THE UNIVERSITY OF
WARWICK

AUTHOR: **Martine J. Barons** DEGREE: **Ph.D.**

TITLE: **What is the added value of using non-linear models to explore complex healthcare datasets?**

DATE OF DEPOSIT: .....

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

AUTHOR'S SIGNATURE: .....

USER'S DECLARATION

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE SIGNATURE ADDRESS

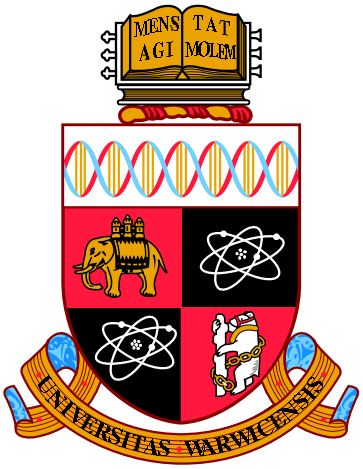
.....

.....

.....

.....

.....



What is the added value of using non-linear models to explore complex healthcare datasets?

by

Martine J. Barons

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Centre for Complexity Science

October 2013

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

No PhD thesis is completed in isolation and I am grateful for all my supportive friends and colleagues. You are too many to acknowledge individually, but I am grateful for each of you.

I acknowledge with enormous gratitude the support of my PhD supervisors Professor Frances Griffiths, Dr Nick Parsons and Professor Margaret Thorogood for lending me their expertise and giving freely of their time, patience, guidance and encouragement not only in my completion of this thesis and the associated research, but in the wider endeavour of becoming a research scientist, my public engagement activities and my involvement in professional bodies the IMA, IOP and RSS.

I would like to thank the staff and students at Warwick's Complexity Science doctoral training centre for making it such a rich and invigorating place to work, with special thanks to its directors Professor Robin Ball and Professor Robert MacKay for their wisdom in designing and developing the centre, and for consistently warm personal encouragement throughout my MSc and PhD years.

I thank the Engineering and Physical Sciences Research Council (EPSRC) for funding the doctoral training centre and my studentship. I also thank the centre, Warwick Medical School, the IMA and Warwick Statistics for additional conference funding.

I thank Dr Sally Turner for the use of the Basingstoke and Alton cardiac rehabilitation data, for hosting my visit to her clinic, for assistance in understanding the

data and its context, and for insightful comments on the research paper currently in preparation. I thank Professor Sallie Lamb for the use of the BeST trial data and for helpful comments on the research paper now submitted.

I express my gratitude to Professor Colin Reeves for helpful discussions and to him and Charles Johnston for supplying the MATLAB® script offered in Reeves and Johnston [2008].

I thank my father, Nigel Weaver, fellow PhD student Pádraig Mac Carron and Colin Reeves for careful proof reading of large parts of this thesis.

I cannot thank enough my husband, Chris, for his loving support, understanding and forbearance throughout the duration of my research studies as I established a new career after our sons had grown up. I thank our family Alex and Holly, Ed and Lauren, Greg and Amy for their support, encouragement and understanding and I thank Greg Barons for the production of Figure 5.1 using Bishop [1996].

I am eternally grateful to Jesus Christ who willingly grants to the undeserving, including me, more than we can ask or imagine.

This thesis was typeset with LaTeX 2$_\varepsilon$[1]

[1] LaTeX 2$_\varepsilon$ is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The style package *warwickthesis* was used.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work I did using an artificial neural network for exploring the BeST data in my MSc mini-project was a useful introduction the the data set and an opportunity to learn about machine learning techniques. None of that introductory work forms part of this thesis.

List of publications including submitted papers.

Martine J. Barons, Nick Parsons, Frances Griffiths, and Margaret Thorogood. A comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioural approach to treatment for non-specific low back pain. In *IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE) 2013*, pages 7-11, 2013b.

Martine J. Barons, Frances E. Griffiths, Nick Parsons, Anca Alba, Margaret Thorogood, Graham F. Medley, and Sallie Lamb. Matching patients to an intervention for back pain: classifying patients using a latent class approach. *Journal of Evaluation in Clinical Practice*, SUBMITTED, 2013a.

Abstract

Health care is a complex system and it is therefore expected to behave in a non-linear manner. It is important for the delivery of health interventions to patients that the best possible analysis of available data is undertaken. Many of the conventional models used for health care data are linear. This research compares the performance of linear models with non-linear models for two health care data sets of complex interventions.

Logistic regression, latent class analysis and a classification artificial neural network were each used to model outcomes for patients using data from a randomised controlled trial of a cognitive behavioural complex intervention for non-specific low back pain. A Cox proportional hazards model and an artificial neural network were used to model survival and the hazards for different sub-groups of patients using an observational study of a cardiovascular rehabilitation complex intervention.

The artificial neural network and an ordinary logistic regression were more accurate in classifying patient recovery from back pain than a logistic regression on latent class membership. The most sensitive models were the artificial neural network and the latent class logistic regression. The best overall performance was the artificial neural network, providing both sensitivity and accuracy.

Survival was modelled equally well by the Cox model and the artificial neural network, when compared to the empirical Kaplan-Meier survival curve. Long term survival for the cardiovascular patients was strongly associated with secondary prevention medications, and fitness was also important. Moreover, improvement in fitness during the rehabilitation period to a fairly modest 'high fitness' category was as advantageous for long-term survival as having achieved that same level of fitness by the beginning of the rehabilitation period. Having adjusted for fitness, BMI was not a predictor of long term survival after a cardiac event or procedure.

The Cox proportional hazards model was constrained by its assumptions to produce hazard trajectories proportional to the baseline hazard. The artificial neural network model produced hazard trajectories that vary, giving rise to hypotheses about how the predictors of survival interact in their influence on the hazard.

The artificial neural network, an exemplar non-linear model, has been shown to match or exceed the capability of conventional models in the analysis of complex health care data sets.

Table of Abbreviations

Abbreviation	Meaning	Explanation
AACVPR	American Association of Cardiovascular and Pulmonary Rehabilitation	Page 56
ACE	Angiotensin converting enzyme	Page 28
AM	Active management	Page 45
ANN	Artificial Neural Network	Page 89
ANOVA	Analysis of Variance	Page 18
BB	Beta blockers	Page 28
BeST	The Back Skills Training Trial	Page 44
BMI	Body Mass Index	Page 22
CBA	Cognitive Behavioural Approach	Page 45
CR	Cardiac rehabilitation	Page 59
EM	Expectation maximisation	Page 67
EQ-5D	EuroQol	Page 23
FA	Fear avoidance	Page 47
FABQ	Fear Avoidance Beliefs Questionnaire	Page 47
HADS	Hospital Anxiety and Depression Scale	Page 47
LBP	Low back pain	Page 45
LCA	Latent Class Analysis	Page 64
MAR	Missing at random	Page 83
MCAR	Missing completely at random	Page 83
MI	Myocardial infarction	Page 27
MLP	Multi-layer perceptron	Page 89
MMPI	Minnesota Multiphasic Personality Inventory	Page 18
MNAR	Missing not at random	Page 83
MVK	Modified von Korff	Page 46
NICE	National Institute for Health and Clinical Excellence	Page 17
PH	Proportional hazards	Page 76
PSE	Pain self-efficacy	Page 47
QALY	quality-adjusted life year	Page 46
QoL	Quality of life	Page 47
RCT	Randomised controlled trial	Page 17
RMQ	Roland Morris Disability Questionnaire	Page 46
SF12	Short Form 12	Page 47
VO_2 max	Measured peak oxygen intake	Page 28

Chapter 1

Introduction

1.1 Complex interventions

A complex intervention is conventionally defined as an intervention that has many parts which may work independently or interdependently (Campbell et al. [2000]). This thesis utilises data from two complex interventions which are described in detail in Chapter 3. This chapter introduces complex interventions and describes how they differ from other intervention types and how their development, implementation and evaluation are conducted using insights from complex adaptive systems concepts.

It is now accepted that many health care interventions are complex (Craig et al. [2008], Emsley et al. [2010], Lancaster et al. [2010], Burton [2012]). The contrast with more simple interventions, such as a drug given to treat a single condition is that in clinical trials to establish efficacy, most sources of variability can be identified and controlled for, either directly or by randomisation (Burton [2012], Hawe et al. [2004]). There is no sharp boundary between simple and complex interventions (Craig et al. [2008]). The greater the difficulty in defining precisely what exactly the active ingredients of an intervention are and how they relate to each other, the greater the likelihood that you are dealing with a complex intervention (Craig et al. [2008]). Some highly complex interventions may comprise a set of individually complex interventions, such as the Sure Start intervention to support families with young children in deprived communities (Craig et al. [2008]).

The Medical Research Council identifies 5 factors which make an intervention complex:

- Number of interacting components within the experimental and control interventions

- Number and difficulty of behaviours required by those delivering or receiving the intervention
- Number of groups or organisational levels targeted by the intervention
- Number and variability of outcomes
- Degree of flexibility or tailoring of the intervention permitted

(Craig et al. [2008]).

In addition to interventions, health problems themselves can have different levels of complexity. Campbell *et al.* (Campbell et al. [2007]) illustrate this with the example of cardiovascular disease: High death rates in people with cardiovascular disease are affected by:

- Disease-Atherosclerosis, risk factors (cholesterol, blood pressure, smoking), co-morbidity
- Patient-Beliefs about lifestyle, adherence to treatment, and symptoms
- Practitioner-Accessibility, prescribing practices, practices in health promotion
- Health service-Availability of effective preventive and therapeutic care
- Policy-Policies on preventive services (tobacco control, diet, exercise, etc)
- Social context-Socioeconomic status, social support.

It is important to recognise that some health problems can occur at multiple levels, if a decision to intervene at one level could be cancelled out or promoted by actions at other levels. For example, improving practitioners' health promotion practices may have no effect on patients' health behaviour if social and environmental factors obstruct response. One example of the importance of context was found in a trial of a secondary prevention programme for cardiovascular disease. Interviews and focus groups showed that because patients viewed heart attacks as self-limited episodes, they were less willing to adopt long term lifestyle changes. It also showed that the ability of practice nurses to provide skilled continuity of care was affected by both their lack of training and by their low status in the primary health care team. The understanding of complex interventions include issues of service and organisation and the frameworks of understanding that inform the behaviour of healthcare users; these are features of the interactions between intervention and environment that characterise complex interventions (Oakley et al. [2006]). This connection between levels means that sometimes the psychological domain is targeted to bring relief

to physical suffering, such as a cognitive behavioural approach to angina (Lewin et al. [2002]) or back pain (Lamb et al. [2010b]), or a cardiovascular rehabilitation intervention, which typically includes both psychological support and appropriate medication, exercise, and education (Turner [2007]). Since there are, by definition, many parts which may work independently or interdependently in a complex intervention, it can be difficult to design such an intervention such that it can be tested for efficacy and the successful parts or combinations identified. To this end, a panel of experts produced a framework for the design and evaluation of complex intervention in health in the year 2000 (Campbell et al. [2000]). They define a 5-stage protocol, which need not be linear, where at each stage the findings may inform the next or a reiteration of the previous stage. The 5 stages were:

- The preclinical stage for exploration of relevant theory
- Phase I modelling stage, for identifying the components of the intervention and their modes of effect.
- Phase II an exploratory trial where the insights from theory and modelling are tested
- Phase III the definitive randomised controlled trial, testing the complex intervention against an appropriate alternative
- Phase IV, the long-term implementation.

This framework was later criticised to be too close to the framework for a drug trial, which is usually a simple intervention. The framework was reviewed and updated in 2008 as experience of evaluating complex interventions accumulated and interest in the methodology had grown (Craig et al. [2008]). Several papers had identified limitations in the framework, recommending, for example, greater attention to early phase piloting and development work, a less linear model of evaluation process, integration of process and outcome evaluation, recognition that complex interventions may work best if they are tailored to local contexts rather than completely standardised, and greater use of the insights provided by the theory of complex adaptive systems (Craig et al. [2008]). However, the 2000 framework provides a useful way to organise an introduction to complex interventions and the issues that surround them.

Preclinical or theory phase

In the preclinical stage, the relevant theory is explored in order to inform the design and select the intervention, to develop a clear hypothesis and to identify major

confounders and strategic design issues. Identifying evidence that an intervention might have the desired effect will include taking relevant evidence from outside health sciences (e.g. theory of organisational change) and may lead to adjustment of the hypothesis. Preliminary work can include simulation, qualitative testing through focus groups, etc. and case studies. Qualitative research can be used to ascertain the mechanism of efficacy and barriers to be overcome. A range of research methods can be used to collect evidence: systematic literature reviews, epidemiological research, and expert opinion to quantify the extent of the problem and identify the groups most at risk and the key modifiable risks. For example, reasons for delayed presentation by patients with symptoms of lung cancer are poorly understood, so epidemiological and qualitative research is being undertaken to identify and quantify determinants and targets that may be amenable to intervention (Campbell et al. [2007]).

Conceptual modelling or mapping can clarify the mechanisms by which an intervention might achieve its aims. The process involves mapping out the mechanisms and pathways proposed to lead from the intervention to the desired outcomes, then adding evidence and data to this map. Modelling of the intervention both depends on, and informs, understanding of the underlying problem. The intervention must engage the target group and affect pathways amenable to change that are identified as important to the problem (Campbell et al. [2007]).

The term 'mediator' is commonly used for a variable on a causal pathway, and 'moderator' for a variable which modifies the strength of part or all of a causal pathway. Complex interventions have, by definition, multiple components and are therefore characterised by complex treatment effect mechanisms with multiple mediators, with the possibility of moderators such as the background characteristics and environment of the patient.(Emsley et al. [2010]).

Optimising combinations of components in the intervention is not a straightforward task and there is no consensus on how to achieve this. Once a conceptual model has been formed, some complex interventions may be amenable to simulations or carefully controlled experimental studies outside the normal clinical setting such as simulated patients. In one study, simulated patients were used to test the intervention with general practitioners, which identified the likely outcomes for a range of patients and allowed general practitioners to comment on how the intervention could be improved. Simulation can also be used to explore the effect of changes in dose on response, and changes in contextual influences. (Campbell et al. [2007]).

Phase I or Modelling phase

The second stage is Phase I when the intervention and the underlying mechanism by which they will influence the outcome are identified. The purpose of this phase is to provide evidence that prediction of the relation between the components and the interaction between them can be made. Clearly, there is some overlap with the previous phase in methodologies used to achieve this; simulated patients, for example, could be used in this phase, too. Preliminary surveys, focus groups and case studies can be used to define components of the complex intervention and detailed descriptions of the setting or intervention variants can inform the design. Potential barriers can be identified through qualitative research, such as whether knowledge or lack of time or resources is the main barrier, and the intervention can be designed appropriately. Some researchers identified barriers to patient participation that included concerns about information and consent; patient preferences for treatment and additional demands such as additional procedures. Barriers to clinician participation included lack of staff and training; concern about the impact on the doctor patient relationship; time constraints and difficulties with the consent procedure (Lancaster et al. [2010]).

Phase II or Exploratory Trial

Phase II of this scheme is the exploratory trial in which a fully defined complex intervention is compared with a suitable alternative. For this, the protocol must be theoretically defensible, reproducible and adequately controlled. The study must also have appropriate statistical power. Lancaster et al (Lancaster et al. [2010]) provide a detailed discussion of statistical issues related to complex interventions in primary care, with particular emphasis on cluster randomised trials which are commonly used with complex interventions. Complex intervention trials often have designs with multiple parts and multiple routes through which an intervention can act and operate. Consequently there are larger number of covariates that may predict outcome, and which need to be considered to avoid confounding than would be found in a simple intervention for the same health problem (Lancaster et al. [2010]).

The purpose of a trial, including the exploratory one, is to evaluate efficacy of the intervention, to identify the mechanisms of action, the components delivering change and the potential for tailoring the complex intervention to individual patients. As such, the complex intervention trial needs to be a sophisticated clinical experiment designed to test the theories motivating the intervention and will also help understand the underlying nature of the clinical problem being treated (Emsley

et al. [2010]).

Key objectives of a pilot study are:

- Test the integrity of the study protocol. This is especially important if multiple sites are to be involved. This includes evaluating the inclusion/exclusion criteria, determining if interim analyses are necessary.
- Sample size calculation. This is to obtain initial estimates, considering variability and an estimate of intra-class correlation coefficient if the main trial is to be cluster randomised.
- Recruitment and consent rates. This is important for planning the length of the study, the strategy for recruitment of practices and participants, how to explain the study in layman's terms, trial the readability of patient information and consent forms.
- Develop and test the implementation of the complex intervention. To determine the optimal duration of delivery of the intervention, the ease of adherence, testing of materials, equipment and techniques, including if self-administration is possible, and whether support, such as an on-call help service is needed
- Determine the acceptability of the intervention to participants, assessors and funders, ascertain if there are any side effects and if costs are feasible and undertake pre-trial modelling of cost-effectiveness.
- Train staff in delivery and assessment procedures, determine inter-rater and intra-rater reliability if applicable, and carry out calibration of instruments. It is also important to trial the data collection, recording and data entry.
- Selection of most appropriate primary outcome measure, and determine if it would be useful to use more than one primary outcome measure or secondary outcome measures or biomarkers. Whether patient-reported outcomes are appropriate and reliable.
- Randomisation procedure implementation, including whether use of a 24 hour randomisation service is needed and its acceptability to participants.
- Pilot data collection forms and/or questionnaires and assess their face/content validity, whether the self-completion should take place at hospital or at home and if the use of postal questionnaires is appropriate or whether home visits and interviewers are required.

- Prepare and plan data collection and monitoring procedures, covering databases, data entry, validation methods, backup, forms for monitoring adverse events and missing data minimisation strategies.

(Lancaster et al. [2010]).

The exploratory phase should ideally itself be randomised to allow assessment of the size of the effect. This initial assessment will provide a sound basis for calculating sample sizes for the main trial. Other design variables can also be established in an exploratory trial. Variability in individual level outcomes may reflect higher level processes and sample sizes may need to be larger to take account of the extra variability and cluster randomised designs considered (Craig et al. [2008]). The findings from a trial of a complex intervention are more generalisable if the trial takes place in the setting in which the eventual intervention is likely to be delivered and with the population likely to be offered the treatment.

In complex interventions, individual randomisation is often not possible, so cluster randomisation frequently is used (in which clusters (or groups) of individuals, rather than individuals themselves, are randomised.). This requires active participation from general practices, nursing homes or households, for example, depending upon the unit of randomisation (Lancaster et al. [2010]). It is also often not possible to conceal the allocation of the treatment from the patient, practitioner and researcher and the potential biases of unblinded trials need to be accounted for. A preference trial design is one where any patient who expresses a preference for one of the alternate treatment regimes is allocated to their preference and others at random, but results from a preference trial design can be hard to interpret. There is evidence that treatments patients prefer have greater effects (Adamson et al. [2008]). The CONSORT statements for individual and cluster randomised trials are helpful and informative for planning randomisation. They cover issues such as procedure, method and level of randomisation. The decision to randomise clusters requires justification because of the associated cost of the increased numbers of patients involved. The avoidance of potential contamination between individuals in different study arms is a common justification. It may also be the case that an individually randomised trial is impossible or that a cluster randomisation would reflect the manner in which an intervention would ultimately be delivered. (Lancaster et al. [2010]).

With this type of design, there is a higher probability that arms in a study will become imbalanced in size, and that baseline covariates will show an imbalanced distribution at the level below the unit of randomisation, usually individuals or possibly, in studies with relatively few clusters, at the level of randomisation itself. There

are two issues to consider for cluster randomised trials to optimise power: ensuring equal numbers of clusters and ensuring equal numbers of patients per treatment arm. In studies where few units of observation are randomised imbalances may lead to a reduction in study power. A design-stage alternative is to eliminate the occurrence of imbalance in chosen covariates by using a restricted randomisation method such as stratification by covariates when there are sufficient randomisation units for this to be practicable (Lancaster et al. [2010]). Eldridge *et al.* (Eldridge et al. [2005]) give examples of different recruitment strategies and discuss what can be done to avoid bias in identifying and recruiting participants to cluster randomised trials, where recruitment may operate at a number of different levels (Lancaster et al. [2010]).

Care must be taken, therefore with eligibility criteria and recruitment of intervention sites and personnel: a trial of complex intervention needs to consider expertise of health professionals as well as investigations, drugs, treatment guidelines, arrangements for discharge and follow up and the organisation, management and skill mix of the unit where the intervention is delivered

Appropriate methods of analysis for individually randomised trials are based on standard methods of statistical analysis, which are well-documented. Whilst there are more complex issues to consider when analysing cluster randomised trials, where researchers have to account for groups of individuals within clusters (Lancaster et al. [2010]), neither of the complex interventions considered in this thesis were assessed through a cluster randomised trial, so details are not given here. Details of the analysis methods used are given in Chapter 4.

During Phase II, it is important to test for effects such as a learning curve, the feasibility of delivering the intervention in different places, consistency of delivery, and the acceptability of the intervention to both providers and patients. If evidence of a learning curve is found, adjustments should be made, such as a run-in period before recruitment to a trial to overcome confounding due to a learning curve in providers.

The feasibility and acceptability of an intervention can be measured by the level of attrition found in the study in those that are willing to participate. A research question that is considered to be important to primary care will usually make recruitment easier, as it will engender greater interest at practice level. It is also important to consider how the study will be perceived to impact on the patient-doctor relationship more generally, and the priority given to the research question in relation to other issues. Information about low participation rates can also provide information about the feasibility and generalisability of likely uptake of the intervention. A systematic review can be used to explore this across several randomised

controlled trials. Relative attrition (attrition in the intervention group divided by the attrition in the control group) has been used as a measure of acceptability that allows the calculation of an overall effect estimate and the study of different levels of attrition based on the population of each trial (Lancaster et al. [2010]).

Suitable outcome measures should be ascertained by the exploratory trial. Outcomes can be measured at the individual level or at the level of a group or cluster, which in primary care research is usually the GP practice. In choosing the primary and secondary outcomes to be measured, it is vital that they are fit for purpose, both theoretically and practically. An example of an interventions in primary care administered at the practice level might be the introduction of a universal parenting programme to prevent early childhood behavioural problems, with main outcomes concerning parenting, child behaviour and maternal mental health. In general, outcome measures need to be valid (i.e. shown to have face/concurrent/predictive validity), to be repeatable (i.e. stable over time when the disease state is not changing), to be reproducible (i.e. when applied by different assessors), and to be objectively measured in situations where self-reporting may be unreliable (for example, self-reported smoking cessation with additional biochemical validation (Lancaster et al. [2010]).

The MRC guidance on developing and evaluating complex interventions states that researchers need to decide which outcomes are most important, which are secondary, and how they will deal with multiple outcomes in the analysis. A single primary outcome and a small number of secondary outcomes are the most straightforward for statistical analysis but may not represent the best use of the data or provide an adequate assessment of the success or otherwise of an intervention that has effects across a range of domains. It is important also to consider which sources of variation in outcomes matter and to plan appropriate subgroup analyses (Craig et al. [2008]).

All interventions need to be cost effective if they are to be used. Model building can be a cost-effective way of evaluating the likely quality-adjusted life-year (QALY) of an intervention from a complex intervention using data from a pilot study, which can also provide estimates of likely effectiveness and cost.

