Altman DG (1997) Approximating statistical functions by using fractional polynomial regression. *Statistician* 46 (3): 411-422

Royston P, Altman DG, Sauerbrei W (2006a) Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 25 (1): 127-141

Royston P, Parmar MKB, Sylvester R (2004) Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Statistics in Medicine* 23 907-926

Royston P, Reitz M, Atzpodien J (2006b) An approach to estimating prognosis using fractional polynomials in metastatic renal carcinoma. *British Journal of Cancer* 94 (12): 1785-1788

Royston P, Sauerbrei W (2004) A new measure of prognostic separation in survival data. *Statistics in Medicine* 23 (5): 723-748

Royston P, Sauerbrei W (2008) *Multivariable Model-building. A pragmatic approach to regression anaysis based on fractional polynomials for modelling continuous variables*. John Wiley and Sons Ltd UK

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York

Saad ED, Machado MC, Wajsbrot D, Abramoff R, Hoff PM, Tabacof J, Katz A, Simon SD, Gansl RC (2002) Pretreatment CA 19-9 level as a prognostic factor in patients with advanced pancreatic cancer treated with gemcitabine. *International Journal of Gastrointestinal Cancer* 32 (1): 35-41

Sarle, W. S. Sarle, W.S. SAS macro: %BOOT. 2000.
Ref Type: Computer Program

SAS Institute Inc., Cary NC USA. SAS 9.1.3. 1999.
Ref Type: Computer Program

Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine* 26 (30): 5512-5528

Sauerbrei W, Royston P, Bojar H, Schmoor C, Schumacher M (1999) Modelling the effects of standard prognostic factors in node-positive breast cancer. *British Journal of Cancer* 79 (11-12): 1752-1760

Schafer JL (1997) *Analysis of Incomplete Multivariate Data*. Chapman and Hall, New York

Schmoor C, Sauerbrei W, Schumacher M (2000) Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine* 19 (4): 441-452

Sener SF, Fremgen A, Menck HR, Winchester DP (1999) Pancreatic cancer: A report of treatment and survival trends for 100,313 patients diagnosed from 1985-1995, using the National Cancer Database. *Journal of the American College of Surgeons* 189 (1): 1-7

Shibamoto Y, Manabe T, Ohshio G, Sasai K, Nishimura Y, Imamura M, Takahashi M, Abe M (1996) High-dose intraoperative radiotherapy for unresectable pancreatic cancer. *International Journal of Radiation Oncology Biology Physics* 34 (1): 57-63

Siddiqui A, Heinzerling J, Livingston EH, Huerta S (2007) Predictors of early mortality in veteran patients with pancreatic cancer. *American Journal of Surgery* 194 (3): 362-366

Silverman DT, Dunn JA, Hoover RN, Schiffman M, Lillemoe KD, Schoenberg JB, Brown LM, Greenberg RS, Hayes RB, Swanson GM, Wacholder S, Schwartz AG, Liff JM, Pottern LM (1994) Cigarette-Smoking and Pancreas Cancer - A Case-Control Study Based on Direct Interviews. *Journal of the National Cancer Institute* 86 (20): 1510-1516

Simon R, Altman DG (1994) Statistical Aspects of Prognostic Factor Studies in Oncology. *British Journal of Cancer* 69 (6): 979-985

Sosa JA, Bowman HM, Gordon TA, Bass EB, Yeo CJ, Lillemoe KD, Pitt HA, Tielsch JM, Cameron JL (1998) Importance of hospital volume in the overall management of pancreatic cancer. *Annals of Surgery* 228 (3): 429-436

Stemmler J, Stieber P, Szymala AM, Schalhorn A, Schermuly MM, Wilkowski R, Helmberger T, Lamerz R, Stoffregen C, Niebler K, Garbrecht M, Heinemann V (2003) Are serial CA 19-9 kinetics helpful in predicting survival in patients with advanced or metastatic pancreatic cancer treated with gemcitabine and cisplatin? *Onkologie* 26 (5): 462-467