Eldridge *et al.* (Eldridge et al. [2005])developed a cost-effectiveness model of a complex falls prevention intervention from pilot study data in order to assess the viability and design of a subsequent trial. Using two models, they first estimated the probability of falling over a 12-month period using a probability tree, and then used a Markov simulation to assess the impact of the programme over time. The probability tree model showed that the intervention would reduce the proportion falling by only

2.8% over a 12-month period. The major reason for this small effect was that less than a quarter of older people at risk of falling were assessed using their screening tool. Sensitivity analyses showed that the only scenarios that produced a substantial increase in the effect of the intervention were those in which all older people are assessed, and this was not cost-effective. They found that even if policy-makers were willing to spend £30,000 per QALY gained, there was only a 40% chance that the intervention would be cost-effective (Lancaster et al. [2010]). The value of well designed pilot or feasibility studies, prior to large multi-centre randomised controlled trials, is illustrated by the feasibility study for the UK Back pain, Exercise, Active management and Manipulation (UK BEAM) trial. The aim was to pilot all aspects of the trial including the intervention, to identify problems in design or execution, to investigate unresolved issues and to demonstrate that the main trial could fulfil its aims, in terms of its design and implementation, and was therefore worthy of funding. Overall, the feasibility study demonstrated that the majority of methods and processes were successful. It identified where changes were required to the trial design or execution and highlighted unexpected problems, allowing further design changes before the start of the main trial. This study proved that pilot studies are vital, especially when evaluating complex interventions, for providing planning information and identifying unanticipated issues in advance of expensive, complex trials (Lancaster et al. [2010]).

Using the information from the exploratory trial, the optimum intervention is developed, possibly in a variety of versions, and consideration given to consistency of delivery and whether there is evidence of improved performance of the intervention over time. A suitable control intervention and outcome measures will have been found.

Phase III the definitive trial

Having established the effect size, sample sizes, inclusion and exclusion criteria, and methods of randomisation, as well as the challenges of complex interventions in the pilot, Phase III, the main trial, can be undertaken (Campbell et al. [2000]). Process evaluation in a trial is an example of the trend to move beyond the rhetoric of quantitative versus qualitative methods. Process evaluations within trials explore the implementation, receipt, and setting of an intervention and help in the interpretation of the outcome results. Process evaluation can help to distinguish between interventions that are inherently faulty (failure of intervention concept or theory) and those that are badly delivered (implementation failure). Process evaluations are especially necessary in multi-site trials, where the "same" intervention may be

implemented and received in different ways (Oakley et al. [2006]). For example, the RIPPLE (randomised intervention of pupil peer-led sex education) study was a cluster RCT designed to investigate whether peer delivered sex education is more effective than teacher delivered sessions at decreasing risky sexual behaviour. By integrating process and outcome data, the ability to interpret results according to empirical evidence was maximised (Oakley et al. [2006]).

With most (simple) interventions, integrity is defined as having the "dose" delivered at an optimal level and in the same way in each site. Complex intervention thinking defines integrity of interventions differently. The issue is to allow the form to be adapted while standardising the process and function (Hawe et al. [2004]). If standardisation is taken to mean that all the components of an intervention are the same in different sites, this treats a complex intervention as if it were a simple intervention. An alternative way of thinking about standardisation is that the fixed aspects of the intervention are the essential functions (the steps in the change process that the elements are purporting to facilitate or the key functions that they are meant to have) and the variable aspect is their form in different contexts. In this way an intervention evaluated in a pragmatic, effectiveness, or real world trial would not be defined haphazardly. In evaluations seeking to identify active ingredients within a complex intervention, strict standardisation may be required and controls put in place to limit variation in implementation, but some interventions are designed to be adapted to local circumstances. In complex interventions, the function and process of the intervention should be standardised not the components themselves. This allows the form to be tailored to local conditions and could improve effectiveness. Intervention integrity is then defined as evidence of fit with the theory or principles of the hypothesised change process. For example, in a trial of a school based intervention to promote health and wellbeing, schools were encouraged to use a standardised process to develop strategies which suited them rather than adopt a fixed curriculum, resulting in widely varied practice between schools (Hawe et al. [2004]). Reports of such studies should include a detailed description of the intervention to enable replication, evidence synthesis, and wider implementation.

Phase IV, the long-term implementation

The final phase is Phase IV which examines the implementation of the intervention into practice. Its purpose is to determine whether others can reliably replicate the intervention and results in uncontrolled settings over the long term. Long term follow-up may be needed to determine whether outcomes predicted by interim or surrogate measures do occur or whether short term changes persist. Although un-

common, such studies can be highly informative.

Summary

In summary, the process of developing and evaluating a complex intervention has several phases, which may not follow a linear sequence. Experimental designs are preferred to observational designs in most circumstances, but are not always practicable. Complex interventions may work best if tailored to local circumstances rather than being completely standardised. Understanding processes is important but does not replace evaluation of outcomes (Craig et al. [2008]). Trials of complex interventions are of increasing importance because of the drive to provide the most cost effective health care. Although such trials pose substantial challenges to investigators, the use of an iterative phased approach that harnesses qualitative and quantitative methods should lead to improved study design, execution, and generalisability of results (Campbell et al. [2000]).

Another definition

One of the more recent developments in the consideration of complex interventions in heath care is the greater use of the insights provided by the theory of complex adaptive systems (Shiell et al. [2008]). Shiell et al. state that although it is rarely delineated, complexity in healthcare has two meanings. In the first it is a property of the intervention, and in the second it is a property of the system in which the intervention is implemented. The first view of complexity, that complex intervention is "built up from a number of components, which may act both independently and inter-dependently" makes it hard to define the "active ingredients" and to be sure which component or combinations of components is more important. This is the more often used definition and the one used above.

The second view is that complexity is considered to be a property of a system not an intervention. A complex system is one that is adaptive to changes in its local environment, is composed of other complex systems (for example, the human body), and behaves in a non-linear fashion (change in outcome is not proportional to change in input, Shiell et al. [2008]). The idea that healthcare is a complex system is discussed in detail in Chapter 7. Whilst the active elements are subject to more variation, making evaluation of complex interventions difficult, economic evaluation is still possible by measuring the resources required and the intended outcomes seen. However, in a complex system it may be difficult to measure and evaluate the spin-off outcomes, such as a slow change in attitude which reaches a point where it creates

the context for change e.g. the ban on smoking in public places (Shiell et al. [2008]).

The first empirical evidence to support the argument that complex interventions are interventions in complex systems was provided by Burton (Burton [2012]). One testable property of complex systems is the presence of characteristic heavy-tailed statistical distributions. Such distributions appear to be ubiquitous in nature and have also been found in healthcare systems. These distributions are very different from the Gaussian (normal) distribution which characterises the distribution of simple effects. If complex interventions are truly interventions in complex systems the effect sizes of these interventions should be expected to show a heavy-tailed distribution, typical of those seen in other complex systems (Burton [2012]).

The effect sizes of a range of complex interventions collected and processed by the methodologically rigorous Cochrane review group was used to reduce the chance that the distribution of effect sizes found was due to the inclusion of methodologically weak studies with high risk of bias. When fitting these data, Burton found the effect sizes had heavy tailed distributions typical of those seen in classical complex systems, supporting the notion of complex interventions as interventions in complex systems (Burton [2012]).

Under either definition, complex interventions offer the possibility of tailoring interventions to patients, and methodological techniques for the analysis of data from complex interventions is needed. This thesis explores different methods for analysing data from two distinct complex interventions.

1.2 Aims of this research

Having established that health care interventions can be complex, the overarching aim of this research was to explore whether using non-linear models can better model health care and so capture the information contained in health care data sets of complex interventions. In this work, the term linear model was used in the usual statistical sense of a model that is linear in the parameters, so that logistic regression and the non-parametric Cox model for survival analysis are linear models. The non-linear model which was used as an exemplar was the artificial neural network model consisting of two layers of adaptive weights with full connectivity between inputs and hidden units and using sigmoidal activation functions (called a multi-layer perceptron). This choice was made because this particular architecture is capable of universal approximation, meaning that it can approximate any continuous function to arbitrary accuracy from a compact region of input space, provided the number of hidden units is sufficiently large, and provided the weights and biases are chosen

appropriately (Nabney [2002]). This means these models have the full capability required for modelling non-linear relationships between variables.

Two exemplar data sets were used, one from each of two common study types in health research, namely the randomised controlled trial and the observational cohort study. Each data set contains information about patients' responses to a complex intervention.

In order to achieve the aims of this research, one goal was to build on the population-level information provided by the standard analysis in the Back Skills Training Trial (BeST) of a cognitive behavioural approach to low back pain (Lamb et al. [2010b]) and investigate whether non-linear models enable individual tailoring to be possible for this treatment. i.e. to discover whether it is feasible to predict which patients will benefit from the treatment. If so, this has potential to inform clinical decision making when recommending treatment for an individual presenting with low back pain. The accuracy of predictions of both linear an non-linear models were compared to test the hypothesis that allowing for non-linear relationships among the explanatory variables leads to improvements in performance in modelling this complex system.

Another goal was to ascertain what predicts long-term survival after a cardiac event or procedure using data from a cardiovascular rehabilitation population. A linear and non-linear model was each fitted using the same explanatory variables and compared to ascertain which gave a better fit to the population survival curve. The two models were then be assessed for their performance to predict hazards for groups of people with the same values of the explanatory variables. By comparing plots that differ in just one covariate, interrelations between the covariates were hypothesised.

In order to achieve the research aim, a number of specific objectives were formulated:

- Can latent class analysis identify sub-sets of patients within a cohort of patients recruited to a clinical trial for non-specific low back pain?
- Is it possible to use linear models and the classes identified by the latent class analysis to tailor interventions for non-specific low back pain to patients?
- How do the performances of the linear and non-linear models compare for

the prediction of patient outcomes following a cognitive behavioural approach intervention for non-specific low back pain?

- What is the potential for these linear and non-linear models to inform policy on the treatment of non-specific low back pain?
- Using a linear model, what predicts long-term survival after a cardiac event or procedure, at population level?
- Can a non-linear artificial neural network be used to model long-term survival after a cardiac event or procedure?
- How do linear and non-linear models compare in modelling the survival and hazard of a population who have experienced a cardiac event or procedure?
- What is the potential for this research to inform policy for those who have experienced a cardiac event or procedure?

In Chapter 2, following, an overview of the literature relevant to the health care applications and the analysis methods employed in this thesis is given. Then the two data sets are introduced in Chapter 3, giving the context, collection and demographics of the cohorts, and a description of previous work using these data sets is given. Following this, the conventional methods, latent class analysis and survival analysis, are described in Chapter 4 and then the artificial neural networks methodology is introduced in Chapter 5. The results of using the linear models and the non-linear models for both data setts is given in Chapter 6 and followed by a discussion of results and the degree of success in achieving the objectives listed above in Chapter 7. The conclusions are drawn in Chapter 8. The appendix contains copies of the published and submitted research papers, diagnostic plots for the multiple imputation of missing data, and a comparison of the survival analysis with the data sets with missing data imputed.

Chapter 2

Literature Review

This literature review is divided into four sections, each being mirrored by a section in the results chapter. The first section is an overview of classification of patients for the allocation of treatments for back pain, and discusses a variety of approaches to classification. The second section explores the use of Artificial Neural Networks (ANNs) to classify patient outcomes. The next section gives and overview of what is known about long and short term survival in cardiac patients, focusing on the covariates that predict survival in those two time scales, and the final section explores the use of ANNs for survival analysis.
The purpose of this literature overview is to provide background for the research question and an indication of where the question fits with the existing literature. It also provides background for the analysis decisions and the interpretation of the results. A thorough search has been made for the latest reviews and studies, however this overview is not claimed to be a systematic review in the formal sense.

2.1 Classification for treatment allocation in back pain

There are a large number of treatments for back pain each of which provide discernible relief for a substantial proportion of patients. Almost everyone experiences non-specific low back pain (LBP) sometime during their lifetime (Dionne [1999]). Each year in the UK, about one third of the population experiences back pain of which 20% consult their Geb=neral practitioner(GP, family doctor. NICE [2009]). There are a number of interventions known to be effective for non-specific low back pain including exercise programmes, manual therapy, acupuncture (NICE [2009]) and a cognitive behavioural approach (Lamb et al. [2010b]). However, these interventions have small mean effects. This has prompted researchers to try to identify

ways of classifying those seeking treatment so that classes of patients can be matched to interventions in order to maximize treatment effect (Kent et al. [2010]). Many different classifications have been developed, including biopsychosocial classifications (McCarthy et al. [2004]), but there is little consensus on their use (Kent and Keating [2005]). For non-specific low back pain of greater than 6 weeks duration, the UK National Institute of Health and Clinical Excellence (NICE) has suggested patient preference should guide the choice of treatment from a range of effective interventions (NICE [2009]). The Latent Class Analysis (LCA) of the BeST data aimed to provide additional guidance for patients and clinicians as whether or not to consider the use of a cognitive behavioural approach (Lamb et al. [2010b], Barons et al. [2013a]).

In a recent systematic review of the role of clinical classification systems in chronic low back pain (Fairbank et al. [2011]), three types of classification are identified in the literature: those relying on clinical descriptor (Diagnostic), those describing prognosis (Prognostic), and those considering treatment response (Treatment-based); the latter is where the LCA of the BeST data belongs. In the review (ibid) they found 28 classification systems in use including five in the treatment-based category. In three of these, patients were classified by a clinician according to where the pain manifest and under what postural conditions, using the McKenzie method. This method was found to have high reliability, in that clinicians trained to use the method matched each others' categorisation in a high proportion of cases (70-100%). However, a 260-patient randomised contol trial (RCT) did not find significant reduction in pain from tailoring treatments based on McKenzie classification. The strength of evidence was classified as high in the review. The Sikorski method of classification had no reliability studies, and the Van Dillen and O'Sullivan methods had no clinical trial studies. The Canadian Back Institute Classification was based on location of pain and improvement or worsening with specific movements. It had good agreement (78%) in a reliability study. In a study with 1,356 patients given classification-tailored treatments and compared with 754 who had usual care, those having the individualised treatment regime specific to this classification system had decreased pain, improved function and less medication use. Evidence for the classification system was rated insufficient, and for the effectiveness, low because there was only one study reporting moderate examiner agreement. The review concluded that there is a need for a classification of low back pain which directs both surgical and non-surgical treatments. In all these studies, classification was done by the clinician and based wholly or in part on observations about pain location and change with movement.

The next section describes studies which did not appear in this review either because they were published later, or were excluded from the review because they are not clinical classifications but statistical partitioning of data.

Delitto's classification-based treatment approach (Hebert et al. [2011]) was compared with usual care in an RCT of 156 patients with sub-acute and chronic back pain, split 74 to classification and 82 to usual care (Apeldoorn et al. [2012]). Classification-based treatment was direction-specific exercise, spinal manipulation or stabilization exercises. Delitto's classification is also based on therapist observations, and the Apeldoorn study found no statistically significant differences between treatment groups. In an observational study, an association between cluster and subjective pain intensity was found in patients who had participated in a chronic pain-management programme (Chapman and Pemberton [1994]). 122 Patients were followed up 6-66 months later and classified into 7 groups using the Minnesota Multiphasic Personality Inventory (MMPI) using repeated measures analysis of variance (ANOVA). Clusters were also associated with occupation in univariate ANOVA.
In a cohort of 301 LBP patients, a hypervigilant sub-group was identified using a standardised clinical examination with good between-therapist agreement and used by physiotherapists (McCarthy et al. [2011]). This means of identification of distinct patient sub-type is accompanied by a suggestion that the hypervigilant might be targeted, but this was not tested. Distinguishing between neuropathic and non-neuropathic pain using the PainDETECT questionnaire with 145 patients in secondary care with 3 to 12 months of LBP did not predict treatment response to a usual care only intervention, but did predict prognosis - those patients with neuropathic pain had more pain, more limited activity and poorer self-rated health at baseline, 3 and 12 months (Morso et al. [2011]).

A variety of work has been published using statistical methods for partitioning data sets of patients with low back pain.

K-means clustering partitions cases into clusters in which each case belongs to the cluster with the nearest mean, with 'nearest' being defined by some distance metric (Hastie et al. [2009]). A k-means algorithm was used on samples of 127 male (57%) and 94 female patients with LBP to form clusters based on their responses to the Symptom Checklist 90 (SCL-90) which measures physical and psychiatric symptoms (Shutty Jr. and DeGood [1987]). The responses were scored in both the

standard 9-dimensions (S-Score) and in a modified 10-dimensions (F-Score) which was formed by an alternative division of response scales. The SCL-90 was administered at baseline and at follow-up 2 weeks to one month later and compliance to whatever treatment was recommended for the individual was recorded as a dichotomous variable. The optimal clustering using the S-Score was 3 clusters for each gender, whilst using the F-score the optimal was 3 clusters for males and 4 for females. The mean scores within each cluster on each of the score was reported. The clusters were numbered by size and 37% of males and 48% of females changed cluster number between scoring schemes. No links of cluster allocation to outcome were tested, and there were no reliability studies. In addition, 65% of the cohort was unemployed with 90% of those attributing their work status to the back pain.

A k-means algorithm was used to divide MMPI scores into 5 clusters for a cohort of 401 patients (60% female) with chronic LBP (McCreary [1985]). The intital clustering was performed on 271 cases, the remaining 130 reserved for cross-validation. Outcome was improvement in pain intensity during the previous week evaluated at baseline and follow-up 6 months to one year later (54% response rate). The same cluster characteristics were found in the cross-validation set. Treatment was a tailored combination of rest, exercises, advice and analgesic medication; the authors do not say how the tailoring differed. The ability to predict improvement in pain intensity from cluster membership was superior in males, and one particular cluster predicted both poor (below average) improvement in pain intensity for men and good (above average) improvement in pain intensity for women, although the authors concede that the small numbers in the separate-gender analyses may make the results unstable. They concluded that a larger percentage of men showed severe psychological disturbances and poor response to medical treatment.

The k-means clustering algorithm was used in an adapted form to learn a division of 21 patients into those who did and did not experience significant pain reduction of LBP using dynamic transcutaneous electrical nerve stimulation (DTENS) (Akhmadeeva et al. [2010]). The algorithm was trained on 74 clinical and paraclinical factors including health-related quality of life and the results of psychological testing. For an independent cohort of 35 patients who all experienced significant pain reduction, the algorithm was able to able to predict significant pain reduction for 70% and possible pain reduction for the remaining 30%. Since the algorithm was not tested on patients whose outcome was little no pain reduction, no assessment of its use in the general population can be made.

Clustering was performed on data from three separate cohorts of LBP pa-

tients (n=170, 82 & 124; Langworthy and Breen [1997]). A technique of sequential testing of squared Euclidean distance was employed to construct dendograms to identify clusters. (A dendogram is a tree diagram used to illustrate the arrangement of the clusters produced by hierarchical clustering.) Two clusters was found to be optimal in all three cohorts and these corresponded to patients with constant pain and patients with cyclic pain (i.e. that is worse at certain times of the day) in each cohort. None of the demographic features (age, sex, social class and job status) was association with cluster membership. The clusters were not used to inform treatment decision.

A longitudinal study of the time course of LBP collected monthly questionnaires about pain intensity from patients for 6 months and used LCA to divide their pain pathways into classes (Dunn et al. [2006]). Demographic information was collected at baseline. At baseline and 12-month follow-up, disability was measured using the Roland Morris Disability Questionnaire (RMQ), HADS was used to identify clinical anxiety and depression and catastrophising was captured with the Coping Strategies Questionnaire. 342 participants returned at least 4 of the questionnaires and were included in the analysis. Four classes were identified: persistent mild (122 patients), recovering (104 patients), severe chronic (71 patients) and fluctuating (45 patients). At 12 months, patients in the persistent mild class had the same pain but improved psychological scores, the majority in the recovering group were pain free, more than half of those in the severe chronic class were off work, and the patients classified as fluctuating were still consulting their practitioner and a persistent one third of those were depressed. There was no investigation of a link between class membership and treatment response, nor of whether patients could be allocated on baseline characteristics alone, in order to inform treatments. Quality of life (QoL) as measured by the SF36 was significantly lower for patients in 2 of 4 latent classes when 1,391 17-year-olds were classified using LCA (Beales et al. [2012]). LBP was assessed using the Nordic questionnaire for musculoskeletal symptoms, and general linear models used to associate class with quality of life. Female patients having LBP, depression and anxiety and male patients having LBP and behavioural disorders had lower QoL than others with or without LBP but without significant comorbidity. The Subgroups for Targeted Treatment (STarT) back screening tool was developed (Hill et al. [2008]) based on referred leg pain, comorbid pain, disability (2 items), bothersomeness, catastrophizing, fear, anxiety, and depression. The tool demonstrated good reliability and validity and was acceptable to patients and clinicians. Patients were classified as low risk, medium risk and (those scoring high on a psychosocial subscale) high risk of persistent disability. STarT was trialled with 851 patients as-

signed randomly 2:1 to intervention and control (Hill et al. [2011]). In the control group, treatment decisions were made on the basis of the physiotherapists' clinical judgment, without knowledge of a participant's STarT Back Tool classification. In the intervention group, the high risk patients received support for psychosocial barriers to recovery in addition to the physiotherapy for symptoms and function received by the medium risk group. All groups in the intervention received advice, a copy of The Back Book, information about exercise facilities and saw a a 15-min educational video entitled 'Get Back Active'. There was a larger improvement in RMQ scores in the intervention group at 4 and 12 months, and the cost was lower (Whitehurst et al. [2012]). The tool is published on the web and is free to download. Some practitioners in the UK and elsewhere have declared themselves to be users and describe the treatment they use. (STarT [accessed 18/05/13])

Many studies have used classification techniques to distinguish between categories of patients with low back pain, and a few have used such classifications to inform treatment regimens. In the few cases where an attempt was made to associate classification with outcome, only pain intensity was considered, and not back pain disability as measured by tools like the RMQ. In the one case where RMQ was used as an outcome, a new back screening tool was being tested and the treatment was related to items that could potentially be modified by treatment options in primary care. Very few studies have used classifications defined solely by patient questionnaires and just one has used the factors interventions may tackle as the basis for classification. The conclusion of the 2011 systematic review largely still stands; there continues to be a need for classification which directs treatments. In particular, research that increases understanding of the complex interaction within people, within complex interventions and between people and the interventions, would be very beneficial, particularly in decision support, design of interventions and tailoring interventions to patients.

2.2 Non-linear relationships and machine learning for back pain

One of the motivations for selecting ANNs to classify LBP data is that LBP has been found to have non-linear relationships with covariates. ANNs have the facility to capture non-linear relationships between variables. In this section, the evidence for non-linear relationships between LBP and covariates is illustrated, and this is followed by the background for the use of ANNs to model LBP.

2.2.1 Non-linear relationships

Previous studies on non-specific low back pain have found non-linear relationships between physical activity and back pain.

In a cross-sectional survey of employees in two Swiss national companies undertaken between 1996 and 1998, frequency of physical activity per week during leisure time had a significant but non-linear relationship to back pain (Lee et al. [2005]). Participation was 41% of all employees giving 10,321 participants (6,251 = 60.6% male, mean age 39.9 ± 10.8 years and 4,070 = 39.4% female, mean age 37.9 ± 11.3 years). Using both a self-administered questionnaire and physical fitness tests, this study assessed participants' general constitution, physical problems, back pain, work environment, stress, smoking habits and physical activity. Participants were asked whether they performed physical activity (running, cycling, aerobics, etc.) in their leisure time at least once a week and on how many days a week. The number of days (one to seven times per week) was taken as an indicator for physical activity. A fitness test was carried out under the supervision of doctors and nurses who visited the various branches of both companies and assessed endurance, flexibility, upper body strength and abdominal musculature, which was then compared with the normal range of the healthy population for those age groups. 4,945 (48%) reported having had mild back pain in the previous 4 weeks, and 696 (7%) reported having suffered from severe back pain. The severe back pain group comprised 340 (48.9%) women and 356 (51.1%) men; the percentage of men was statistically significantly higher. No association between body mass index (BMI: weight in kilogrammes divided by (height in metres)^2) and back pain was found. In univariate analyses, flexibility, upper body strength, abdominal strength, age (especially over 55 years), smoking category, and the sum of the stress factors were statistically significant. Stress factor scores for smokers, non-smokers and ex-smokers were significantly different. Frequency of physical activity per week had a significant but non-linear relationship to back pain: no or little activity (none to two times a week), but also intensive physical activity (six to seven times a week), was associated with back pain. However, moderate activity (three to four times a week) was associated with less back pain (χ^2=30.86, P<0.001). A similar relationship was found for intense back pain and physical activity (χ^2=31.78, P<0.001). In multivariate analyses, back pain prevalence of any intensity decreased with increasing age and a weak upper body and personal stress increased the likelihood of back pain, as did smoking and ex-smoker status. The variables gender, obesity, strength of the abdominal musculature and frequency of physical activity were insignificant in multivariate analysis.

A previous, similar workplace cross-sectional study had uncovered an interaction between ergonomic work variables and poor psychosocial work environment (a composite of workload, work content and social support, Linton [1990]). In this study, a relationship between regular exercise and back pain was not found, but back pain was not categorised or clearly defined by minimum pain intensity or duration, and the authors admit that this may have masked the effect. Heavy lifting, monotonous work, uncomfortable posture, vibration, age, workload, work content and social support was each associated with low back pain. In addition, an interaction between psychosocial environment and each of monotonous work, lifting and posture was found to be of greater significance than the constituent variables alone. Furthermore, the relationship between age and psychosocial factors was found to be nonlinear, with the under-30 and over-50 age groups more strongly affected by psychosocial environment than the intermediate age groups.
EuroQol (EQ-5D) measures health-related quality of life and is composed of 5 dimensions, each with 3 levels and a visual analogue scale. To account for the differences between national characteristics, the scoring of the 5 dimensions is calibrated by calculating specific national coefficients in the linear model that relates the dimension levels with the visual analogue scale to produce a single-number summary. Using a cohort of 633 patients with low back pain, the Spanish version of EQ-5D was shown to give statistically different model coefficients for those with LBP than the general population (Zamora et al. [2007]). It was also shown that a non-linear model (in which the regression coefficients noticeably indicated the presence of non-linear effects in the different dimensions of the EQ-5D instrument) gave a more accurate prediction of the visual analogue scale from the scores of the 5-dimensions.

In addition to the non-linear relationships between physical activity and back pain, previous studies on non-specific LBP have found non-linear relationships between anxiety and pain (Beesdo et al. [2009]) and between anxiety and depression (Katerndahl [2009]). A systematic review (Ramond et al. [2011]) found that LBP is a complex condition and that depression, psychological distress, passive coping strategies and fear-avoidance beliefs were psychosocial factors sometimes found to be independently linked with transition from acute to chronic non-specific LBP in the adult general population, whereas most social and socio-occupational factors were not. The authors concluded that psychosocial factors should be considered as momentary and partial indicators of more complex and dynamic distress, which required tailored management. Therefore, appropriate treatments for back pain are of great importance.

2.2.2 Machine learning for back pain

Artificial Neural Networks (ANNs) belong to a class of learning methods that developed separately in the distinct fields of statistics and artificial intelligence, based on essentially identical models and are described in detail on page 89. They can be represented by a network diagram like Figure 5.1, and their architecture is described by the number of layers, and the number of units in each layer e.g., 6-3-2 is a 3 layer network with 6 input units (explanatory variables), 3 hidden units (units in the single hidden layer) and 2 output units (response variables). The transfer or activation function is the function applied to the weighted sum of the inputs to a unit to provide the output of the layer.

Whilst clinical applications of ANNs are widespread (Dybowski and Gant [2001]), the literature on using ANNs to model LBP data is sparse. A 3-layer feed-forward ANN with sigmoidal transfer functions (242-20-3 architecture) was used to classify patients into healthy, LBP sufferers and malingerers (Gioftsos and Grieve [1996]). The training data was derived from 36 people, a similar number from each category, with malingerers represented by actors pretending to have LBP. Each study participant's movements were video recorded from the side whilst they stood up from and sat down on a chair. The changes in flexion-extension angles of the hip, knee and lumbar spine were captured electronically and the horizontal and vertical foot pressure measured with a force plate in front of the chair. Five episodes of sitting and standing were captured and the normalised pressures and angles used as inputs to the ANN. The root means squared error was the loss function and leave one out cross validation was used to evaluate the performance of the trained network, giving 31 (86%) correct. The video recordings were used to allow 9 physiotherapists to classify the same participants into the 3 categories. They found it easy to discriminate between the normal and abnormal patterns, despite the fact that they had not been trained to assess patients by watching them on monitors, but hard to distinguish LBP cases from malingerers. Each physiotherapist's accuracy was statistically significantly lower than the ANN. Another study used a 3-layer ANN with sigmoidal activation (or transfer) functions (3-3-1 architecture) to classify patients into pain / no pain 6 months after hospitalisation (Hallner and Hasenbring [2004]). The training data was derived from 71 people who were treated conservatively or by surgery and had their pain intensity measured at the beginning of hospitalisation and 6 months later. The number of neurons in the hidden layer were reduced until the accuracy began to decline. Inputs to the ANN were psychosocial variables, with the three with highest predictive power used in the final version (depression, suppressive

behaviour, thoughts of suppression). Accuracy was 83% with sensitivity 78% and specificty 97%, with no reported validation. The authors conclude that this could aid the early detection of risks for chronicity which could lead to improved cognitive, behavioural and emotional management of patients to aid the avoidance of chronicity. Another 3-layer ANN (378-n-5, n not given) was used to classify patients referred to surgical spine clinic into one of 5 diagnoses using pain drawings used routinely as a diagnostic tool (Sanders and Mann [2000]). Types of pain were marked on a human outline and a clinician examined the patterns to see if they follow a dermatomal pattern, a segmented field of skin innervated by a spinal nerve. The drawings were digitised using a video camera and segmented into 85 regions derived from low back pain dermatomes and gross anatomical regions. The number of pixels with a pain mark was normalised by the number of pixels in the region. Training was carried out on 200 samples and testing on 50, and overfitting avoided by early stopping. The ANN sensitivity was 49% which was statistically equivalent to the physicians (51%) and was a statistically significantly improvement over discriminant analysis (46%). The authors conjecture that low number of cases and an inadequate image capture might explain the disappointing sensitivity.

ANNs have been employed to aid diagnosis of genuine back pain, to assess risk of chronicity and to diagnose the origin or type of back pain. No literature has been found that uses ANNs to predict the outcome of a specific treatment for back pain based on baseline measurements of the patient characteristics that the intervention was designed to tackle. Similarly, no literature has been found addressing the usefulness of ANN prediction of outcome to tailor treatments to patients.

2.3 Cardiovascular Survival

Coronary heart disease is estimated to cost the UK economy £33bn (Euro40bn; $53bn) annually, of which just under half is due to direct health costs (Allender et al. [2008]). It remains the leading cause of mortality in the United Kingdom, with coronary heart disease accounting for 18% of all deaths in men and 13% in women (Scarborough et al. [2010]). The UK has been experiencing dramatic falls in death rates from coronary heart disease in recent years due to the fall in smoking prevalence (Hardoon et al. [2008]) and due to improvements in treatment, particularly secondary prevention (Unal et al. [2004], Hughes et al. [2011]).

There are a number of factors known to be associated with the development of cardiovascular disease (CVD). A recent study, based on the Framingham Offspring cohort, assessed the effect of risk factors measured at baseline on the long term (30-year) risk of developing hard cardiovascular disease (hard CVD is used to mean coronary death, myocardial infarction and fatal and non-fatal stroke) in those free of CVD and cancer at baseline examined using Cox regression (Pencina et al. [2009]). In a secondary model full CVD was used as outcome. Considering the extensive length of follow-up and the potential bias due to the competing risk of non-cardiovascular mortality in the prediction of long-term risk, the estimates were adjusted for the competing risk of non-CVD mortality as those who die of non-cardiovascular causes are ineligible for development of CVD events. Standard CVD risk factors (male sex, age, systolic blood pressure, anti-hypertensive treatment, total and HDL cholesterol, smoking and diabetes) were highly significant in the multivariate model. BMI was weakly significant in the final model, but in a simplified office-based risk model in which BMI replaced the lipids it was highly significant along with all other risk factors. In time-dependent analysis updating all variables approximately every 4 years, all standard risk factors remained significantly related to the hard CVD outcome with hazard ratios similar to those obtained in 30-year risk models. BMI lost its entire impact in time-dependent model. The authors conclude that this finding illustrates how the effect of BMI is mediated through other risk factors: it was present in 30-year risk model when the follow-up is extended for a long period from the baseline but then it impacted the individual risk factors, and after controlling for this impact in time-updated models, BMI lost its significance.

2.3.1 Primary prevention

Primary prevention is preventing illness in someone who does not currently have the illness. There is no single cause for coronary heart disease, although a cluster of risk factors has been identified which make individuals more prone to developing it. Some are unalterable, such as being male or inheriting a family history, but other known risk factors such as smoking, depression, raised cholesterol and sedentary lifestyle can be modified.

Smoking cessation reduced risk for coronary heart disease substantially more than cholesterol lowering (Critchley and Capewell [2003]). The evidence was sufficient to infer a causal relationship between smoking and coronary heart disease (Surgeon General [2004]) and 11,500 deaths from ischaemic heart disease among those over 65 in England were estimated to be smoking attributable (Twigg et al. [2004]). IMPACT, a widely used and replicated epidemiological model was used to synthe-

sise estimates stratified by age, gender, and area deprivation quintiles for the English population aged 25 and older between 2000 and 2007. IMPACT is an epidemiological model used to explain the contributions of population-level risk factor changes (incidence reduction) and uptake of evidence-based treatments (case fatality reduction) to the change in CHD deaths between two points in time. This model suggested that approximately half the recent CHD mortality fall in England was attributable to improved treatment uptake. This benefit occurred evenly across all social groups. The single largest contribution to the overall CHD mortality decrease came from a population fall in systolic blood pressure amongst those not on hypertensive medications with relatively small gains from hypertension therapies. Furthermore, moderate declines in smoking levels were actually greater in deprived areas (Bajekal et al. [2012]).
A recent Cochrane review into the reduction and alteration of fat in the diet (Hooper et al. [2012]) concluded that total mortality and cardiovascular mortality were unaffected by reduction of fat in the diet; there was a small reduction in cardiovascular risk on modification of dietary fats but not reduction in total fats.
The most successful CVD primary prevention strategies have been reduction in smoking and blood pressure and uptake of medications.

2.3.2 Secondary prevention

Secondary prevention is used to mean methods to diagnose and treat existent disease in early stages before it causes significant morbidity.
While many studies have reported on the effects of treatment on short term case fatality (Gale et al. [2008], Vale et al. [2011b], Greenhalgh et al. [2010], Nordmann et al. [2005], Elfstrom et al. [2012], Mikhail [2005], Kodama et al. [2009], Singh [2003]) there is much less evidence published on factors associated with the long term survival (greater than 5 years) of individuals who have experienced a coronary event (e.g. myocardial infarction, MI, heart attack) or procedure (e.g. coronary artery bypass graft, CABG, percutaneous coronary intervention, PCI) (Grundtvig [2012], Grundtvig et al. [2011], Hannan [2012], Kavanagh et al. [2002]) and predictors of long- and short-term survival differ (Filardo et al. [2012], Fox et al. [2008], Shahian et al. [2012]). This review investigates what is know about risk factors for short and long term survival for such individuals.

Long term (5-year) prognosis after hospital admission for MI is improving (Dudas et al. [2012]). Preoperative risk factors, which are good predictors of short-term outcomes, have been found to contribute little information to the prediction of long-term survival in CABG patients (Filardo et al. [2012]). Traditional predictors

of early survival in CABG patients over the age of 65 do not affect long-term survival, but late mortality is increasingly associated with chronic diseases and health behaviours (Shahian et al. [2012]). Patients who were defined as having limited functional status before undergoing CABG did not have a different long-term all-cause mortality from those who did not (Cervera et al. [2012]). Patients undergoing CABG or valve surgery and who have prolonged stay in intensive care also have poorer long-term survival (Elfstrom et al. [2012]). The change in definition of MI which was introduced in 2000 (Antman et al. [2000], Thygesen et al. [2012]), diagnosed many more patients with MI than under the previous definition. However, the long-term survival of patients diagnosed with MI under either scheme is not significantly different (Grundtvig [2012]). Fitness as measured peak oxygen intake, VO_2 max, (most accurate measure of exercise capacity) was found to be a predictor of long-term mortality of both men and women; even moderate fitness conferred a 50% reduction in cardiac mortality (Kavanagh et al. [2002], Kavanagh et al. [2003]).

Statins are a class of drugs used to lower cholesterol levels by inhibiting the enzyme HMG-CoA reductase, which plays a central role in the production of cholesterol in the liver. Statins have gained a pivotal role in the primary and secondary prevention of coronary artery disease, and are thought to improve perioperative outcomes in patients undergoing cardiac surgery. A recent Cochrane review (Liakopoulos et al. [2012]) identified eleven randomised controlled trials with a total of 984 patients undergoing cardiac surgical procedures, and found that preoperative statins reduced postoperative atrial fibrillation and was associated with a shorter stay in the intensive care unit. Statins failed to influence short-term perioperative mortality and there was no reduction in MI. Early statin treatment (initiation within 14 days) for patients with acute coronary syndrome (ACS) reduced the occurrence of unstable angina at 4 months following ACS, but did not reduce death, MI or stroke up to 4 months following ACS (Vale et al. [2011a]).
Beta blockers (BB) and angiotensin-converting-enzyme inhibitor (ACE inhibitor) are used in patients experiencing acute cardiovascular events. The effect of early treatment (within 24 hours of event) with these and other drugs on short-term mortality was investigated in a Cochrane review (Perez et al. [2009]). Whilst Nitrates reduced all-cause mortality during the first two days following acute MI, ACE inhibitors had a significant effect only at 10 days and Beta blockers and calcium channel blockers had no effect on mortality at 2, 10 or 30 days. There was no information on their effect on mortality risk at 5 years or beyond. In a systematic review of secondary prevention of ischaemic cardiac events, survival benefits of BB were similar

in men and women. The highest absolute benefit from beta-blockers was found in people who were over 50 years of age, with a higher heart rate at study entry, with a history of MI, angina pectoris, hypertension, or treatment with digitalis, and with transient signs or symptoms of mechanical or electrical failure in the early phases of MI (Skinner and Cooper [2011]). The same review advised that ACE inhibitors should be considered for secondary prevention in all patients after acute MI with left ventricular systolic dysfunction and in high-risk patients without left ventricular systolic dysfunction. It further suggested that after CABG, ACE inhibitors should be initiated with caution in low-risk patients without left ventricular dysfunction. The benefits of ACE inhibitors in very low-risk patients with a previous MI and without left ventricular dysfunction may warrant further investigation. If this advice is followed, those taking ACE inhibitors are those at increased risk.

2.3.3 Cardiac rehabilitation

Following a cardiac event or procedure, patients are routinely referred to Cardiac rehabilitation (CR, Bethell et al. [2009]). Cardiac rehabilitation is the process by which patients with cardiac disease are encouraged and supported to achieve and maintain optimal physical and psychosocial health and is a complex intervention consisting of three core elements: education, exercise training and psychological support (Heran et al. [2011], Jolliffe et al. [2009], Bethell and Mullee [1990], Bethell et al. [1983], Bethell et al. [2009], Plüss et al. [2008], Taylor et al. [2004], Turner et al. [2003], Turner et al. [2002], Unal et al. [2004], West et al. [2011]). In this section, literature describing what is known about the various elements of cardiac rehabilitation, especially long-term effects, is discussed.

The education element of cardiac rehabilitation

A recent review of the education element of CR concluded that there is some evidence to suggest that education may improve health-related quality of life, and that it may reduce costs by reducing subsequent use of health care, but there was no strong evidence of an effect of education on all-cause mortality, cardiac morbidity or hospitalisation (Brown et al. [2011]).

Depression and anxiety in psychological rehabilitation

Psychological interventions showed small to moderate effects on depression and anxiety following a cardiac event in a Cochrane review which also found that in smaller

studies (with some evidence of small-study bias) there was a modest positive effect of psychological interventions on cardiac mortality (Whalley et al. [2011]). Overall, there was no strong evidence that psychological intervention reduced total deaths, or risks of revascularisation and non-fatal infarction. Relative to placebo, antidepressants produced no change in cardiovascular function in heart disease, in respiratory function in lung disease, or in vital signs or laboratory tests in cancer, although were significantly likely to improve depression (Gill and Hatcher [2000]). Depression 7 days after MI has been associated with 18-month cardiac mortality (Frasure-Smith et al. [1995]) but only medical history related to heart disease and was controlled for in this study. One meta-analysis of observational studies concluded that depression is more strongly associated with short-term than long-term mortality (5 years and longer) and that more studies of long term mortality are needed (Barth et al. [2004]). A later observational study of 588 followed up for up to 8 years after MI found no association between depression and cardiac mortality, whether measured just before MI or at 12 months follow-up (Dickens et al. [2007]). Fitness is known to affect depression (Turner [2007]) but was not measured in this study. Another meta-analysis of prospective studies found that the presence of depressive symptoms after MI was not uncommon and was associated with a 2- to 2.5-fold increased risk of impaired cardiovascular outcome within 2 years (van Melle et al. [2004]). The authors state that confounding factors are possible and a causal inference cannot be drawn. A subsequent prospective study employed a case-control design and assessed depression using a structured interview (i.e. met diagnostic criteria for either major depression, minor depression, or dysthymia on the Depression Interview and Structured Hamilton) in contrast to patient questionnaires (e.g. HADS). This study found that depression was an independent risk factor for death 5 years after an acute MI and minor depression was associated with an increased risk of death, although the authors stated it was not known whether treating depression can improve survival or not (Carney et al. [2008]). Again only medical items directly associated with MI were controlled for, and fitness was not measured. A systematic review of depression screening tools in CVD patients found that sensitivity ranged from 39% to 100% (median, 84%) and specificity ranged from 58% to 94% (median, 79%). The screening tools have a known high false positive rate and a clinical interview is necessary to confirm a diagnosis of depression. Additionally, this systematic review found depression treatment with medication or cognitive behavioral therapy resulted in modest reductions in depressive symptoms in those with CVD, and there was no evidence that depression treatment improved cardiac outcomes (Thombs et al. [2008]). The failure of psychological interventions to reduce mortality suggested the association

is not causal (Dickens et al. [2007]).

A LCA of the course of depression symptoms found that there is a class of patients whose risk of a new cardiovascular risk is raised (Kaptein et al. [2006]). Scores on the Beck Depression Inventory for 475 patients (385 men [80%], 90 women, mean age 60.6 years) who had MI were assessed during hospitalisation and at 3, 6, and 12 months. The prevalence of depressive symptoms was 22.7% during hospitalization. The analysis identified 5 classes: no depressive symptoms (56.4%), mild depressive symptoms (25.7%), moderate and increasing depressive symptoms (9.3%), significant but decreasing depressive symptoms (4.6%), and significant and increasing depressive symptoms (4.0%). Participants in this last class had a statistically significantly higher risk of a new cardiovascular event compared with those without depressive symptoms. Controlling for baseline cardiac status and sociodemographic data did not alter the association; fitness was not measured and there was no long-term folow-up.
Depression predicted failure to complete cardiac rehabilitation (Casey et al. [2008], Turner [2007]).

Exercise in cardiac rehabilitation

Exercise-based cardiac rehabilitation has been found to reduce overall mortality, cardiovascular mortality and hospital admissions in the short and medium term. The vast majority of studies have less than two years follow-up, and only three had longer than five years follow up. These three studies are described in this section.
The National Exercise and Heart Disease Project examined whether a supervised exercise programme improved 19-year survival in 30- to 64-year-old male MI patients (Dorn et al. [1999]). This 3-year multi-centre randomised clinical trial was conducted in the United States from 1976 to 1979 and involved 651 men with neither hypertension nor significant comorbidity and with a ability to exercise at an intensity level $\geq$ 3 metabolic equivalents (METs) and a supine resting diastolic blood pressure <100 mm Hg. After completion of a 6-week, low level exercise program run-in period (those failing to complete were excluded), the men were randomly assigned to the exercise-treatment group (n=323) or non-exercising control group (n=328). The exercise prescription consisted of brisk physical activity in the laboratory for 8 weeks, exercising 1 hour per day, 3 days per week. Thereafter activities consisting of 15 minutes of continuous jogging, cycling, or swimming, were followed by 25 minutes of recreational games until the end of the study period. The duration of individual programs ranged from 6 weeks to 48 months, and follow-up periods ranged from 24

to 60 months. The research found that the exercise group had non-significantly reduced mortality risks early in the follow-up period. Benefits diminished as time since participation increased, which suggested that the protective mechanisms associated with the program may be short term. Analyses were performed with intention-to-treat and by the end of 2 years, 23% of the treatment group had stopped attending sessions and did not report exercising elsewhere, whereas 31% of the control group reported they were exercising regularly. Contamination between groups over time could also explain the diminished effects, because increased maximal physical work capacity (as determined by exercise testing) provided survival benefits up to 19 years.

In a study of 95 consecutive men aged 49 to 60 years suffering a first, recent MI study, La Rovere *et al.* report the impact on survival, during a 10-year follow-up period, of producing an exercise-induced increase in baroreflex sensitivity (BRS, causing changes in heart rate to maintain blood pressure) of ≥ 3 ms/mm Hg. The men were were randomly assigned to a 4-week endurance training period (n=49) or to no training (n=46). The exercise sessions were 30 minutes, 5 times a week for a 4-week training period and consisted of calisthenics and stationary bicycle ergometry. Exclusion criteria were atrial fibrillation or abnormal sinus node function, insulin-dependent diabetes, exercise-induced myocardial ischemia, and hypertension. There was a marginally statistically significant reduction in cardiac deaths in the trained group. However, those within both groups who had significantly improved BRS also had statistically significantly lower mortality; an improvement in exercise capacity not paralleled by significant changes in BRS was not accompanied by a better prognosis (La Rovere et al. [2002]).