Stocken DD, Buchler MW, Dervenis C, Bassi C, Jeekel H, Klinkenbijl JHG, Bakkevold KE, Takada T, Amano H, Neoptolemos JP (2005) Meta-analysis of randomised adjuvant therapy trials for pancreatic cancer. *British Journal of Cancer* 92 (8): 1372-1381

Stocken DD, Hassan AB, Altman DG, Billingham LJ, Bramhall SR, Johnson PJ, Freemantle N (2008) Modelling prognostic factors in advanced pancreatic cancer. *British Journal of Cancer* 99 (6): 883-893

Stone CJ (1986) Comment: Generalized additive models. *Statistical Science* 1 312-314

Stone CJ, Koo CY (1985) Additive splines in statistics. *Proceedings of the Statistical Computing Section, ASA* 45-48

Storniolo AM, Enas NH, Brown CA, Voi M, Rothenberg ML, Schilsky R (1999) An investigational new drug treatment program for patients with gemcitabine - Results for over 3000 patients with pancreatic carcinoma. *Cancer* 85 (6): 1261-1268

Talar-Wojnarowska R, Gasiorowska A, Strzelczyk J, Janiak A, Malecka-Panas E (2003) Prognostic factors in the operative and palliative treatment of pancreatic cancer. *Neoplasma* 50 (5): 383-387

Tas F, Aykan F, Alici S, Kaytan E, Aydiner A, Topuz E (2001) Prognostic factors in pancreatic carcinoma - Serum LDH levels predict survival in metastatic disease. *American Journal of Clinical Oncology-Cancer Clinical Trials* 24 (6): 547-550

Terwee CB, Van Dijkum EJMN, Gouma DJ, Bakkevold KE, Klinkenbijl JHG, Wade TP, van Wagensveld BA, Wong A, van der Meulen JHP (2000) Pooling of prognostic studies in cancer of the pancreatic head and periampullary region: The triple-P study. *European Journal of Surgery* 166 (9): 706-712

The R Foundation for Statistical Computing. R version 2.7.2. 2008.
Ref Type: Computer Program

Therneau TM, Grambsch PM (2000) *Modeling Survival Data. Extending the Cox Model*. Springer

Trigui B, Barrier A, Flahault A, Huguier M (2000) Prognostic factors in advanced pancreatic carcinoma. Multivariate analysis and predictive score of survival. *Annales de Chirurgie* 125 (7): 625-630

Tsuruta K, Okamoto A, Egawa N, Kamisawa T, Karasawa K, Takahashi T (2001) Survival benefits of adjuvant chemotherapy with oral doxifluridine (5 '-DFUR) following radiotherapy in patients with unresectable pancreatic cancer. *Journal of Surgical Oncology* 78 (3): 202-207

Ueno H, Okada S, Okusaka T, Ikeda M (2000) Prognostic factors in patients with metastatic pancreatic adenocarcinoma receiving systemic chemotherapy. *Oncology* 59 (4): 296-301

Wagner M, Redaelli C, Lietz M, Seiler CA, Friess H, Buchler MW (2004) Curative resection is the single most important factor determining outcome in patients with pancreatic adenocarcinoma. *British Journal of Surgery* 91 (5): 586-594

Watanabe I, Sasaki S, Konishi M, Nakagohri T, Inoue K, Oda T, Kinoshita T (2004) Onset symptoms and tumor locations as prognostic factors of pancreatic cancer. *Pancreas* 28 (2): 160-165

Yasue M, Sakamoto J, Teramukai S, Morimoto T, Yasui K, Kuno N, Kurimoto K, Ohashi Y (1994) Prognostic Values of Preoperative and Postoperative Cea and Ca19.9 Levels in Pancreatic-Cancer. *Pancreas* 9 (6): 735-740

Yeo CJ (1998) Pancreatic cancer: 1998 update. *Journal of the American College of Surgeons* 187 (4): 429-442

Ziske C, Schlie C, Gorschluter M, Glasmacher A, Mey U, Strehl J, Sauerbruch T, Schmidt-Wolf IGH (2003) Prognostic value of CA 19-9 levels in patients with inoperable adenocarcinoma of the pancreas treated with gemcitabine. *British Journal of Cancer* 89 (8): 1413-1417