The Alton cardiac rehabilitation trial, an exercise-only trial, randomized 200 men who had suffered acute MI to either a 3 month supervised exercise course or an advice only group and followed them up for 11 years (Bethell et al. [1999]). The included patients were age 65 or under and were recruited between December 1979 and March 1984 and were excluded if they lived more than 25 miles from Alton, if they had medical or orthopaedic problems that precluded their taking part in the exercise course, if they had insulin dependent diabetes mellitus or were in atrial fibrillation, if they had previously been through the course, or if they were on the investigator's personal general practice list (Bethell and Mullee [1990]). The follow-up questionnaire response rate fell from approximately 90% of survivors at year 1 to 72% at year 11. There had been 54 exclusions and 28 pre-randomization deaths among the original 311 patients and a further 12 died and 17 were excluded between

randomization and the first exercise test, leaving 200 for the study. Eighteen were excluded on medical grounds and these would have included some sicker patients with poor prognoses. There was no significant difference between the two groups for non-fatal reinfarction nor for long-term mortality. It did not reduce the risk of total MI, CABG or PCI. One or more previous myocardial infarcts, increasing age and low initial fitness were among the significant predictors of cardiac death.

All these studies included only men without significant comorbidity and under the age of 65 who do not fully reflect the typical characteristics of patient population who have experienced a cardiac event or procedure.

Efficacy of cardiac rehabilitation

A recent Cochrane review (Heran et al. [2011]) analysed 47 studies randomizing 10,794 patients to exercise-based cardiac rehabilitation or usual care. It found that cardiac rehabilitation did not reduce the risk of total MI, CABG or PTCA. Despite the inclusion of more recent trials, the population studied in this review was still predominantly male, middle aged and low risk. It concludes that well-designed, and adequately reported RCTs in groups of coronary heart disease (CHD) patients more representative of usual clinical practice are still needed.

There have been recent claims that the inclusion of historic trials has biased the Cochrane reviews in favour of cardiac rehabilitation (West et al. [2011]). Based on the RAMIT multi-centre RCT, West *et al.* suggested that there was no difference between patients referred to CR and those given usual care. The RAMIT study included 1,813 patients in England and Wales recruited between 1997 and 2000 with 903 allocated to comprehensive cardiac rehabilitation and 910 to usual care without cardiac rehabilitation. They found there was no significant difference between the two groups either at 2 years or at 7-9 years in all-cause mortality, morbidity, health service use, health-related quality of life (as measured by SF36), psychological general well-being and cardiovascular risk factors at 1 year. The evidence in favour of CR in the Cochrane review of 2000 is said to be heavily weighted by early trials when mortality was high. The improvement in mortality following CR is not seen when pooling just those trials in the Cochrane review that followed the WHO trial in 1983. This may be because the establishment of coronary care units, care regimens and secondary prevention medication have so altered the context of CR. This paper calls into question the value of cardiac rehabilitation as practiced in the UK at present. However, West acknowledges that extended cardiac rehabilitation training

(Plüss et al. [2011]), HF-ACTION Randomized Controlled Trial (O'Connor et al. [2009]), and GOSPEL Italian RCT (Giannuzzi et al. [2008]) may have benefits that current UK cardiac rehabilitation does not.

Plüss et al followed up for five years participants in a single-centre prospective randomised controlled trial of standard cardiac rehabilitation versus expanded cardiac rehabilitation (Plüss et al. [2011]). The expanded rehabilitation group experienced reduced cardiovascular events. Of the original 828 candidates, there were just 224 patients in the trial, split 113 to standard and 111 to expanded rehabilitation. The most common reason for exclusion was age over 75 years (308) followed by living outside the hospital catchment (101), loss to screening (25), severe co-morbidity (17) and language difficulties (15). The patients were recruited between May 1999 and May 2002 and all had recent acute MI or CABG. The expanded CR included all the elements of standard care (physical training, cardiologist counselling session, 'heart school' education, individual counselling as needed and smoking cessation) and in addition a five-day stay at the 'patient hotel' with physical training and information, 22 group stress management session over a year, and three cooking sessions with diet counselling. Most patients were taking secondary prevention medications (Beta-blockers, aspirin, statins, and ACE inhibitors). Follow-up was by national registry and so was nearly 100%. Cox proportional hazards models were compared, using the primary end points of cardiovascular mortality, acute MI, or readmission to hospital for CVD. The result of the study was that the expanded cardiac rehabilitation cohort had a hazard ratio (HR) of 0.47 for non-fatal MI and 0.76 for ischaemic stroke. The total number of hospitalization and the number of days of hospitalization were each significantly lower for the expanded CR group. This contrasted with the one-year follow-up which found no difference in biochemical risk markers or exercise performance between the two groups (Plüss et al. [2008]). Fitness was not measured.

The GOSPEL RCT investigated the value of a multi-factorial intervention consisting of one-to-one support held monthly from month 1 to month 6, then every 6 months for 3 years (Giannuzzi et al. [2008]). Each session consisted of 30 minutes of supervised aerobic exercise, plus lifestyle and risk factor counseling lasting at least 1 hour and reinforcement of preventive interventions lasting approximately 30 minutes. To improve adherence, a booklet was distributed and the support of family members was encouraged in ad hoc meetings. There were 3,778 patients recruited after MI between January 2001 and December 2002 randomised 1,620 to intervention and 1621 to usual care, after losses due to exclusion and loss to follow

up. Exclusion criteria were age over 75, poor short-term prognosis, disease limiting exercise and logistic difficulties. The targets of the intervention strategy were to give up smoking, adopt a healthy Mediterranean diet, increase physical activity up to at least 3 h/wk at 60% to 75% of the mean maximum heart rate, BMI of 25 or less, blood pressure of 140/85mm Hg or lower, total cholesterol level of 200 mg/dL or lower, low-density lipoprotein (LDL) cholesterol level lower than 100 mg/dL, blood glucose level of 110 mg/dL or lower, and haemoglobin A1c (HbA1c) level lower than 7.0% in subjects with diabetes. The intensive intervention decreased CV mortality plus nonfatal MI and stroke by 33% (95% CI, 0.47-0.95; P=0.02), cardiac death plus nonfatal MI by 36% (95% CI, 0.43-0.94; P=0.02), and nonfatal MI by 48% (95% CI, 0.31-0.86; P=0.01) with respect to usual care. Total mortality, sudden death, and total stroke decreased, although not significantly. The authors conclude that the GOSPEL intervention was effective in decreasing the risk of several important CV outcomes, particularly nonfatal MI, although the overall effect was small (Giannuzzi et al. [2008]). Fitness was not measured.

The HF-ACTION RCT was designed to investigate the safety and efficacy of exercise training for those with heart failure (O'Connor et al. [2009]). 2,331 patients were randomised to usual care or usual care plus aerobic exercise training, consisting of 36 supervised sessions followed by home-based training. The trial was delivered at 82 centres in 3 countries and follow-up was at clinic visits every 3 months for the first 2 years and yearly thereafter for up to 4 years. Cardiopulmonary exercise testing and a 6-minute walk test were performed at the 3-, 12-, and 24-month follow-up visits. Exclusion criteria included major comorbidities or limitations that could interfere with exercise training, recent or planned major cardiovascular events or procedures, performance of regular exercise training, or use of devices that limited the ability to achieve target heart rates. The 6-minute walk test (distance walked in 6 minutes) was also performed at the 3-year and final visits. Patients made their final visit at the end of the study follow-up period or at 4 years. To provide comparable levels of attention from study personnel in the 2 randomized arms, all patients were to be called and asked about their exercise every 2 weeks for the first 9 months, monthly until 24 months of follow-up, and quarterly thereafter. At all time points, approximately 30% or more of the patients in the exercise training group exercised at or above the target exercise minutes per week. In the exercise group, 37 patients had at least 1 hospitalization due to an event that occurred during or within 3 hours after exercise. In the usual care group, 22 patients had such a hospitalization, despite not undergoing a formal exercise programme. There was no significant difference

in the number of deaths (189 [16%] in the exercise group vs 198 [17%] in the usual care group). The authors concluded that exercise training resulted in nonsignificant reductions in the primary end point of all-cause mortality or hospitalization. Fitness was not measured.

Home-based cardiac rehabilitation is an attractive alternative to no cardiac rehabilitation for elderly patients (over 65) with coronary heart disease who decline participation in centre-based CR (Oerkild et al. [2012]). In an RCT, a physiotherapist made 2 home visits in a 6-week period in order to develop an individualised exercise programme that could be performed at home and surrounding outdoor area. Risk factor intervention, medical adjustment, physical and psychological assessments were offered at baseline and after 3, 6 and 12 months to both intervention and control groups. The primary outcome was 6 min walk test. Secondary outcomes were blood pressure, body composition, cholesterol profile, cessation of smoking, health-related quality of life, anxiety and depression. The study population was characterised by high age (median age 77 years, range 65 to 92 years) and high level of comorbidity. Patients receiving home-based CR had a significant increase in the 6 min walk test at 3 months, whilst the usual care group did not significantly improve, but with no significant differences between the groups. At 12 months follow-up, there was a similar decline in 6 min walk test in both groups (55.2m and 52.1m for intervention and control respectively). Participation in home-based CR improved exercise capacity among elderly patients with coronary heart disease, but there was no statistically significant difference between the home intervention and the control group. In addition, no statistically significant difference was found in the secondary outcomes and during the follow-up (mean 4.5 years), there was no difference in all-cause mortality. When intervention ceased, the initial increase in exercise capacity was rapidly lost. This cohort had little depression so that statistically significant improvement was not possible.

A systematic review of home-based CR interventions compared to usual care and centre-based supervised rehabilitation found that home-based cardiac rehabilitation for low-risk patients does not have significantly poorer outcomes compared to centre-based programmes. However, duration of follow-up is short and there are only limited data on mortality rates and from well conducted RCTs (Jolly et al. [2006]).

A recent narrative review of CR concludes that observational studies in the new millennium consistently conclude that CR does significantly reduce mortality

even in the context of improved medical and revascularization strategies (Dobson et al. [2012]). The authors suggest that advances in the quality of cardiac rehabilitation programmes and European-wide guidelines have led to improved survival.

Long-term survival after a cardiac event

In this section, studies investigating long-term survival ($\geq$ 5 years) after a cardiac event are discussed and the factors known to affect long-term survival identified.

In a recent study, long-term survival figures were evaluated based on EuroSCORE, a widely utilized as a pre-operative risk prediction tool (O'Boyle et al. [2012]). A number of the risk factors in EuroSCORE have previously been identified by Cox regression analysis as significant factors with regard to long-term survival, leading to some authors utilizing EuroSCORE as a possible predicator of long-term survival despite its derivation being based on in-hospital operative mortality. Only patients who had undergone an isolated CABG were included and long-term follow up was from 6 months to 12 years, mean 6.6 years. Cox regression analysis demonstrated that age, diabetes, ejection fraction, BMI, dialysis, logistic EuroSCORE, creatinine kinase myocardial isoenzyme (CKMB), left internal mammary artery (LIMA) usage and peripheral arterial disease were significant factors affecting long-term survival. Except for diabetes, BMI, CKMB and LIMA usage, all these variables were incorporated in the EuroSCORE, however their weighting will vary. The authors concluded that this explains why EuroSCORE was a reasonable approximation as a predictor for long-term survival, but the incidence of diabetes, CKMB and LIMA usage will affect its accuracy.

Non-high-density lipoprotein cholesterol (non-HDLC) can predict the risk of cardiovascular events among general population without coronary heart disease and may be a practical predictor of long-term cardiac death in patients with CHD after CABG (Fukushima et al. [2012]). A more active lifestyle is significantly associated with improved survival in elderly (over 70 years) CABG patients. The nonlinearity of the relation suggests that more sedentary patients could have the most benefit on survival by increasing their exercise lifestyle habits. The improved outcome is explained by both cardiac and overall mortality reduction (Rengo et al. [2010]). In a study of 1,158 (84%)men and 215 women aged 35 to 64 years, followed-up for 12 years for non-fatal and fatal CHD events and all-cause mortality, the overall survival of men who underwent CABG was similar to the survival of the corresponding background population for about ten years but started to worsen after that (Ketonen et al. [2008]). The CHD mortality of men who had undergone the operation

was clearly higher than in the background population. Among women, the all-cause mortality after CABG was about twice the expected mortality in the corresponding background population. In Cox proportional hazards models age, smoking, history of MI, BMI and diabetes were significant predictors of mortality. Fitness was not measured.

In a study of 10,268 patients, multivariate logistic regression analyses showed that underweight was associated with higher early mortality after CABG surgery (van Straten et al. [2010]). Multivariate Cox regression analyses revealed morbid obesity as an independent predictor of late mortality. A study to analyze the impact of varying BMI on early clinical outcome and long-term survival in a group of patients who underwent CABG and/or cardiac valve procedures found there was no association between BMI and hospital mortality in the entire patient population (Rahmanian et al. [2007]). Multivariate analysis revealed obesity as an independent predictor of hospital mortality in patients who underwent valve surgery. Obesity was associated with an increased risk for sternal infection , whereas underweight correlated with postoperative bleeding. Underweight was an independent predictor for decreased long-term survival. Fitness was not measured.

The Basingstoke and Alton study

There is a need for studies of survival after a cardiac event or procedure that that measure fitness and can identify the relative importance of fitness and BMI. Survival in these patients has been improved by secondary prevention and especially medication. There is some evidence that CR improves survival, especially programmes that include a significant exercise component. Few studies have followed up patients in the long term and very few have measured fitness.

2.4 Artificial Neural Networks for Survival

Artificial neural Networks (ANNs) have been widely used in clinical applications (Dybowski and Gant [2001]), predominantly in classification problems i.e. to assign the patient to one of a small set of classes based on their measured features. ANNs have recently been used for survival analysis, but most of the literature has been devoted to treating the problem as a classification one or making piecewise or discrete approximations of the hazard function (Bottaci et al. [1997], Tangri et al. [2008], Regnier-Coudert et al. [2012], Biglarian et al. [2013]). This literature review

will focus largely on use of ANNs for regression, where time is used as a continuous variable and a survival probability distribution is produced. The key contributions in this field have been made by Ruth Ripley, Rashmi Joshi and the collaboration group including Antoni Eleuteri, Azzam Taktak, Paulo Lisboa and Bertil Damato.

2.4.1 The continuous-time approach

The use of artificial neural networks in non-linear survival analysis began in 1992 when Ravdin and Clark used them to predict the survival probability of breast cancer patients, given the prognostic variables, using time as an input variable (Ravdin and Clark [1992]). Ravdin and Clark first introduced the idea that time to event or censoring should be coded as one of the prognostic variables. Their data set consisted of 1,373 breast cancer patients followed up for a minimum of 36 months. The prognostic variables included patient age and a number of medical measurements of the tumour. Some of the variables were transformed by taking the logarithm of the variable plus a constant in order to approximately equate the median and mean of the distributions. All prognostic variables were normalised to lie in $[-1, 1]$ for use with the *tanh* transfer (or activation) function. The data were split into training, validation and testing sets of size 500, 453 and 420 cases respectively. Time was divided into periods where 10% decrement in survival was seen, i.e. 12 months, 18 months, 27 months, 40 months and 60 months, by which time 50% of cases had died. These time intervals were coded 1 to 5 and predictions were made for survival at these time intervals. In order to correct the bias introduced by 231 of the cases within time 5 (maximum) having died in the interval and 45 having survived, the authors selected at random 45 of the 231 of the deceased cases for inclusion and deleted the remainder to match the 50% survival shown on the Kaplan-Meier estimate. The multi-layer perceptron (MLP) used had a 9-7-1 architecture and was optimised using the back-propagation algorithm (described on page 93). The weight vectors were stored at intervals during the training and the resulting network varaints were evaluated using the validation set. The optimal number of training iterations was determined by calculating the goodness of fit for the prognostic indices using the global χ^2 statistic of a Cox regression on the same transformed variables using the validation data set. The influence of the individual inputs was investigated by omitting each in turn and measuring the global χ^2 statistic. The patients predicted to have excellent and poor early and late prognosis by the network were identified and their Kaplan-Meier survival estimates used to confirm their qualitative categorisation. There was no attempt to produce confidence intervals, use alternative

learning rules or to optimise over network architectures. Although this paper used time in a piecewise approximation of the hazard, it was the first to demonstrate that using time as a prognostic variable, an ANN can make a set of predictions for given patients over time.

Brian and Ruth Ripley summarised the use of neural networks as non-linear statistical methods in survival analysis (Ripley and Ripley [2001]). They point out the pitfalls of using neural networks (overfitting, sensitivity and specificity) and that standard statistical models can be more robust in that they can be built up from simple models, with each stage of added complexity being tested for a significant improvement in fit. There is no analogue for this in fitting ANNs.
The performance of 7 different neural network models were compared using a data set of times to first relapse in days since surgery for 1,335 patients with primary ductal breast cancer (Ripley et al. [2004]). That study used a MLP and back propagation with a quasi-Newton method for optimisation. In three models time to first recurrence was categorical and in the remaining 4 it was continuous.
The first discrete time model predicted probability of relapse within 5 years using a standard classification neural network. The following 2 models estimated the probability of relapse in the time periods less that one year, one to two years, two to three years, three to five years, and greater than five years. The first of these two discrete time models ignored the ordering of the time periods and fitted the model using a softmax neural network (see page 94). The second discrete time model incorporated the ordinality of the outcomes, and was the first time such a method was used for survival analysis with ANNs. The continuous time models were three extensions of linear models (log-logistic, log-normal, and proportional hazards) and a time-varying extension of the proportional hazards model. The authors observed that use of back-propagation of errors was not possible in these cases as the log partial likelihood was not a sum over data examples. By ordering the patients in reverse order of time of relapse, the risk set always consisted exactly of the patient under consideration plus all those previously processed. This allowed the partial sums to be accumulated as the patients were processed in order and allowed for efficient programming of the exact derivatives and their use in standard optimization algorithm. The performance of the seven models was compared on a breast cancer data set of 1,335 women obtained from the Imperial Cancer Research Fund Medical Oncology Unit in Oxford. There were 680 cases with no missing data and missing values for the remainder were estimated using multiple regression on available variables. For patients censored before the end of the observation period, the authors

estimated the conditional probability that the censored patient would have survived to the end of the observation period conditional on the fact that they survived until their censoring time. 276 patients suffered relapse in the observation period (Ripley [1998]). Comparison of results with the same analysis using only the complete cases produced similar results. The basis for comparison of the models was the prediction of relapse within 5 years using the sum of the log probabilities, accuracy, sensitivity and specificity on the cross-validation data and based on a probability cut off of 0.5; five-fold cross validation was used since the data set was too small to split into a training and testing set. The size of the artificial neural networks' hidden layers and the value for the weight decay parameter were also assessed by cross validation. These tests showed that the binary model and the continuous time models did best, the ordinal model second best and the multiple time model least well. Accuracy of the best models was around 78%. The sensitivity (correct prediction of relapses) ranged between 31 and 41% and the specificity (correct prediction of non-relapses) 90-94%. The authors conclude that the improvement due to non-linearity is not very great in this case, and that the binary model is a useful system to predict risk of relapse in newly diagnosed patients.

Joshi and Reeves have reported work on artificial neural networks for malignant melanoma prognosis (Joshi et al. [2003], Joshi et al. [2005], Reeves and Johnston [2008] and Joshi [2004]). The data comprised 1,946 patients diagnosed with malignant melanoma between 1987 and 1996, of which 1,160 (60%)were female, 786 were male and 1,628 (84%) were living at follow up to 31st December 1999 (Joshi et al. [2005]). Cox survival model analysis showed violations of the proportional hazards assumption, so the data was stratified in the non-proportional variables, and a stratified Cox model was fitted. This identified 4 significant prognostic variables with between 2 and 5 categories, making 120 possible combinations. Since the ANN approach was to enter all prognostic variables into the network as binary variables, the 5-level variable was dropped leaving 24 combinations. These were the inputs to the ANN along with bias and scaled time as event time divided by maximum event time. The ANN had 3 hidden nodes; 5 nodes was known to produce superior results but was computationally intensive. The data was split into the 24 subsets each with a uniform distribution of responses to the 4 prognostic variables and the ANN weights were optimised by minimising the log-likelihood, with weight decay regularisation to avoid over-fitting. A monotonic survival function is ensured by constraining the weights relating to the time input from the hidden layer to the output are of different sign to the weights from the inputs to the hidden layer. The optimal brain

surgeon technique was used to assess the significance of each of the 34 weights and to prune the network to 24 weights. The Hessian of the log likelihood was required for this procedure, and could also be used to produce the large sample approximate variance-covariance matrix. The ANN model was assessed by plotting the survival probability for each of the 24 subsets against the equivalent Kaplan-Meier estimate, the stratified Cox model and a log-normal model and this showed that the ANN most closely matched the data. The 10-year survival probabilities for the Kaplan-Meier, Cox, ANN and another ANN optimised by minimising least squares were compared for each subset, and the sum of the absolute differences compared to the Kaplan-Meier. By this measure, the least squares ANN was most accurate, followed by the Cox model. The standard errors on the ANN weights were constructed using the $weight \pm 1.96 \times \sqrt{jj\ component\ of\ Fisher\ Information\ Matrix}$. Confidence intervals on the output of the ANN were produced using a bootstrap approach on the multivariate normal distribution simulated by Cholesky decomposition of the variance-covariance matrix of the optimised ANN.

2.4.2 A Bayesian approach

The collaboration group including Antoni Eleuteri, Azzam Taktak, Paulo Lisboa and Bertil Damato have produced a number of papers since 2003. In Eleuteri et al. [2003], a standard sigmoidal activation function was used in a multi-layer perceptron (MLP) for the covariates, except for time, which had instead a logarithmic activation function. The only input to this unit through the weight was the time unit; neither biases nor other inputs fed through this unit. The purpose of this was to ensure that the survival function at time zero was 1 and at $+\infty$ was zero. In order to ensure that the survivor function was non-increasing as time increased, they constrained the weights associated with the time input to be non-negative. Instead of a conventional approach to training which minimizes an error function with a single weight vector, the authors use a Bayesian scheme which considers a probability distribution over weights. For the unconstrained weights, a zero mean Gaussian prior was selected, whilst the positivity constraint weights had an Levy prior, which has positive support and leads to a stable, non-degenerate joint distribution of the weights out of the activation functions. This strategy was tested on data on 12 covariates from 1,776 patients with colon cancer. The data was split 700 for testing and 1,076 for training and the outputs for the ANN and a Cox model compared to the Kaplan-Meier estimate. The authors concluded that their novel ANN architecture performed as well as the Cox PH model.

The Bayesian approach to modelling survival with an ANN has been further developed by this group to include automatic relevance determination (PLANNARD, Taktak et al. [2006]), adjustment for skew and labeling of missing data (Lisboa et al. [2003]), modelling of the log of the hazard rate function (CHENN, Eleuteri et al. [2007a], Eleuteri et al. [2007b]), applications in Choroidal Melanoma (Damato et al. [2008]), colon cancer (Dolgobrodov et al. [2007]), Ocular melanoma (Taktak et al. [2008]), competing risks (Lisboa et al. [2009]), and model selection using Genetic Algorithms (Ambrogi et al. [2007]).

Artificial neural networks have been used in medical applications, often for decision support (Dybowski and Gant [2001]) including myocardial infarction detection (Haraldsson et al. [2004]), medical diagnosis (Kononenko [2001]), pathological staging of prostate cancer (Regnier-Coudert et al. [2012]). The use of ANNs to model survival is not widespread. The literature that exists has been overwhelmingly in the classification context, where survival for a specified time is a binary prediction. The use of survival time as one of the inputs during training, and the consequent ability to produce a survival probability distribution curve is much less frequently used, and has been used only on Cancer data sets. Whilst there have been exemplar hazards produced by these ANNs (Ripley et al. [2004], Lisboa et al. [2009]), producing hazard rates for specific subgroups of patients and comparing them with hazard rates produced by other survival models to generate hypotheses has not been done. Artificial neural network (ANN) modelling has been used to model time to relapse in cancer patients (Joshi et al. [2003], Joshi [2004], Ripley et al. [2004], Joshi et al. [2005]), but modelling the long-term survival of patients who have has a cardiac event or procedure with ANNs is new. There is a need to extend continuous time ANNs to survival on non-cancer data sets (Joshi [2004]). The ability of ANNs to produce hazard rates has not been exploited either by comparison to rates provided by other models, or to produce research hypotheses. This research has addressed these omissions, selecting the Joshi methodology in order to make a direct comparison.

The following chapter gives details of the two data sets used in this research and the work that has already been done using those data sets.

Chapter 3

Data

The purpose of this chapter is to give details of the two data sets that have been used in this thesis, and the analysis and results of the original study teams who collected them. The methods used in my re-analysis of the data for this research are detailed in chapters 4 and 5 (starting on pages 64 and 87 respectively) and the results of my analyses in the Results chapter starting on page 103.
The two data sets are: the Back Skills Training Trial, designed and undertaken by Professor Sallie Lamb and colleagues at Warwick Clinical Trials Unit, which is a multi-centred randomised controlled trial of a primary-care based cognitive behavioural program (a complex intervention) for low back pain; The Basingstoke and Alton Cardiovascular Rehabilitation data set is an observational study designed and undertaken by Dr. Sally Turner who collected and analysed it with a view to determining the role of depression and fitness on survival in a cohort attending a community cardiovascular rehabilitation programme (a complex intervention).

3.1 Introduction to the Back Skills Training Trial Data

3.1.1 Introduction

The Back Skills Training Trial (BeST) was a large, randomised controlled trial of a cognitive-behavioural approach to the relief of back pain, conducted by Warwick Trials Unit and published in *The Lancet* (Lamb et al. [2010b]). The motivation for the trial was the significant prevalence and cost to the economy of lower back pain coupled with evidence that psychological risk factors play an important role in the progression of lower back pain. Starting in 2000 there has been a major change in the management of lower back pain in primary care to active management which is advice to engage in physical activity and avoid bed rest, to take appropriate medication

and to take a positive attitude. This advice was summarised in a booklet called 'The Back Book', routinely handed to patients being treated for lower back pain. Cognitive behavioural approaches for back pain were first introduced in secondary care for chronic and severe lower back pain. Trials of the cognitive behavioural approach for subacute and chronic low back pain produced a mixed picture, and variable adherence to the principles of cognitive behavioural approach was hypothesised as an explanation. The aim of the BeST trial was to develop and test a group-based cognitive behavioural approach intervention for lower back pain that could be accessed from UK NHS primary care (Lamb et al. [2010a]).

3.1.2 Methods

A complex intervention is defined as an intervention that has many parts which may work independently or interdependently (Campbell et al. [2000]). The BeST trial compared active management (AM) with a complex intervention comprising active management plus group treatment using a cognitive behavioural approach (AM+CBA), and found that AM+CBA was effective in treating subacute and chronic low back pain (LBP) in both short and long term, and that it was cost-effective.
Training was provided to health care professionals in order to deliver both treatments effectively. To deliver the active management consistently, primary care nurses were given a 1-hour training session on the best practice for the management of LBP, and asked to cascade this within their practices. To deliver the active management plus cognitive behavioural management consistently, other nurses, physiotherapists, psychologists, and occupational therapists attended a 2-day course, supported with remote mentoring.

701 patients with at least moderately troublesome back pain of at least 6 weeks' duration were recruited from 56 general practices in 7 regions in the UK between April 2005 and April 2007. The patients were randomised 2:1 in favour of the cognitive behavioural approach arm of the trial. Originally, randomisation was to have been balanced 1:1 between the two treatment arms, but such a randomisation would have produced too few participants to run the cognitive behavioural approach groups, so it was switched to 2:1 randomisation in favour of the AM+CBA arm. Those in the control group received active management which is best practice care, including advice to keep active, to use painkillers, and take a positive outlook. They were also given a copy of 'The Back Book', which was designed by LBP experts to reinforce these messages. Those in the intervention group received this same treatment, but in addition were invited to attend a series of six cognitive behavioural

group sessions with about eight participants in each. In the group sessions, the topics of goal setting, pacing, challenging beliefs, managing pain, and improving communication with health professionals were covered. Compliance with treatment was defined as attendance at least three of the six sessions. Patients recruited to the trial had subacute and chronic lower back pain that was at least moderately troublesome and of at least 6 weeks duration and had attended general practice, were at least 18 years of age, and had not been managed previously in a cognitive behavioural programme.
Demographic and clinical data were collected at the pre-randomisation stage, including date of birth, sex, lower back pain symptoms in the last six weeks, frequency of pain in the last six weeks, ethnic origin, age when left full-time education, and employment details. The two primary outcomes were the Roland Morris Questionnaire (RMQ, Roland and Morris [1983], Roland and Fairbank [2000]) and the Modified Von Korff (MVK, Korff and Anthony [1982]) scale.

The Roland Morris Questionnaire (Roland and Fairbank [2000]) is the most widely used measure of lower back pain disability in primary care trials. It contains 24 yes / no responses relating to a range of functions commonly affected by lower back pain, and the score is the sum of the responses. There were concerns expressed by the trial report authors that the RMQ does not conform to the assumptions of scaling and normality of distribution that underpin its use in statistical analysis. It has been shown to be differentially sensitive at low, mid and high ranges, with better sensitivity in the mid-range (Lamb et al. [2010a]). Previous trials have adopted a clinically significant difference between groups of 2.5 RMQ points.

The Modified Von Korff Scale assesses two dimensions, pain and disability, associated with back pain in the last four weeks. It is made up of six items, each of which is rated on a scale from 0 (no pain or disability) to 10 (worst pain or disability). The first three items relate to disability and the interference of pain with daily activity, recreation and ability to work. The last three items relate to worst pain, average pain and rating of back pain today. RMQ is scored by summing the responses and so has a maximum score of 60 (Korff and Anthony [1982]).

Secondary outcomes were occupational disability and limited activity days, participant satisfaction, psychological and behavioural measures, quality of life (QoL), health economics (including quality-adjusted life year (QALY)) and resource use. The psychological and behavioural measures were included because they measure

constructs hypothesised to lie on the causal pathway of effect, and might provide some explanation as to why the treatment was or was not effective.

The Fear Avoidance Beliefs Questionnaire (FABQ) is a measure of the degree of fear of pain and disability, and the avoidance of physical activities that can result (Vlaeyen and Linton [2000]). Each item is scored on a Likert scale 0 to 6 and measures fear avoidance beliefs about work and fear avoidance beliefs about physical activity. The latter was selected for use as it has generic applicability. The sum score has a maximum of 24, with a higher score indicating a greater degree of fear and avoidance beliefs.

Pain self-efficacy (PSE) is a measure of a patient's confidence to carry out a range of activities despite their back pain. There are ten items on a Likert scale each with responses 0 to 6, sum score up to 60, with higher score indicating the higher attribute of good pain self-efficacy (Nicholas [2007]).

Quality of life is measured using the Short Form 12 (SF12) a measure of health related quality of life widely used in back pain trials. There is a manual for scoring the physical and mental components on a scale of 1 to 100 for each and designed to have a mean of 50 and a standard deviation of 10 in a representative sample of the U.S. population, so that a score greater than 50 represents above-average health status (Ware et al. [1996]).

The Hospital Anxiety and Depression Scale (HADS) is a self-report rating scale designed to measure both anxiety and depression. It consists of two sub scales, each containing seven items on a 4-point scale (ranging from 0-3). The HADS is scored by summing the ratings for the 14 items to yield a total score, and by summing the ratings for the 7 items of each sub scale to yield separate scores for anxiety and depression (Lisspers *et al*, 2007).

Troublesomeness was reported as a response to the question 'How troublesome has your back been during the past 6 weeks?' on a single likert scale, with options 'Not at all troublesome', 'Slightly troublesome', 'Moderately troublesome', 'very troublesome', and 'Extremely troublesome'. Those whose responses were 'Not at all troublesome' or 'Slightly troublesome' were not recruited to the trial.

3.1.3 Missing values

The data losses occurred because not all participants responded to every question, therefore some of the responses needed to calculate the scores were missing. The complete cases all match the trial HTA report (Lamb et al. [2010a]), where it reports them:
Fear Avoidance 662 complete cases, as on page 239 of HTA report
PSE 676 complete cases, as on page 240 of HTA report
RMQ baseline 700 complete cases, as on page 238 of HTA report
RMQ 12months 498 complete cases, as on page 238 of HTA report
SF12 687 complete cases, consistent with page 241 of HTA report (Complete cases for individual questions is not given)
Appendix 12 of HTA report does not include details on the following:
HADS anxiety 686 complete cases
HADS depression 693 complete cases
Troublesomeness 638 complete cases

Taking the complete cases only in a sequential manner we now arrive at 407 complete cases altogether; accumulated missingness is detailed in table 6.1 on page 107 in Chapter 6.
Where items making up the RMQ were missing, the mean was taken of the responses that were given and multiplied by 24, consistent with published research (Kent and Laurisden [2011]). For all other variables, a missing item response meant the variable was considered missing.

3.1.4 Previous results

Follow up was measured at 3, 6 and 12 months and outcomes were summarised as the change from baseline score. At 12 months, the mean change from baseline in the RMQ score was 1.1 (95% CI 0.39 to 1.72) in the control (AM) group and 2.4 (95% CI 1.89 to 2.84) in the AM+CBA group. The difference of 1.3 (95% CI 0.56 to 2.06) is statistically significant ($p = 0.01$), and can be delivered at low cost to the provider. 1,465 people of the 9,771 initially identified appeared to be eligible, and after excluding those who no longer had back pain, or who had only infrequent pain, or pain that was not troublesome, or were pregnant, 705 were randomised of whom four did not provide baseline data, leaving 701 for the trial analysis. Of these, 70% experienced pain every day, 55% had moderately troublesome pain and the remainder had very or extremely troublesome pain.

The average age of the trial participants was 54 years, 60% were women, most had left full-time education before age 15, and half were currently working, mostly full-time; most of those not working had retired. Stratified block randomisation was carried out by an independently administered telephone randomisation service at the MRC clinical trials unit in London.

There was no statistically significant difference in the follow-up questionnaire return rates between the two arms of the trial at any of the observation points nor any evidence of a systematic difference in the baseline characteristics of participants who provided follow-up data and those who did not. Attendance at all six of the cognitive behavioural sessions was achieved by 25% of the participants allocated to the AM+CBA arm, 63% attended at least three sessions and were considered to have received the basic elements (adhered), and just over 10% did not attend any. There were no differences between the patients who were adherent and those who were not in any of RMQ score, troublesomeness of pain, fear avoidance beliefs, sex or Modified Von Korff disability at baseline. Adherent patients had slightly lower MVK pain scores at baseline (mean difference 4.5, 95% CI 0.95 to 8.19) and were older (average 4.5 years older, 95% CI 1.8 to 7.5). There were 62 groups run with mean size of 8 (SD 1.62, range 4-12) and a range of times were offered to meet the needs of participants. There was no evidence of either group or therapist effects.

Improvements in the AM only arm were, on average, 1.1 RMQ points, with change occurring between baseline and three months and no further improvement thereafter. The change in the CBA arm was almost double by three months, and the treatment difference continued to widen at six and twelve month follow-up points. The mean treatment difference was 1.1 Roland Morris Questionnaire points at three months, 1.4 at six months and 1.3 at twelve months, and all were statistically significant.

The MVK disability scale showed improvement over both arms of the trial, with the improvement in the active management arm occurring between baseline and three months and declining thereafter, and the cognitive behavioural approach arm showing a greater improvement. The mean difference between the two arms was 4.3% at three months, 8.1% at six months, and 8.4% at twelve months. The MVK pain scale also showed improvement over both arms, with the active management arm improving gradually over the twelve months, and greater improvements

in the cognitive behavioural approach, particularly at three months. The difference between the two arms was 6.8% at three months, 8% at six months and 7% at twelve months, and all were statistically significant.

The SF12 physical subscale showed improvements in both arms of the trial in the first six months, but greater improvement in the cognitive behavioural approach arm, and in the active management arm there was no difference from baseline by twelve months. The mean difference between the two arms was 2.2 at three months, 1.8 at six months, and 4.1 at twelve months, all statistically significant. The SF12 mental subscale showed no significant improvement or differences between the two arms at twelve months. There were statistically significant improvements in the cognitive behavioural arm at six months, but by twelve months the improvement disappeared. Fear avoidance beliefs did not change in the active management arm but substantial improvements occurred in the CBA arm between 0 and three months and were maintained to twelve months. The mean difference at three months was 2.6, 3.1 at six months and 3.0 at twelve months, and the differences between treatments were significant at all time points.

Pain self-efficacy improved in the cognitive behavioural arm, with peak improvement at six months maintained at twelve months. In the active management arm, there were no discernible changes in pain self-efficacy, the mean difference at three months was 3.2, 4.1 at six months and 3.8 at twelve months.

Distributions of patient scores is given in Table 3.1. Subgroup analysis revealed that fear avoidance beliefs at baseline was not associated with treatment effects measured by the RMQ, but there was a statistically significant interaction between baseline fear avoidance and outcomes measured by MVK for disability, with those not fear avoidant at baseline having a larger treatment effect. However these observations were not consistent across all primary outcomes. The treatment effect was larger for people with moderately troublesome back pain as opposed to severe pain. Active management had little or no effect for these patients which accounts for the difference, and the interaction was not significant. There was no difference in the results between the observed case analysis and an analysis based on multiple imputation, and the findings appeared insensitive to the method of dealing with missing data.

Variable	Male		Female		Total	
Number	161	(39.6%)	246	(60.4%)	407	(100%)
Mean age in years (sd)	54.2	(14.7)	53.9	(14.3)	54.0	(14.0)
	N	%	N	%	N	%
Age group under 40 years	31	11.9	44	17.9	75	18.4
Age group 40-49 years	34	30.7	54	21.9	88	21.6
Age group 50-59 years	35	30.7	60	24.4	95	23.4
Age group 60-69 years	45	37.9	58	23.6	103	25.3
Age group 70 years and over	16	19.5	30	12.2	46	11.3
Treatment allocation						
Active management only	51	31.7	75	30.5	126	31.0
Active management and CBA	110	68.3	171	69.5	281	69.0
RMQ improvement of 3+ points						
Achieved	64	39.8	117	47.6	181	44.5
Not achieved	97	60.2	129	52.4	226	55.5
Anxiety						
Not anxious	91	56.5	86	35.0	177	43.5
Borderline	43	26.7	80	32.5	123	30.2
Anxious	27	16.8	80	32.5	107	26.3
Depression						
Not depressed	113	70.2	144	58.5	257	63.1
Borderline	40	24.8	75	30.5	115	28.3
Depressed	8	5.0	27	11.0	35	8.6
Pain self efficacy						
Very low	6	3.7	20	8.1	26	6.4
Low	25	15.5	36	14.6	62	15.0
Moderate or better	30	80.8	190	77.3	320	78.6
Troublesomness						
Somewhat	2	1.3	1	0.4	3	0.7
Moderately	82	50.9	113	45.9	195	47.9
Very	57	35.4	102	41.5	159	39.1
Extremely	20	12.4	30	12.2	50	12.3
interference with social activities						
All of the time	3	1.9	6	2.5	9	2.2
Most of the time	15	9.3	30	12.2	45	11.0
Some of the time	50	31.1	81	32.9	131	32.2
A little of the time	30	18.6	46	18.7	76	18.7
None of the time	63	39.1	83	33.7	146	35.9

Table 3.1: The numbers and proportions of patients in each category of each score in the BeST trial.

3.1.5 Discussion

This was a large scale randomised controlled trial which demonstrated the long-term effectiveness and cost-effectiveness of a cognitive behavioural approach in treating subacute and chronic lower back pain, with benefits lasting at least to twelve months and the cost per QALY less than half that of most other interventions for lower back pain. The intervention was delivered by NHS staff, demonstrating that it could be delivered within the NHS, which was the aim. Since the trial sample was recruited from a range of general practices, the external validity is good. The sample is representative of those who will accept an invitation to a cognitive behavioural approach, and of the ethnic mix within the UK, and there was no upper age limit imposed.

The study was powered to be definitive and the statistical significance of the comparisons across a range of measures supported the rejection of the null hypothesis that there was no difference between the groups. Just over half the people randomised to the cognitive behavioural approach reported improvements, which is more than would be expected from natural recovery or from provision of active management alone. The trial team hypothesise that this approach could be adapted to treat a range of musculoskeletal pain.

3.2 Introduction to Basingstoke and Alton Cardiac Rehabilitation data

3.2.1 Introduction

The Basingstoke and Alton cardiovascular rehabilitation data set is an observational study of an unselected cohort of 2,714 patients passing through routine NHS rehabilitation service in the 10-year period between 1st January 1993 and 31st December 2002 (Turner et al. [2002]). This cohort has now been followed up for an average of 11.5 years (range between one day and 18 years and three months), providing 11,871 person-years of follow-up. The study original hypothesis was formed and the data were collected by Dr Sally Turner (ST) during her employment as a programme coordinator at Basingstoke and Alton cardiac rehabilitation centre, and the preliminary analysis was performed in the course of her PhD research (Turner [2007]). ST describes the motivation of her research in terms of a particular patient who was being treated medically for clinical depression and who enrolled in cardiac rehabilitation following his discharge from hospital with a diagnosis of myocardial infarction (heart attack). This patient attended the rehabilitation programme regularly for three months, and on completion not only had his fitness improved, but he enjoyed improvements in his depression such that his antidepressant medication had been reduced. This case prompted ST to look more closely at the outcomes collected at this cardiac rehabilitation centre and to investigate whether baseline fitness or changes in fitness as a result of fitness training, or baseline psychological scores and changes in these would predict the prognosis of cardiovascular rehabilitation patients.

The study hypotheses formulated by ST were that risk of mortality in this coronary population:

- is predicted by baseline fitness, fitness on completion of cardiovascular rehabilitation programme, and the change in fitness during the programme, where fitness is measured by peak exercise performance on an exercise treadmill or bicycle ergometer;
- is predicted by the severity of depressive symptoms at baseline, at time of discharge from the programme, or the degree of change in the depression score as measured by the Hospital Anxiety and Depression Scale (HADS);
- is associated with a statistically significant interaction between baseline fitness and baseline depression scores, between fitness and depression scores at grad-

uation from the programme, and between change in fitness level and change in depressive symptoms by graduation from the programme.

Cardiac rehabilitation is the process by which patients with cardiac disease are encouraged and supported to achieve and maintain optimal physical and psychosocial health. Patients were referred to the Basingstoke and Alton cardiovascular rehabilitation programme by cardiologists and general practitioners at Basingstoke and North Hampshire Hospital and neighbouring general hospitals and primary care trusts. The programme was provided at three locations, the hospital physiotherapy gymnasium at Basingstoke, in a community centre at Tadley, 9 miles north of the hospital, and at purpose-built premises at Alton, 16 miles south of the hospital. There are four phases of cardiovascular rehabilitation: Phase I is when the patient is in hospital and being supported, along with their family, by specialist medical staff including the provision of going home advice and/or a self-help manual to assist convalescence; Phase II is the period immediately after hospital discharge, whilst surgery wounds heal, and the advice from Phase I is put into practice; Phase III is the active physical recovery provided through an individualised, incremental exercise prescription together with continuation of lifestyle advice and risk factor monitoring as provided in the Phase I and II; Phase IV involves patients complying with healthy lifestyle and risk factor control in the long term, and lasts for life.

This programme offers Phase I, III, and IV. The Phase III active physical recovery component lasted between six weeks and six months according to need, and is offered to all myocardial infarction (heart attack) patients discharged from the hospital, and to patients recovering from cardiac surgery or revascularisation and who lived within the catchment area which extended to a 25-mile radius of the Alton premises. The Phase III programme on offer at Basingstoke and Alton cardiac rehabilitation centre assessed patients for physical and psychological health at the beginning and end of the programme. They began a supervised aerobic exercise class once or twice a week, with home aerobic exercises in between. The supervised sessions comprised circuit training for 40 minutes, with those patients needing to rest between different aerobic exercises switching to strength and endurance exercises for 'active recovery'. Besides the exercise programme, a health education and stress management component was offered, to which patients' spouses or partners were also invited. This component covered relaxation techniques and a health education programme (understanding coronary heart disease, cholesterol, healthy eating, blood pressure, the benefits of regular physical activity, smoking advice, cardiac medications) and stress management.

3.2.2 Method

The Basingstoke and Alton cardiac rehabilitation data consists of baseline measures, taken at the first visit, of:

- fitness
- anxiety and depression
- a list of medications
- blood pressure
- age
- gender
- reason for referral
- former exercise habit
- employment status
- occupation
- quality of life score
- diabetes
- other co-morbidities
- blood cholesterol
- triglycerides
- thyroid function
- postcode (to be related to index of deprivation)
- weight
- Height (routinely recorded starting in 1998)

At the end of the rehabilitation, when the patient had reached the required ability to complete the exercises in the supervised session without recourse to active recovery, measurements were made of:

- fitness
- anxiety and depression
- blood pressure
- a list of medications
- the number of sessions attended
- any adverse events during rehabilitation

The date and cause of death were also recorded, being obtained automatically from the medical research department of the Office for National Statistics (National Statistics). These are coded using a binary indicator variable as Cardiac deaths or otherwise. Descriptive statistics and preliminary survival analysis using a Cox proportional hazards model was performed by ST and presented in her PhD Thesis (Turner [2007]).

Risk stratification

All patients who are recruited to exercise-based CR undergo risk stratification during initial assessment. In the Basingstoke and Alton CR, patients' risk level was assessed as standard. Exercise testing was a part of this process. Risk stratification enables an appropriate and individualised exercise prescription to be planned for patients that reflects the severity of cardiac illness, co-morbidity and current medical state. The American Association of Cardiovascular and Pulmonary Rehabilitation [(ACVPR) was the first to lay down criteria for risk stratification. Comorbidity as measured by the D'Hoore co-morbidity index were calculated for each Basingstoke and Alton CR patient (D'Hoore et al. [1996]).

Risk stratification criteria for cardiac patients (AACVPR 1999)

LOW RISK

- Uncomplicated MI, CABG, angioplasty or atherectomy
- Functional capacity equal to or greater than 6 METS 3 or more weeks after clinical event
- No resting or exercise induced myocardial ischaemia manifested as angina and/or ST segment displacement
- No resting or exercise-induced complex arrhythmias
- No significant left ventricular dysfunction (Ejection fraction equal to or greater than 50%)

MODERATE RISK

- Functional capacity less than 5- 6 METS 3 or more weeks after clinical event
- Mild to moderately depressed left ventricular function (Ejection fraction 31-49%)
- Failure to comply with exercise prescription
- Exercise induced ST-segment depression of 1-2mm or reversible ischaemia defects (echocardiography or nuclear radiography)

HIGH RISK

- Severely depressed left ventricular function (Ejection fraction equal to or less than 30%)
- Complex ventricular arrhythmias at rest or appearing or increasing with exercise
- Decrease in systolic blood pressure of >15mmHg during exercise or failure to rise consistent with exercise workloads
- MI complicated by Congestive Heart Failure, cardiogenic shock and/or complex arrhythmias
- Patients with severe CHD and marked (>2mm) exercise induced ST-segment depression
- Survivor of a cardiac arrest

D'Hoore co-morbidity index

Table 3.2 gives details of the D'Hoore co-morbidity index.

D'Hoore co-morbidity index	
Weight	Condition
1	Myocardial infarct*
	Congestive heart failure*
	Peripheral vascular disease
	Dementia
	Cerebrovascular disease †
	Chronic pulmonary disease
	Connective tissue disease
	Ulcer disease
	Mild liver disease‡
2	Hemiplegia ‡
	Moderate/severe renal disease (end stage) §
	Diabetes
	Any tumour ♣
	Leukaemia ♣
	Lymphoma ♣
3	Moderate or severe liver disease
6	Metastatic solid tumour

Table 3.2: *Myocardial infarct and congestive heart failure were omitted from the index because they are included in the AACVPR risk stratification for events.
†includes patients with history of stroke or history of cerebrovascular disease.
‡Mild liver disease and hemiplegia were omitted from index because it could not be quantified in Zoghbi database
§Includes patients with end stage renal disease
♣ Labelled as one category (malignancy)

3.2.3 Missing values

Details of the number and percentage of missing values are given in tables 6.18, 6.19 and 6.20 on pages 146, 147 and 148 in Chapter 6.

3.2.4 Previous results

Of the 2,714 patients who entered the study with a baseline fitness score, 1,398 completed cardiac rehabilitation (CR) and had an exit fitness score taken using the same protocol (bicycle or treadmill, Turner [2007]). There were statistically significant differences between patients who completed the CR programme and those who did not. There were twice as many current smokers in the group who did not complete CR. A greater proportion of patients with elementary occupations did not complete, and a higher proportion of managers and senior officials did complete. The patients who completed CR included 1.47% who reported no social support at all, whilst 4.48% of the non-completers reported no social support at all. Those who did not complete CR had a lower median fitness and greater median depression and anxiety than those who completed CR.

The mean age of the cohort was 62 years, with females (mean age 64.7 years) statistically significantly older than males (mean 61.0). The majority of patients were male (79.8%). Females were less fit than males (median VO_2 of 14.0ml/kg/min compared with 20.2ml/kg/min) and had higher median HADS anxiety scores (7 cf 6) and depression (4 cf 3). 4.6% of the cohort showed signs of being clinically depressed at baseline with 10.6% borderline on the depression scale. Myocardial infarction (MI) was the most common reason for referral (53.5%) with coronary artery bypass graft (25.4%) next and then angioplasty (PCI) (9.3%, a further 3.7% with MI plus PCI). Current smoking was reported in 8.9% of the patients, with a further 30.6% who had given up smoking within the previous year.

Comparison of those measures taken both before and after CR showed a 16.8% improvement in mean fitness, but systolic and diastolic blood pressure rose significantly (the expectation is that systolic blood pressure falls with increasing fitness). Almost half the patients categorised as low fitness at baseline moved into the medium fitness category by the end of CR, and just over half from the medium into the high fitness category, whilst very few moved to a lower fitness category. Anxiety and depression median scores and mean weight decreased significantly, and perception of social support rose.

Variable	Male		Female		Total	
Number	1320	(86.3%)	209	(13.7%)	1529	(100%)
Mean years of follow-up (sd)	11.3	(3.8)	11.1	(3.6)	11.3	(3.7)
Mean age in years (sd)	61.0	(9.4)	62.9	(9.0)	61.3	(9.4)
	N	%	N	%	N	%
Age group under 50 years	158	11.9	19	9.1	177	11.5
Age group 50-59 years	405	30.7	50	23.9	455	29.8
Age group 60-69 years	500	37.9	85	40.7	585	38.3
Age group 70 years and over	257	19.5	55	26.3	312	20.4
Diagnostic Category						
Myocardial Infarction (MI)	673	51.0	108	51.7	781	51.1
Coronary Artery Bypass Graft (CABG)	382	28.9	51	24.4	433	28.4
Percutaneous Coronary Intervention (PCI)	124	9.4	22	10.5	146	9.5
MI + PCI	56	4.3	7	3.3	63	4.1
Angina	61	4.6	19	9.1	80	5.2
Other cardiac	24	1.8	2	1.0	26	3.8
Smoking history						
Never smoked	347	26.3	93	44.4	440	28.8
Not for 10 years+	430	32.6	30	14.4	460	30.1
Not for 1-10 years	56	4.2	9	4.3	65	4.3
Recent quitter	407	30.8	64	30.6	471	30.8
Current smoker	80	6.1	13	6.3	93	6.0
D'Hoore Co-morbidity score						
None	968	73.3	141	67.5	1109	72.5
1 (least)	150	11.4	22	10.5	172	11.2
2	168	12.7	42	20.1	210	13.7
3	21	1.6	3	1.4	24	1.6
4 (most)	13	1.0	1	0.5	14	1.0
Diagnosis of diabetes	158	12.0	29	13.9	187	12.2
Family history of CHD	613	46.4	115	55.0	728	47.6
Weight at baseline						
A under 75kg	407	30.8	144	68.9	551	36.0
B 75-90kg	608	46.1	39	18.7	647	42.3
C over 90kg	305	23.1	26	12.4	331	21.7
Medications						
ACE inhibitor No	665	50.4	88	42.1	753	49.2
ACE inhibitor Yes	655	49.6	121	57.9	776	50.7
Aspirin No	45	3.4	12	5.7	57	3.7
Aspirin Yes	1275	96.6	197	94.3	1472	96.3

Table 3.3: Baseline values for patients at recruitment to the programme.

Variable	Male		Female		Total	
	N	%	N	%	N	%
Statin No	455	34.5	57	27.3	512	33.5
Statin Yes	865	65.5	152	72.7	1017	66.5
Beta blockers No	727	55.1	108	51.7	835	54.6
Beta blockers Yes	593	44.9	101	48.3	694	45.4
Occupation						
Managers & senior officials	236	17.9	16	7.6	252	16.5
Professional Occupations	143	10.8	11	5.3	154	10.1
Associate Professional	145	11.0	25	12.0	170	11.1
Administrative & secretarial	125	9.5	67	32.1	192	12.6
Skilled trade	362	27.4	13	6.2	375	24.5
Personal service	23	1.7	34	11.5	47	3.1
Sales and customer	26	2.0	16	7.6	42	2.7
Process, plant & machines	155	11.7	12	5.7	167	10.9
Elementary occupations	105	8.0	25	12.0	130	8.5
Fitness						
High baseline	590	44.7	29	13.9	619	40.5
Mid baseline	509	38.6	81	38.7	590	38.6
Low baseline	221	16.7	99	47.4	320	20.9
Depression at baseline						
Not depressed	1162	88.0	160	76.6	1322	86.5
Borderline	113	8.6	36	17.2	149	9.7
Depressed	45	3.4	13	6.2	58	3.8
Anxiety at baseline						
Not anxious	930	70.5	122	58.4	1052	68.8
Borderline	251	19.0	42	20.1	293	19.2
Anxious	139	10.5	45	21.5	184	12.0
Median baseline estimated						
VO_2 ml/kg / min	21.0		15.5		20.1	
(10th, 90th percentiles)	(13.09,	29.70)	(8.38,	24.50)	(11.0,	29.2)

Table 3.4: baseline values of patients at recruitment to the programme, continued.

3.2.5 Discussion

Before the work of ST, the evidence base for cardiovascular rehabilitation consisted of studies on mainly male patients in the middle age bracket with a low risk for a further cardiac event, and could not be generalised to females, to the elderly or to those with significant co-morbidity. Improved treatments and advancing diagnostic techniques have changed the profile of the cardiac rehabilitation population. The definition of myocardial infarction was changed again in 2005 (Thygesen et al. [2012]), with increased diagnosis including more patients than under the old, narrower definition. Many patients now have a diagnosis before suffering a myocardial infarction and have revascularisation (bypass) surgery or angioplasty and stenting (where the artery is opened up and a stent inserted to prevent it narrowing again). These patients have a swifter initial recovery period and can progress more quickly to Phase III exercise, and undertake a greater intensity of exercise. In addition, post-event medication has improved, with the use of cholesterol reduction with statins, ACE inhibitors, beta-blockers and anti-platelet therapy becoming standard and helping secondary prevention of coronary heart disease.

This cohort of patients was more representative than that of previous studies because it was an observational study of all those coming through the Basingstoke and Alton Cardiac Rehabilitation programme over a period of ten years. These include both elderly and female patients and those with significant co-morbidity, unlike the previous studies, which were small randomised controlled trials and meta-analyses. This cohort was also larger than other studies with 2,714 patients included in the study, giving it greater credibility.

This work was limited by the need to measure fitness using a treadmill or bicycle ergometer, rather than with sophisticated lab equipment. However, the equipment used is typical of the facilities used in NHS clinics, giving the work practical applicability. A further limitation was the small numbers of patients exhibiting signs of clinical depression using HADS, which limits the power of the data set to detect the influence of depression in the mortality of this cohort. Statins were not widely available until the mid 1990s and ACE inhibitors were less commonly prescribed until mid-way through the study period. There is variability in the drug protocol for patients which is time-dependent, further limiting the applicability. Nevertheless, this data provides an important opportunity for understanding a typical cardiac rehabilitation cohort, especially since most of the previous studies have restricted eligibility for participation in ways which make the population under consideration atypical.

The Cox proportional hazards survival model was built using the 385 deaths from all causes (25.2% of 1,529) including 192 (12.6%) from cardiovascular causes as at March 2011. Age, gender, diagnosis, co-morbidity score, fitness category after CR and fitness category before CR were all significant in explaining both all-cause and cardiovascular deaths. There was evidence that depression increased mortality, but this became not significant when fitness category was added to the model.

The next chapter introduces the methods used in the re-analysis of these two data sets.

Chapter 4

Methods 1: Conventional Methods

'The study of complex systems is about understanding indirect effects. Problems that are difficult to solve are hard to understand because the causes and effects are not obviously related' New England Complex Systems Institute (NECSI Team [2000]).

Health care data are often complex because, although the effects are seen, the mechanism which connects causes to effects is not well understood, for example the connection between patient attitude of mind and physical illness and symptoms. There is usually structure in medical data due to correlations between data collected at similar times or locations, and especially from the same patient. Data may also vary over time and space, and in a medical context this can mean slow changes in treatments over time due to new discoveries, equipment and policies, variation in treatment protocols between countries or hospitals, and even new diseases or redefinitions of existing ones. The two medical data sets we consider here are both complex in most of these ways. They are also both observations from complex interventions, treatments having many parts which may act independently or interdependently, as detailed on page 1 (Campbell et al. [2000]). In order properly to evaluate the contribution made by the artificial neural network (ANN) approach to complex health care data, it is first necessary to analyse the data using conventional methods to establish a baseline against which to compare the performance, advantages and disadvantages of the ANNs. In this chapter, the conventional methods used on each data set are explained. The results of conventional analysis were used as a basis for comparison of the outputs of the non-linear ANN models, to make a valid judgment as to the additional information provided by adding the non-linearity in the model by this method.

The first commonly used method is latent class analysis, a model-based approach to

dividing units of observation (often people) into subgroups. The subgroups cannot be observed directly, but are observed indirectly by means of two or more observed variables. This technique is applied to the Back Skills Training trial (BeST) data to ascertain if there are 'types' of patients whose response to this treatment differs. This is then a basis for comparison of the ability of the ANN to distinguish between patients with a positive response to the treatment and those without.
The second conventional method is standard survival analysis using the Cox proportional hazard model and this is applied to the Cardiovascular Rehabilitation (CR) data. This is then a basis for comparison of the ability of the ANN to produce a survival curve and to predict, both in the short and long term, the survival probability for a given patient. The further conventional method described here and used to explore the data is multiple imputation for missing data. This is required in the analysis of the cardiovascular rehabilitation data to assess the impact of missing data in real data sets. These will not be modelled directly by the artificial neural network, but their impact on predictions was assessed.

4.1 Latent Class Analysis

Latent Class Analysis (LCA) is a method for identifying sub classes of people used in many fields, for example ADHD and comorbid symptoms in a population sample of adolescent female twins (Neuman et al. [2001]), longitudinal cohort study on back pain (Dunn et al. [2006]), symptom profiles of persons who experience traumatic events (Breslau et al. [2005]), satisfaction-with-life-domain profiles (Eid et al. [2003]), identifying clinically distinct subgroups of self-injurers among young adults (Klonsky and Olino [2008]), eating disorder phenotypes (Keel et al. [2004]), developmental trajectories of crime (Eggleston et al. [2004]), ADHD symptoms in a school sample of Brazilian adolescents (Rohde et al. [2001]), human herpes virus 8 assay performance and infection prevalence in sub-Saharan Africa and Malta (Engels et al. [2000]), course of depressive symptoms after myocardial infarction and cardiac prognosis (Kaptein et al. [2006]), child behaviour checklist anxiety and depression in children and adolescents (Wadsworth et al. [2001]), country and consumer segmentation of financial product ownership (Bijmolt et al. [2004]), symptoms associated with chronic fatigue syndrome and fibromyalgia (Sullivan et al. [2002]), distinguishing phenotypes of childhood wheeze and cough (Spycher et al. [2008]), underage problem drinking of 16 to 20 year olds (Reboussin et al. [2006]).

4.1.1 Introduction to latent class analysis

The objective of LCA is to categorise people into classes using the observed variables and to identify those items which best distinguish between classes (Nylund et al. [2007]).

LCA was introduced as a way of formulating latent variables from dichotomous survey items in data about attitudes (Lazarsfeld and Henry [1968]). LCA models identify an error-free latent class variable, which is unobserved and categorical, by measuring a number of observed response variables, which are assumed to be subject to error. Statistical analysis based on latent variable models attempts to separate the latent variable and the measurement error (Collins and Lanza [2010]). LCA models assume that the latent variable is categorical, in contrast to factor analysis which posits continuous latent variables. Factor analysis is a variable-oriented approach where the focus is to identify the relationship between variables which, it is supposed, applies across all people. LCA is a person-oriented approach which seeks to identify subtypes of individuals who exhibit similar patterns of individual characteristics. In this application the aim is to identify patient types and their likely responses to the complex intervention tested in the BeST trial. For this reason, the patient-oriented LCA approach was used.

LCA assumes that each observation (patient) is a member of one, and only one, of $\mathcal{C}$ latent classes. The classes are not observed directly, but observed via the responses of the individuals given to the variables measured. The categorical latent variable which we seek to discover, $\mathcal{L}$, has $c = 1 \dots \mathcal{C}$ latent classes. The prevalence of each latent class is the probability of membership in latent class c of latent variable $\mathcal{L}$, and is denoted γ_c. Since the latent classes are mutually exclusive and exhaustive, the prevalences sum to one

$$\sum_{c=1}^{\mathcal{C}} \gamma_c = 1 \tag{4.1}$$

In general, the latent class model has $j = 1 \dots J$ observed variables and observed variable j has $r_j = 1 \dots R_j$ response categories. Cross tabulating the J variables forms a contingency table with $W = \prod_{j=1}^{J} R_j$ cells. The entire range of possible combinations of responses forms an array of response patterns Y with W rows and J columns. Each response pattern y is associated with a probability $P(Y = y)$ and $\sum P(Y = y) = 1$. The item-response probability $\rho_{j,r_j|c}$ is the probability of response r_j being given to observed variable j, conditional on membership in latent class c. For example, the probability of a patient being classified 'clinically anxious', given they have been assigned to a certain class. Since each patient provides only one of

the response alternatives to variable j, the vector of item-responses for a particular variable conditional on a particular latent variable sums to 1

$$\sum_{r_j=1}^{R_j} \rho_{j,r_j|c} = 1 \ \forall \ j \tag{4.2}$$

If y_j represents element j of response pattern y, and indicator function $\mathcal{I}(y_j = r_j) = 1$ when the response to variable $j = r_j$, and 0 otherwise, then the probability of observing a particular vector of responses is given by

$$p(Y = y) = \sum_{c=1}^{\mathcal{C}} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{\mathcal{I}(y_j=r_j)}. \tag{4.3}$$

The log likelihood function that is maximised with respect to $\rho_{j,r_j|c}$ and γ_c in a data set with N observations is

$$ln\ L = \sum_{i=1}^{N} ln \sum_{c=1}^{\mathcal{C}} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{\mathcal{I}(y_j=r_j)}. \tag{4.4}$$

The expectation-maximisation (EM) algorithm (Dempster et al. [1977]) is often used to estimate the parameters in LCA models. Finding a maximum likelihood solution requires taking the derivatives of the likelihood function with respect to all the unknown values, resulting in a set of interdependent equations in which solving for the parameters requires the latent variables and vice-versa. This is overcome by an iterative algorithm which estimates the parameters and the latent variables by first picking values for one and then using those values in solving for the other. Then those solutions are used in finding better estimated for the first, and so on. This is repeated until the log likelihood ceases to increment beyond some arbitrarily small value (Linzer and Lewis [2011a])

There are few modelling assumptions associated with LCA; the one fundamental assumption is that, conditioned on the latent variable, the observed variables are independent. This is called the assumption of local independence, and underpins equation (4.3) so that the responses are conditioned on the classes and not also on each other. This is not to say that the the observed variables in the data set to be analysed are independent, but that the relations among the observed variables are explained by the latent classes. The latent classes are assumed nominal and their categorical indicators (the variables used to identify the classes) have a joint multinomial distribution. It is therefore unnecessary to assume a distributional form, for example, a multivariate normal distribution on the model.

4.1.2 Model selection

Since item-response probabilities are not regression coefficients but conditional probabilities, it is necessary to examine the pattern of item-response probabilities across all response alternatives for the variable and across all latent classes and not just a single probability when determining the strength of the relation of the observed variable and the latent variable (Collins and Lanza [2010]). There are two criteria which together reveal a strong relation between an observed variable j, and a latent variable $\mathcal{L}$, namely an array of item-response probabilities that are close to 0 or 1 corresponding to the variable j and a distribution for the conditional probabilities $\rho_{j,r_j|c}$ that varies across the latent classes. When the observed variable and the latent variable are independent, the conditional probabilities are the same as the variable's marginal proportions for each response, i.e. $\rho_{j,r_j|c} = P_{j,r_j}$. When they are not independent, we have $\rho_{j,r_j|c} \neq \rho_{j,r_j|c'}$ for some c and c'. If $\rho_{j,r_j|c} = 1$ for $r_j = k$, then $\rho_{j,r_j|c} = 0 \; \forall \; r_j \neq k$ i.e. conditional on membership of class, a particular response (k) can be determined with certainty. An item response probability of 1 clearly reflects a high degree of certainty. An item response probability, r_j, of 0 also reflects a high degree of certainty, and if there are just two response alternatives, then the other response must have probability 1. In such a case, 0 reflects as much certainty as 1, but if there are more alternatives, a 0 probability for one response does not give any information on the probability of other responses. Where the marginal response proportions are close to 0 and 1, then independence between the observed variable and the latent variable cannot be discounted, hence the requirement for both criteria in determining the strength of the relation.

Homogeneity and latent class separation are helpful aids to interpretation of the latent classes. Latent class c is highly homogeneous when members of c are highly likely to provide the same observed response pattern, implying this response pattern is characteristic of latent class c. If the response patterns seen in class c are highly variable, then there is low homogeneity. Recalling equation (4.3) and that $\gamma_c = P(\mathcal{L} = c)$,

$$p(Y = y|\mathcal{L} = c) = \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{\mathcal{I}(y_j=r_j)}. \tag{4.5}$$

Homogeneity of latent class c is perfect when $\rho_{j,r_j|c}$ is 0 or 1 for all variables j and all response categories r_j. In this case, there is a unique response pattern y' for which $P(Y = y'|\mathcal{L} = c) = 1$ when $\rho_{j,r_j|c} = 1 \; \forall \; \rho \neq 0$ and $P(Y = y'|\mathcal{L} = c) = 0$ for all the remaining response patterns. Good latent class separation occurs when the pattern of item-response probabilities across indicator variables clearly differentiates

among the latent classes. Perfect latent class separation is seen when for each latent class c' there is a unique response pattern y' for which $P(Y = y'|\mathcal{L} = c') = 1$ and $P(Y = y'|\mathcal{L} = c) = 0 \; \forall \; c \neq c'$. Homogeneity and latent class separation are based on the same quantities and a high degree of latent class separation implies a high degree of homogeneity since if for latent class c', $P(Y = y'|L = c') = 1$ then it follows that $P(Y = y'|\mathcal{L} = c) = 0 \; \forall \; c \neq c'$. Perfect homogeneity for latent class c' requires $P(Y = y'|\mathcal{L} = c') = 1$ but this does not give any information about the homogeneity of the other classes; $1 \geq P(Y = y'|\mathcal{L} = c) \geq 0$ holds for $c \neq c'$. Homogeneity and latent class separation are closely related to uncertainty in probability of membership in class c conditional on response pattern y, $P(\mathcal{L} = c|Y = y)$. To obtain an expression for this, equations (4.3) and (4.5) are substituted into Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4.6}$$

which in the LCA context becomes

$$P(\mathcal{L} = c|Y = y) = \frac{P(Y = y|\mathcal{L} = c)P(\mathcal{L} = c)}{P(Y = y)} \tag{4.7}$$

substituting equations (4.3) and (4.5),

$$P(\mathcal{L} = c|Y = y) = \frac{(\prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{\mathcal{I}(y_j=r_j)})\gamma_c}{\sum_{c=1}^{\mathcal{C}} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{\mathcal{I}(y_j=r_j)}}. \tag{4.8}$$

Equation (4.8) can be used to obtain a vector of classification probabilities for each individual based on that individual's observed response pattern and the latent class prevalence and item-response probabilities from any latent class model fit to the data. The vector will include the probability of membership in each of the $\mathcal{C}$ latent classes for that individual (Collins and Lanza [2010]) and will sum to 1. These posterior probabilities tend to be large for one single latent class and small for all others where there is both strong latent class separation and strong homogeneity, i.e there is little classification uncertainty. When either homogeneity or class separation is weak, there is greater classification uncertainty seen as posterior probabilities which are more similar across latent classes.

4.1.3 Latent class analysis in R

The R package *poLCA* (R Development Core Team [2011]) was used to build the Latent class models. The poLCA package uses the expectation-maximisation and Newton-Raphson methods to find maximum likelihood estimates of the model parameters, the proportions of observations in each class, γ_c, and the probabilities of observing each response to each manifest variable conditional on latent class, $\rho_{j,r_j|c}$, (Linzer and Lewis [2011b]). In order to ensure a global maximum likelihood has been obtained, rather than a local one, use was made of the facility within the poLCA package to re-estimate the model automatically a specified number of times and then retain the models with the greatest likelihood. Ten replications were specified and a maximum of up to 10,000 iterations, to ensure convergence to the parameter estimates that produce the global maximum likelihood. Several repeats of model-fitting with these settings produced models with consistent maximum log likelihood, parsimony measures AIC (equation (4.23)) and BIC (equation(4.24)), likelihood ratio (G^2) (Equation (4.22)), and Pearson's χ^2 goodness of fit for observed versus predicted cell counts. The optimal number of latent classes is not automatically determined by the package but specified by the user, so in the case of the BeST data, models with two to six models were fitted and compared using the goodness of fit statistics listed above. Evaluation of different information criteria for assessing the number of latent classes has been investigated through a simulation study (Yang [2006]). The performance of the various criteria was affected by sample size and number of latent classes in the underlying simulated data. AIC out-performed BIC at the sample sizes and for the range of latent class numbers investigated here. But in a subsequent simulations study, BIC was found to perform more accurately (Nylund et al. [2007]). Therefore, both are considered in the analysis of the BeST data given on page 108

4.2 Survival Analysis

Survival analysis is the phrase used to describe the analysis of data in the form of times from a well-defined time origin until the occurrence of some particular event or end-point. In medical applications, these often correspond to recruitment of a patient into the study and an event such as recurrence of illness or patient death respectively. This approach is widely described as time-to-event analysis, since the methodology used in modelling survival of patients can also be used in other applications, such as failure times of plant and machinery in reliability analysis, time to drop-out from a course of study in education research, etc. Survival (or reliability) data are said to be censored when the event of interest is not observed. Often censoring occurs when the data from a medical study, where the end-point of interest is patient death, are to be analysed at a point in time when some of the individuals are still alive. Censoring can also occur when the survival status of one or more individuals is not known at the date of analysis because they have been lost to follow-up, perhaps because they have relocated and their current address is unknown to the study team, or because they died of another cause, such as a road traffic accident rather than the event of interest. Through a modelling approach to the analysis of survival data, the survival experience of a group of patients, and how it depends on the values of one or more explanatory variables, whose values have been recorded for each patient at the time origin, can be explored (Collett [2003]).

4.2.1 Introduction to survival analysis

Informative censoring

An important assumption in the analysis of censored survival data is that the actual survival time of an individual, t, is independent of any mechanism that causes an individual's survival time to be censored at time c, where $c < t$. This means that if we consider a group of individuals, all of whom have the same values of relevant prognostic variables, an individual whose survival time is censored will be representative of those at risk at the censoring time if the censoring process operates randomly. When survival data are to be analysed at a predetermined point in calendar time, or at a fixed interval of time after the origin for each patient, the prognosis for individuals who are still alive can be taken to be independent of the censoring, so long as the time of analysis is specified before the data are examined. Informative censoring is the case where the censoring is related to the survival time, such as if patients in a trial comparing treatments were withdrawn from one arm of the study because that treatment caused life-threatening side effects. The survival

rates for that treatment would appear larger than they were, leading to an incorrect estimate of the treatment difference. In such a case, sensitivity analysis is performed to compare the original data analysis to analysis which first assumes the censored patients were high risk and experienced the event right after censoring, and then assumes they were low risk and survived the longest. If informative censoring begins only after a significant period of time, survival up to onset of censoring cause can be analysed. The assumption of uninformative censoring can be examined by plotting observed survival times against the values of explanatory variables, distinguishing censored from uncensored survival times. If there is a greater proportion of censored survival times in patients with a particular range of values of explanatory variables, there is evidence of informative censoring.

Kaplan-Meier estimate of the survivor function

The first step in the analysis of censored survival data is often to calculate the Kaplan-Meier estimate of the survivor function, $\hat{S}(t)$, from the data. It can be shown that the Kaplan-Meier estimator maximizes the generalized likelihood over the space of all distributions so its evaluation on large data sets gives a good qualitative description of the true survival function (Eleuteri et al. [2003]). To obtain a Kaplan-Meier estimate, a series of time-intervals is created, each containing a single death time. Each death event is assumed to have occurred at the beginning of the time interval. The Kaplan-Meier estimate of the survivor function is based on the assumption that the r death times of the n individuals in the sample occur independently of one another. If n_j individuals are alive just before d_j deaths occur at time t_j then the estimated probability of survival through the time interval from $t_j - \delta$ to t_j, where δ is small and contains only one death, is $\frac{(n_j - d_j)}{n_j}$. Then, the estimated survivor function at any time, t, in the kth constructed time interval from $t_{(k)}$ to $t_{(k+1)}$, $k = 1, 2, \ldots, r$, where $t_{(r+1)}$ is defined to be ∞, will be the estimated probability of surviving beyond $t_{(k)}$. This is actually the probability of surviving through the interval from $t_{(k)}$ to $t_{(k+1)}$ and all the preceding intervals, and leads to the Kaplan-Meier estimate of the survivor function, which is given by:

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{(n_j - d_j)}{n_j} \tag{4.9}$$

for $t_{(k)} \leq t < t_{(k+1)}$.

Basic equations of survival modelling

Survivor function is the probability that the survival time T is greater than some value t:

$$S(t) = P(T \geq t). \tag{4.10}$$

$F(t)$ is the cumulative density function of T, and $f(t)$ is the probability density function of T:

$$F(t) = P(T < t) = \int_0^t f(u)du = 1 - S(t), \tag{4.11}$$

so

$$f(t) = \frac{dF(t)}{dt} = -S'(t). \tag{4.12}$$

The hazard function is the risk or hazard of death at time t, and is derived from the probability that an individual dies at time t conditional on their having survived up to that time. This conditional probability is expressed as probability per unit time by dividing by the time interval δt to give a rate (sometimes called the hazard rate, the force of mortality, or the instantaneous death rate), and the hazard function, $h(t)$, is the limiting value as δt tends to zero. This leads to:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\left(logS(t)\right). \tag{4.13}$$

The cumulative hazard, $H(t)$ is

$$H(t) = \int_0^t h(u)du = -logS(t), \tag{4.14}$$

so that

$$S(t) = exp\left(-H(t)\right). \tag{4.15}$$

When fitting survival models to data, estimates of the unknown parameters are found by maximising the logarithm of the likelihood.

The likelihood function for randomly censored data

When the censoring times are not informative, the likelihood function is derived (Collett [2003]) as follows:
The likelihood function for randomly censored data of n individuals, with observed time t_i for ith individual, $i = 1, 2, \ldots, n$, and event indicator Δ which takes the value $\Delta = 1$ if t_i is an event time and $\Delta = 0$ if the event of interest is censored.
Let T_i be the random variable associated with *event* time of the ith individual and

C_i be the random variable associated with time to censoring. Then the value t_i is an observation on the random variable $\tau_i = min(T_i, C_i)$. The density function of T_i is $f_{Ti}(t)$ and the survivor function $S_{Ti}(t)$. Similarly, the the random variable associated with censoring time C_i has density function $f_{Ci}(t)$ and survivor function $S_{Ci}(t)$.

The probability density distribution for the pair (τ_i, Δ_i) for censored observations, so that $\Delta = 0$ is described by

$$p(\tau_i = t, \Delta_i = 0) = P(C_i = t, T_i > t).$$

This joint probability is a mixture of continuous and discrete components but to simplify the presentation, $P(T_i = t)$, for example, will be understood to be the probability density of function of T_i. Assuming independence of the event time distribution T_i from censoring time C_i,

$$\begin{aligned} P(C_i = t, T_i > t) &= P(C_i = t)P(T_i > t) \\ &= f_{C_i}(t)S_{T_i}(t) \end{aligned},$$

and so

$$P(\tau_i = t, \Delta_i = 0) = f_{C_i}(t)S_{T_i}(t).$$

Similarly, if the observations are not censored, so that $\Delta = 1$

$$\begin{aligned} P(\tau_i = t, \Delta_i = 1) &= P(T_i = t, C_i > t) \\ &= P(T_i = t)P(C_i > t) \\ &= f_{T_i}(t)S_{C_i}(t) \end{aligned},$$

again assuming that the distributions of C_i and T_i are independent.

Combining these results, the joint probability, or likelihood of the n observations, $t_1, t_2, \ldots, t_n$, is

$$\prod_{i=1}^{n} (f_{T_i}(t_i)S_{C_i}(t_i))^{\Delta_i} (f_{C_i}(t_i)S_{T_i}(t_i))^{(1-\Delta_i)}$$

which can be written as

$$\prod_{i=1}^{n} (f_{C_i}(t_i)^{(1-\Delta_i)}S_{C_i}(t_i))^{\Delta_i} \times \prod_{i=1}^{n} (f_{T_i}(t_i)^{\Delta_i}S_{T_i}(t_i))^{(1-\Delta_i)}.$$

Under the conditions of non-informative censoring, the first product will not contain any parameters that are relevant to the distribution of survival times, and so can be regarded a constant. The likelihood of the observed data is therefore proportional

to the second product:

$$L = \prod_{i=1}^{n} (f_{T_i}(t_i)^{\Delta_i} S_{T_i}(t_i))^{(1-\Delta_i)}. \tag{4.16}$$

Log-likelihood is

$$log\ L = \sum_{i=1}^{n} [\Delta_i\ log(f(t_i)) + (1 - \Delta_i)\ log(S(t_i))] \tag{4.17}$$

which can be maximised to fit a survival model to data.

4.2.2 Cox proportional hazards model

The Kaplan Meier estimate of the survival function is a non-parametric approach. In most medical studies, survival data is supplemented by physiological variables and lifestyle factors which can be used as explanatory variables. Statistical modelling is used to explore how the survival experience of a group of patients depends on the explanatory variables, whose values have been recorded for each patient. In the analysis of survival data, the risk or hazard of death at any time after the origin of the study is the centre of interest. As a consequence, the hazard function is modelled directly in survival analysis.

The most widely used survival specification which takes into account the system features is the proportional hazards model. The two most common approaches are either to choose a parameterized functional form for the baseline hazard and then make use of Maximum Likelihood to find values for those parameters, or to make a Maximum Likelihood estimation of only the feature dependent part of the model without fixing the baseline hazard, and derive a non-parametric estimate for the baseline hazard (Eleuteri et al. [2003]).

There are two broad reasons for modelling survival data. One objective of the modelling process is to determine which combination of potential explanatory variables affects the form of the hazard function, $H(t)$. In particular, the effect that the treatment has on the hazard of death can be studied, as can the extent to which the other explanatory variables affect the hazard function. Another reason for modelling the hazard function is to obtain an estimate of the the hazard function itself for an individual, in contrast to regression analysis where the mean response is modelled.

This hazard function may be of interest in its own right, but in addition, from the relationship between the survivor function and hazard function described in equation (4.15), an estimate of the survivor function can be found for the individual. This will in turn lead to an estimate of quantities such as the median survival time, which will be a function of the explanatory variables in the model. The median survival time could then be estimated for current or future patients with particular values of these explanatory variables. The resulting estimate could be particularly useful in devising a treatment regimen, or in counselling a patient about their prognosis.
When the hazard of death at any time for any individual in one group is proportional to the hazard at that time for a similar individual in another group, the assumption of proportional hazards may be employed. The proportional hazard model proposed by Cox in 1972 is based on the assumption of proportional hazards, and since no particular form of probability distribution is assumed for the survival times, the model is a semi-parametric model.
Let the set of values of the explanatory variables in the proportional hazards model be the vector $x = (x_1, x_2, \ldots, x_p)'$. When the values of all the explanatory variables that make up the vector x are zero, the hazard function is called the baseline hazard function, $h_0(t)$. The hazard function for the ith individual can then be written

$$h_i(t) = \nu(x_i)h_o(t), \tag{4.18}$$

where $\nu(x_i)$ is a function of the values of the vector of explanatory variables for the ith individual, and can be interpreted as the hazard at time t for an individual whose explanatory variable is $x_i \neq 0$ relative to an individual whose explanatory variable is $x_i = 0$. Since the relative hazard $\nu(x_i)$ cannot be negative, it is convenient to write it as $exp(\eta_i)$ where (η_i) is a linear combination of the explanatory variables. The general proportional hazards model then becomes

$$h_i(t) = exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi})h_0(t) \tag{4.19}$$

so that the log of the hazard ratio is a linear regression:

$$ln\left(\frac{h_i(t)}{h_0(t)}\right) = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}. \tag{4.20}$$

The estimation of the β parameters in the linear component of a proportional hazards models is all that is required to draw inferences about the effect of explanatory variables in the model on the hazard function. These can then be used to estimate the hazard function itself, and the corresponding survivor function. The

estimates of the baseline hazard, survivor and cumulative hazard functions can be used to obtain corresponding estimates for an individual patient with vector of explanatory variables x_i.

Factors can be included in the Cox model in the same way as variates, except that one level of the factor has to correspond to the baseline hazard, and so the other levels of the factor are expressed in proportion to this reference level. Similarly, interactions between factors can be included along with their relevant main effects, as can mixed terms formed of an interaction between a factor and a variate. Fitting the Cox model entails estimating the unknown coefficients (β_i) of the explanatory variables using the method of maximum likelihood to give the maximum likelihood estimates. Cox showed that the relevant likelihood function for the proportional hazards model is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\beta' x_{(j)})}{\sum_{l \in R_{(t_{(j)})}} exp(\beta' x_l)}. \tag{4.21}$$

where $R_{(t_{(j)})}$ is the set of individuals at risk at time t_j, i.e those alive and uncensored at a time just prior to t_j. The proportional hazard model for survival data assumes the hazard function is continuous and under such an assumption, tied survival times are not possible. In practice, the recording of deaths to the nearest day, month or year can give rise to tied survival times. Kalbfleisch and Prentice gave the appropriate likelihood function for this case (Kalbfleisch and Prentice [2002]), but often a simpler approximation to the likelihood function is used. This work uses the Breslow method for tied times (Breslow [1975]). Having estimated the coefficients (βs), the approximate 95% confidence interval can be calculated using the standard normal distribution. Evidence that the value of β is not zero is given if this interval does not contain zero, and testing the null hypothesis $\beta = 0$ using the statistic $\frac{\hat{\beta}}{se\hat{\beta}}$ using the standard normal distribution gives a small p-value, also provides such evidence. Equivalently, the Wald test uses the square of this statistic compared to the percentage points of a χ^2 distribution. In the usual case where there are a number of explanatory variables, the effect of each term depends on the other terms currently in the model. If another model is built using a subset of the terms in the original model, it is said to be parametrically nested within the original model. Nested models can be compared using the ratio of their likelihoods, and

$$-2\ log\left(\frac{\hat{L}(1)}{\hat{L}(2)}\right) \tag{4.22}$$

follows a χ^2 distribution with degrees of freedom equal to the difference between the number of independent β parameters between the nested models, q. Comparison of models, which need not be nested, can be made using Akaike's information criterion AIC or Bayesian information criterion (BIC) which both take account of the number of parameters and are defined as:

$$AIC = -2\ ln\ \hat{L} + 2q, \tag{4.23}$$

$$BIC = -2\ ln\ \hat{L} + q\ ln\ (n) \tag{4.24}$$

The smaller the value of the AIC or BIC, the better the model fit; the value of the information criteria will rise when unnecessary parameters are added to a model, although the AIC penalises the number of parameters less strongly than does the BIC.

The usual recommended strategy for model selection is first to use each variable as a lone predictor and ascertain which models are a significant improvement on the null model, $h_0(t)$. These variables are then fitted together in a single model, and nested models formed by removing each variable in turn are tested, with those leading to a significant decrease in the $-2\ log\ \hat{L}$ retained in the model, and the remainder removed. Variables which were not significant as sole predictors are then added one at a time and the same statistic is used to determine if they have become significant in the presence of other variables. Only after this are interactions and other higher order terms (such as powers of variates) tested for inclusion. Higher order terms are used to capture non-linearity suspected of existing in the system. When a model has been selected by this optimisation procedure, the coefficients of each variate are interpreted as the logarithms of the ratio of the hazard of death to the baseline hazard. The linearity assumption means that the hazard ratio, for example between a patient aged 80 relative to to one age 75, is the same as the hazard ratio between a patient aged 20 relative to one aged 15. If this assumption does not hold, the variate can be divided into levels of a factor and these can be tested for inclusion in the model instead.

Although the proportional hazards model finds widespread applicability in the analysis of survival data, there are relatively few probability distributions for the survival times that can be used with this model. Moreover, the distributions that are available, principally the Weibull and Gompertz distributions, lead to hazard functions which increase or decrease monotonically.

A model that encompasses a wider range of survival time distributions is the accelerated failure time model. In circumstances where the proportional hazards assumption is not tenable, models based on this general family may prove to be fruitful. Again, the Weibull distribution may be adopted for the distribution of survival times in the accelerated failure time model, but some other probability distributions are also available.

One other general family of survival models, known as the proportional odds model, may be useful in some circumstances. Although any continuous distribution for non-negative variables might be used, the properties of the log-logistic distribution make it a particularly attractive alternative to the Weibull distribution. The lognormal, gamma and inverse Gaussian distributions are sometimes used in accelerated failure time modelling.

One limitation of the Weibull hazard function is that it is a monotonic function of time. However, for example, following a heart transplant, a patient faces an *increasing* hazard of death over the first 10 days or so while the body adapts to the new organ. The hazard then decreases with time as the patient recovers. The log-logistic distribution is so called because the variable $\log T$ has a logistic distribution, a symmetric distribution whose probability density function is very similar to that of the normal distribution.

The gamma distribution, like the Weibull distribution, includes the exponential distribution as a special case. Indeed the gamma distribution is quite similar to the Weibull distribution and and inferences based on either model will often be very similar. The generalised gamma distribution is actually more useful and includes the Weibull and lognormal distributions as special cases.
The inverse Gaussian is a flexible model that has some important theoretical properties, but the complicated form of the survivor function makes this distribution difficult to work with.
When the number of observations in a single sample is reasonably large, an empirical estimate of the hazard function could be obtained using the life table or Kaplan-Meier methods. A plot of the estimated hazard function may then suggest a suitable parametric form for the hazard function.

Model Checking

If the model fitted to the observed data is satisfactory, then a model-based estimate of the survivor function $\hat{S}_i(t_i)$ for the ith individual at t_i, the survival time of that individual, will be close to the corresponding true value $S_i(t_i)$. This suggests that if the correct model has been fitted, the values $\hat{S}_i(t_i)$ will have properties similar to those of $S_i(t_i)$. Then the negative logarithms of the estimated survivor functions, $-log\ \hat{S}_i(t_i),\ i = 1, 2, \ldots, n$ will behave as n observations from a unit exponential distribution. These estimates are the Cox-Snell residuals. If the observed survival time for an individual is right-censored, then the corresponding value of the residual is also right-censored. The residuals will therefore be a censored sample from the unit exponential distribution, and a test of this assumption provides a test of model adequacy. Modified Cox-Snell (Cox and Snell [1968]) residuals take account of the fact that censored observations lead to residuals that cannot be regarded on the same footing as residuals derived from uncensored observations.

Modified Cox-Snell residuals can be refined further into Martingale residuals (Therneau and Grambsch [2000]) with zero mean when uncensored. Unlike the above, Martingale residuals are similar to residuals encountered in other areas of data analysis in that it is the difference between the observed number of deaths and the corresponding estimated expected number on the basis of the fitted model. They are not, however, symmetrically distributed about zero, so Therneau et al. introduced Deviance residuals (Therneau et al. [1990]), which are symmetrically distributed about zero.

Two disadvantages of the residuals described above are that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. Both these disadvantages are overcome in the Schoenfeld residuals (Schoenfeld [1982]). This residual differs from the others in one other important respect: that there is not a single value of the residual for each individual, but a set of values, one for each explanatory variable included in the fitted Cox regression model. Schoenfeld residuals are standard for time-to-event analysis.
We need to assess the validity of the β values and the proportionality assumption in evaluating Cox model fit. The functional form of the covariates can be checked using the Martingale residuals obtained from fitting the null model (no covariates). Plotting these against the values of each covariate in the model then this should display the functional form required for the covariate, as shown by Therneau. In particular, a straight line plot indicates that a linear term is needed.

Influential observations can be identified by examining the extent to which the estimated parameters in the fitted model are affected by omitting the observation. Testing the assumption of proportional hazards can be performed as follows: if the log-cumulative hazard functions for individuals with different values of their explanatory variables are plotted against time, the curves so formed will be parallel if the proportional hazards model is valid. In terms of assessing the overall fit of a model, a plot of the deviance residuals against the risk score gives information on observations that are not well fitted by the model, and their relation to the set of values of the explanatory variables. Plots of residuals against survival times, the rank order of the survival times or explanatory variables may also be useful.

4.3 Missing Data

4.3.1 Introduction to analyses with missing data

Standard statistical analysis methods have largely been developed for use with rectangular data sets comprising rows of cases or observations and columns of variables (Little and Rubin [2002]). A common approach to cases that have some observations missing is to omit such cases from the analysis, and work with the remainder as if it were the entire data set. This is called 'complete-case analysis'. In situations where missing values are confined to a tiny percentage of cases, this is likely to leave the conclusions drawn from the analysis unchanged. However, in many real-world applications, cases with missing observations comprise a substantial proportion of the collected data. The reasons why the data is missing dictates how it may be treated. If a survey respondent declined or omitted to answer some of the questions, or the measurement equipment in an experiment suffered a failure, then it is reasonable to suppose that there are actual underlying values that would have been observed had the data gathering mechanism performed better. In contrast, if an opinion survey asks a respondent to choose between two or more options, it is not clear whether non-response is due to omission or to indecision, in the latter case possibly representing a 'don't know' stratum of the population (Little and Rubin [2002]).

It is possible to code the non-responses as missing and even to have several codes to distinguish between 'don't know', 'refuse to answer', or 'out of legitimate range'. Some statistical packages default to complete-case analysis and simply exclude cases that have data missing in any of the variables involved in the analysis.

Complete-case analysis is generally inappropriate since the investigator is usually interested in making inferences about the entire target population, rather than just the portion of the target population that would provide responses on all relevant variables in the analysis.

Missing data *patterns* describe which values are observed and which are missing in the data matrix. Missing data *mechanisms* concern the relationship between missingness and the values of variables in the data matrix.
Some example of missing data patterns are :

- Univariate missing data where the missing values all occur in a single variable, e.g. values are missing for an entire plot in an agricultural experiment.
- Unit and item non-response survey data:
  Item non-response occurs where values are missing on particular items, e.g. in a questionnaire and typically have a haphazard pattern.
  Unit non-response occurs where variables are all observed or missing on the same set of cases, e.g. where some of the respondents cannot be contacted for part of the survey.
- Attrition in longitudinal studies having a monotone missing data pattern:
  Participants drop out over time so all data is available for some variables, but later collections show increasing missingness.
- File matching problem with two sets of variables never jointly observed. When variables are never observed together, parameters relating to the association between them cannot be estimated.
- Latent variable indicator variables with manifest variable patterns that are never observed.

It is assumed that the missingness indicators in all these cases hide true values that are meaningful for analysis, so that it makes sense to fill in (impute) the missing values of the unobserved variables.
Mechanisms that lead to missing data are a different issue to patterns of missing data. Whether the fact that variables are missing is related to the underlying values in the data set informs the analyst which missing data methods are appropriate. In 1976, Rubin formalised the crucial role of mechanism (Little and Rubin [2002]) by treating the missing-data indicators as random variables and assigning them a distribution as follows:

Define the complete data $Y = (y_{ij})$ and the missing data matrix $M = (M_{ij})$. The missing data mechanism is characterised by the conditional distribution of M given Y and unknown parameters ϕ i.e. $f(M|Y, \phi)$

- Missing completely at random (MCAR) : Missingness does not depend on the values of the data Y, missing or observed.

$$f(M|Y, \phi) = f(M|\phi) \ \forall \ Y, \phi \tag{4.25}$$

- Missing at Random (MAR): Missingness depends only on the components Y_{obs} of Y that are observed and not on the components that are missing, Y_{miss}. This is equivalent to saying that the behaviour of two units who share observed values have the same statistical behaviour on the other observations, whether observed or not.

$$f(M|Y, \phi) = f(M|Y_{obs}, \phi) \ \forall \ Y_{miss}, \phi \tag{4.26}$$

- Missing not at random (MNAR) The distribution of M depends on the missing values in the matrix Y. Even accounting for all the available observed information, the reason for observations being missing still depends on the unseen observations themselves.

4.3.2 Imputation and multiple imputation

Imputing values by any method and then assuming analyses for complete data sets can be used with impunity is a mistake (Little and Rubin [2002]). Imputations are means or draws from a predictive distribution of the missing values and require a method of creating a predictive distribution for the imputation based on the observed data. The two generic approaches are explicit modelling (formal statistical model e.g. multivariate normal with explicit assumptions) and implicit modelling (algorithm supplies an underlying model with implicit assumptions which need to be recognised and assessed for reasonablity).
Explicit methods:

- Mean Imputation: means from responding units are used, can be weighted.

- Regression Imputation: regression of the missing item on items observed for the unit, usually calculated from units with both observed and missing variables present.

- Stochastic regression imputation: replaces missing values by a value predicted

by regression plus a residual, drawn to reflect uncertainty about the missing value.

Implicit modelling:

- Hot deck imputation: substituting individual values drawn from 'similar' responding units. Literature on the theoretical properties of this approach are sparse (Little and Rubin [2002]).
- Substitution: used for unit non-response at the field work stage; an alternative unit is selected into the sample to replace the nonresponsive one. The substituted units are responders and may differ systematically from non-responders.
- Cold deck imputation: substitutes a constant value from an external source, e.g. a previous survey.
- Composite methods: e.g. combination of predicted mean for a regression with a residual randomly chosen from empirical residuals.

The weaknesses of these is that the inferences about parameters based on the filled-in data do not account for imputation uncertainty. Multiple imputation addresses both this drawback and the loss of efficiency.
To estimate imputation uncertainty:

- Apply explicit variance formulae that allow for non-response.
- Modify the imputations so that valid standard errors can be computed from a single filled-in data set.
- Apply the imputation and analysis procedure repeatedly to re sampled versions of the incomplete data. Uncertainty is estimated from the variability of point estimates of parameters from a suitable samples drawn from the original sample (bootstrap or jackknife, Efron and Tibshirani [1993]).
- Create multiply imputed data sets that allow additional uncertainty from imputation to be assessed. This provides consistent standard errors under broad classes of imputation procedures. Complete-data estimates and standard errors from each imputed data set are combined with between-imputation uncertainty derived from variability in estimates across the data sets.

Multiple imputation refers to the procedure of replacing each missing value by a vector of $D \geq 2$ imputed values. The D values are ordered in the sense that D completed data sets can be created from the vectors of imputations; replacing each

missing value by the first component in its vector of imputations creates the first completed data set, replacing each missing component with the second component in its vector creates the second completed data set and so on. Standard complete-data methods are then used to analyse each data set. When the D sets of imputations are random draws from the predictive distribution of the missing values under a particular model for non-response, the D complete-data inferences can be pooled to form one inference that properly reflects uncertainty due to non-response under that model (Little and Rubin [2002]). When the imputations are from two or more models for non-response, the combined inferences under the models can be contrasted across models to display e sensitivity of inference to models for non-response, a particularly critical activity when non ignorable non-response is being entertained.

4.3.3 Multiple imputation on the cardiovascular rehabilitation data

The missing data were imputed using the R package *MICE* (R Development Core Team [2011], van Buuren and Groothuis-Oudshoorn [2011]) which used chained equations and a Gibbs sampler to impute plausible values where data were missing and then used Rubin's rules to produce pooled estimates of parameters from the completed data sets. First it is necessary to ascertain whether the assumption that the data were missing at random (including missing completely at random) was suspect. Second, the form of the imputation model must be specified for each variable which had data missing, including capturing known relationships between the covariates. Next the set of variables to be included as predictors in the imputation was specified. In this case all of those found to be significant univariate predictors of survival were included. If any of the variables to be imputed were sum scores or other functions of incomplete variables, passive imputation could be implemented. The number of iterations of the algorithm needs to be specified and the convergence of the Gibbs sampler checked by plotting the mean and variance of the parameters against the iteration number and checking that different imputed values mix freely and without trend. The number of imputed data sets to be produced needs to be specified. In the cardiovascular rehabilitation work, the number of imputed data sets was 20, and the number of iterations was also set to 20, which is of the order recommended in the MICE documentation; this produced plots showing convergence of the Gibbs sample to similar values for each imputed variable within the data sets. Setting this too low can lead to p-values that are too low (van Buuren and Groothuis-Oudshoorn [2011]). If θ is a vector of unknown parameters of the multivariate distribution of the complete data Y, the chained equations methodology obtains the posterior distribution

of θ by sampling iteratively from the conditional densities:

$$\begin{aligned} P(Y_1 &\mid Y_{-1}, \theta_1) \\ &\vdots \\ P(Y_p &\mid Y_{-p}, \theta_p) \end{aligned} \tag{4.27}$$

where $\theta_1 \ldots \theta_p$ are the parameters of the conditional densities, $Y_1 \ldots Y_p$ are the p incomplete variables and $Y_{-j} = (Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p)$. The plausible data values imputed into each completed data set were drawn from a distribution specifically modelled for each missing entry. The estimates for the quantities of interest, in this case the parameters of the Cox model, were pooled into a single estimate and its variance is estimated. For quantities that were approximately normally distributed, within- and between-imputation variance could be given.

For the cardiac rehabilitation data predictive mean matching was selected as the imputation model for continuous variables and for factors with 2 levels logistic regression and with more levels, polytomous regression.
Details of the analysis results are found in Chapter 6 on page 103, and the diagnostic plots of the multiple imputation in Chapter 9 on page 252.

Chapter 5

Methods 2: Artificial Neural Networks

There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. (Breiman [2001])

The purpose of this chapter is to give an introduction to artificial neural networks (ANNs), a form of machine learning, including their use in survival analysis. The motivation for using artificial neural networks in these settings is that they have the capacity to capture non-linearities in the relationships between the variables and the outcomes for patients, and this may prove more appropriate than purely linear models.
Pattern recognition encompasses a wide range of information processing problems, such as the classification of hand written characters, identifying faces, and medical diagnosis and ANNs are often applied to such problems. Many humans solve these in a seemingly effortless fashion, but computer solution has proved to be immensely difficult in many cases (Bishop [1996]). ANNs have been applied in the study of cancer, diagnosis of pulmonary embolism, lung disease, breast cancer and differentiating benign from malignant pulmonary nodules (Joshi et al. [2003], Veropoulos [2001], Fukushima et al. [2004], Eng [2002], Wu et al. [1993], Matsuki et al. [2002]).

Statistical pattern recognition is a well established field which recognises the probabilistic nature both of the information to be processed and of the appropriate form in which the results should be expressed. The probability of the pattern

belonging to class $\mathcal{C}_k$ after the pattern vector $\mathbf{x}$ has been observed is given by the posterior probability $\mathcal{P}(\mathcal{C}_k|\mathbf{x})$. Where estimation of the probability density is impractical, suitable alternative discriminant functions can often be determined from the training data (Bishop [1996]). There are two separate stages in the classification process. The first is inference where training data is used to determine values for the posterior probabilities. The second stage is decision making where these are used to make decisions such as assigning a new data point to one of the possible classes.

The high dimensionality of pattern recognition problem data makes it impractical to store every possible combination of values, so a classifier must be designed to generalise, i.e. to classify correctly a previously unseen vector. Such systems can be described as a mapping from a set of input variables, $x_1, \ldots, x_d$ to output variables y_k, modelled in terms of a mathematical function containing a number of tunable parameters whose values are determined with the help of the data. The general form is $y_k = y_k(\mathbf{x}; \mathbf{w})$ where $\mathbf{w}$ denotes the vector of parameters, called weights in the context of ANN models. Both regression and classification problems can be seen as particular cases of function approximation; in classification problems the task is to assign new inputs to one of a number of discrete classes or categories, whilst in regression problems, the outputs represent the values of a continuous variables (Bishop [1996]).

Machine learning can be defined as follows: A computer program is said to learn from experience E with respect to some class of tasks T and a performance measure P, if its performance at tasks T, as measured by P, improves with experience E (Mitchell [1997]). Machine learning can be supervised or unsupervised. Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances (Kotsiantis [2007]). If one supposes that (X, Y) are random variables represented by some joint probability density $\mathcal{P}(X, Y)$, then supervised learning can be formally characterised as a density estimation problem where one is concerned with determining properties of the conditional density $\mathcal{P}(Y|X)$ (Hastie et al. [2009]). In unsupervised machine learning one has a set of N observations of a random p-vector X having joint density $\mathcal{P}(X)$. The goal is directly to infer the properties of the probability density $\mathcal{P}(X)$ without the help of a supervisor or teacher providing correct answers (Hastie et al. [2009]).

5.1 Artificial Neural Networks

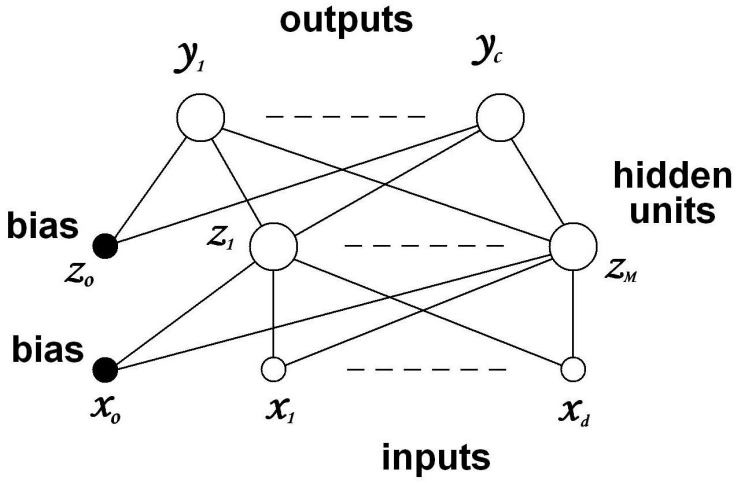


Figure 5.1: The Multi-layer Perceptron. (Bishop [1996])

Artificial Neural Networks belong to a class of learning methods that developed separately in the distinct fields of statistics and artificial intelligence, based on essentially identical models. The term ANN has come to encompass a large class of models and learning methods. Multi-layered networks having threshold or sigmoidal activation functions are called multi-layer perceptrons (MLP)(Bishop [1996]). An activation (or transfer) function acts on the linear function of the input variables to give a discriminant function. If the activation is the identity function the network collapses to a linear model in the inputs.

The importance of neural networks in the context of practical applications is that they offer a very powerful and very general framework for representing non-linear mappings from several input variables to several output variables, where the form of the mapping is governed by a number of adjustable parameters. The process of determining values for these parameters on the basis of the data is called learning or training, and for this reason the data set of examples is generally referred to as a training set. ANN models, as well as many conventional approaches to statistical pattern recognition, can be viewed as specific functional forms used to represent the mapping, together with particular procedures for optimising the parameters in the mapping. In fact, ANNs often contain conventional approaches as special cases. ANNs are an extension of conventional techniques, which builds on the many powerful results that the field of machine learning offers (Bishop [1996]).

ANN models represent non-linear functions of many variables in terms of superpositions of non-linear functions of a single variable, called hidden functions or hidden units. The hidden units are adapted to the data as part of the training process, so the number of such functions needs to grow only as the complexity of the problem grows. For a given number of hidden units, the number of free parameters typically grows linearly or quadratically with the dimensionality, d, of the input space, compared to d^M growth for a general Mth-order polynomial. When the output variables are continuous, the pattern recognition is a regression problem, and when discrete, a classification problem. In regression problems, the artificial neural network approximates the regression function, whilst in classification, the outputs of a artificial neural network can be interpreted as approximations to posterior probabilities of class membership.

MLPs with sigmoidal node functions are the most commonly used artificial neural networks. Each hidden node produces a hyperplane boundary in the multidimensional space containing the input data. The output node smoothly interpolates between those boundaries to give decision regions of the input space occupied by each class of interest. With a single logistic output, multilayer perceptrons can be viewed as non-linear extensions of logistic regression and, with two layers of weights, they can approximate any continuous function (Dybowski and Gant [2001], Bishop [1996]).

MLPs are regression or classification models typically represented by a network diagram as in figure 5.1. Some describe these as non-linear statistical models (Hastie et al. [2009]), whilst others argue that for a model to be statistical, assumptions about the variability of the data need to be made (Breiman [2001]). These tools are especially effective in high signal to noise settings where prediction without interpretation is the goal. They are less effective in problems where the goal is to describe the physical process that generated the data and the roles of the individual inputs. Each input enters into the model in many places, in a non-linear fashion. In general, the difficulty of interpretation has limited their use in fields like medicine where the interpretation of the models is very important.

In this study the MLP used consisted of two layers of adaptive weights with full connectivity between inputs and hidden units, and used sigmoidal activation functions. This architecture is capable of universal approximation, meaning that it can approximate any continuous function to arbitrary accuracy from a compact

region of input space provided the number of hidden units is sufficiently large, and provided the weights and biases are chosen appropriately (Nabney [2002]). There is a direct correspondence between a network diagram and its mathematical function, and general network mappings can be developed by considering more complex network diagrams. In this project, attention was restricted to feed-forward networks, which have the property that there are no feed-back loops in the network. In general, a network is feed-forward if it is possible to attach successive numbers to the inputs and all hidden and output units such that each unit received connections only from inputs or units having a smaller number. Feed-forward networks have the property that the outputs can be expressed as deterministic functions of the inputs, and so the network represents a multivariate non-linear functional mapping. Note that if the activation functions of all the hidden units in a networks are taken to be linear, then for any such a network there is an equivalent network without hidden units, since the composition of successive linear transformations is itself a linear transformation. If the number of hidden units is smaller than either the number of input units or the number of output units, then the linear transformation that such network generates is not the most general possible, as information is lost in the dimensionality reduction at the hidden units.

In a network with d inputs, M hidden units and c output units, the analytic function corresponding to the network can be expressed as follows:
The output of the jth hidden unit is formed with a weighted linear combination of the d input values and adding a bias (i.e. constant term) to give

$$a_j = \sum_{i=1}^{d} w_{ji}^{(1)} x_i + w_{j0}^{(1)}. \tag{5.1}$$

Here $w_{ji}^{(1)}$ denotes a weight in the first layer going between input i and hidden unit j, and $w_{j0}^{(1)}$ denotes the bias for the hidden unit j. By including an additional input variable $x_0 = 1$, the bias term for the hidden units can be absorbed into the weights to give

$$a_j = \sum_{i=0}^{d} w_{ji}^{(1)} x_i. \tag{5.2}$$

The activation of hidden unit j is then obtained by transforming the linear sum in (5.2) using an activation function $g_1(\cdot)$ to give

$$z_j = g_1(a_j). \tag{5.3}$$

The outputs of the two-layer network are obtained by transforming the activations of the hidden units using a second activation function. For each output unit, k, a linear combination of the outputs of the hidden units is constructed. Absorbing the bias as before gives

$$a_k = \sum_{j=0}^{M} w_{kj}^{(2)} z_j. \tag{5.4}$$

The activation of the kth output unit is obtained by transforming this linear combination using a non-linear activation to give

$$y_k = g_2(a_k) \tag{5.5}$$

where the notation $g_2(a_k)$is used to emphasise that the activation of output units need not be the same function as used for the hidden units. Combining (5.2), (5.3), (5.4) and (5.5) the explicit expression for the complete function represented by the network is

$$y_k = g_2 \left(\sum_{j=0}^{M} w_{kj}^{(2)} g_1 \left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i \right) \right) \tag{5.6}$$

The sigmoid activation function is

$$g(z) = \frac{1}{1 + (\frac{1}{b} e^{-\frac{z}{a}})} \tag{5.7}$$

where a is the slope parameter and b shifts the curve up or down. For fitting survival functions b is set to unity (Reeves and Johnston [2008]).

The weights are at first specified using domain knowledge or applied randomly to the inputs if there is none, and the output compared to the target output as given in the training data set. If the weights are near zero the operative part of the sigmoid is roughly linear, so setting weights near zero to start means the model starts out nearly linear and becomes non-linear as the weights increase; individual units localise to directions and introduce non-linearities as needed. An error function is defined to measure the degree to which the prediction and target differ. The credit assignment problem is the determination of which of the hidden units should be regarded as responsible for generating the error. When the activation functions of the network are differentiable, the activations of the output units become differentiable functions of both the input variables, and of the weights and biases. Evaluation of the derivatives of the error with respect to the weights allows the weight values which minimise the error function to be found (Bishop [1996]). The most common algorithm for evaluating these derivatives of the error function is called error back-

propagation. Most training algorithms involve an iterative procedure for minimising the error function, adjusting the weights in a sequence of steps: first evaluating the derivatives, then use the derivatives to compute the adjustments to be made to the weights. Then the errors are evaluated using the new weights, and the procedure repeated. Each iteration is one training cycle. One of the simplest and most common training algorithms is called gradient descent. In the batch version of the algorithm, the weight vector is initialised randomly and then the vector is iteratively updated in the direction of the greatest rate of decrease of the error. Weight changes are accumulated over an entire presentation of the training data (an epoch) before being applied. In the on-line or sequential version of the algorithm, the error function gradient is evaluated for just one pattern at a time, and the weights updated after the presentation of each training example (Bishop [1996]). The label 'on-line learning' can be misleading as it implies that learning may occur during use in the field. However, with both algorithms training is normally done off-line during a separate phase with controlled data sets (Wilson and Martinez [2003]). Nevertheless, the on-line approach does offer the opportunity to update the training with a small collection of new patterns, for example as clinical practices slowly change, without having to delay until an entirely new data set is collected.
A network with too many weights will overfit the data at the global minimum of the error function, so early stopping or weight decay regularisation is used (see page 99). However, it is better to have too many hidden units rather than too few or the model may not have enough flexibility to capture non-linearities and the extra weights can be shrunk toward zero if appropriate regularisation is used.
Multiple minima may be present in the error function so one option is to choose a number of starting values for the weights and select the solution giving the lowest penalised error

5.2 Artificial Neural Networks for Survival Analysis

Recall the survival analysis equations on page 73 and equation (5.6) above and the log likelihood equation

$$log\ L = \sum_{i=1}^{n} [\Delta_i\ log(f(t_i)) + (1 - \Delta_i)\ log(S(t_i))]. \tag{5.8}$$

Expressions for $S(t)$, the survivor function and $f(t)$, the probability density function of survival time T, are required. A single layer artificial neural network has survival

function

$$S(t_i) = g_2 \left(\sum_j w_{kj}^{(2)} \; g_1 \left(\sum_i w_{ji}^{(1)} x_i \right) \right) \tag{5.9}$$

Now if

$$z(t_i) = \sum_j w_{kj}^{(2)} \; g_1 \left(\sum_j w_{ji}^{(1)} x_i \right) \tag{5.10}$$

so that

$$S(t_i) = g_2 \left(z(t_i) \right) \tag{5.11}$$

then

$$z(t_i) = g_2^{-1} \left(S(t_i) \right) \tag{5.12}$$

and

$$S'(t_i) = g_2' \left(z(t_i) \right) z' \left(t_i \right) \tag{5.13}$$

from equation 5.11 by the chain rule and from 5.10

$$z'(t_i) = \sum_k g_1' \left(\sum_j w_{ji}^{(1)} x_i \right) w_j^{(2)} w_{pj}. \tag{5.14}$$

Define $x_d = t$ so $\frac{d}{dt} \sum_j w_{ji}^{(1)} x_i = w_{jd}$ giving

$$S'(t_i) = g_2' \left(z(t_i) \right) \sum_j g_1' \left(\sum_j w_{ji}^{(1)} x_i \right) w_{kj}^{(2)} w_{jd}^{(1)} \tag{5.15}$$

and $f(t_i) = -S' \left(t_i \right)$. The use of artificial neural networks for survival analysis has a relatively recent history (Ripley et al. [2004], Joshi et al. [2003], Joshi [2004]). Both classification and regression networks can be used for survival modelling. In classification, it is the probability of survival or time to relapse up to a given time point that is modelled. It is possible to choose a single time point, e.g. relapse within 5 years which is a simple binary output unit, or a series of time intervals, such as relapse in less than one year, one to two years, two to three years, etc. The multiple category cases may be fitted using a softmax function on the outputs which produces outputs which lie in the range [0,1] and sum to 1 so can be interpreted as probabilities. The Softmax function (Hastie et al. [2009]) is

$$y_k = \frac{e^{a_k}}{\sum_{k'} e^{a_{k'}}}. \tag{5.16}$$

Since the probabilities must add up to 1 for N categories, only $N-1$ of them can vary independently. The natural ordering of these times can be taken into account. In the regression model, the log likelihood is optimised (Joshi et al. [2005]) and the probabilities of survival can be expressed as a survival curve by predicting on a vector of times.

5.3 Optimising the Network

5.3.1 Minimising the error function

Non-linear activation functions are often used in artificial neural networks, and since hidden and output units perform different roles, the choice of activation function for the output units may differ from that for the hidden units. In a general feed-forward network, each unit computes a weighted sum of its inputs in the form

$$a_j = \sum_i w_{ji} z_i$$

where z is as defined in equation (5.3). Suppose that the error function can be written as a sum over all patterns in the training set of an error defined for each pattern separately,

$$E = \sum_n E^n$$

and that E^n is differentiable function of the network variables so that

$$E^n = E^n(y_1, \ldots, y_c).$$

The derivatives of the error function E with respect to the weights and biases in the network can be expressed as sums over the training set of the derivatives for each pattern separately. E^n depends on weight w_{ji} only via the summed input unit a_j to unit j. The derivative of E^n with respect to a weight w_{ji} can be obtained using the chain rule for partial derivatives

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}}. \tag{5.17}$$

The required derivative is obtained by multiplying the value of $\frac{\partial E^n}{\partial a_j}$ for the unit at the output end of the weight by the value of $\frac{\partial a_j}{\partial w_{ji}}$ for the unit at the input end of

the weight. For the output units

$$\frac{\partial E^n}{\partial a_k} = g'(a_k)\frac{\partial E^n}{\partial y_k} \tag{5.18}$$

by the chain rule on equation (5.3). For the hidden units we have

$$\frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k}\frac{\partial a_k}{\partial a_j} \tag{5.19}$$

where the sums run over all units k to which unit j sends connections, and using the fact that variations in a_j give rise to variations in the error function only through variations in the variables a_k. This leads to the back propagation formula

$$\frac{\partial E^n}{\partial a_j} = g'(a_k)\sum_k w_{kj}g'(a_k)\frac{\partial E^n}{\partial a_k} \tag{5.20}$$

which allows the evaluation of errors recursively. The use of the logistic sigmoid as an activation function is computationally efficient here, since its derivative can be expressed in the form

$$g'(a) = g(a)(1 - g(a)). \tag{5.21}$$

The dervatives of the error function with respect to the weights obtained in this way form the Jacobian matrix of partial derivatives. The derivatives of the outputs with respect to the inputs can also be calculated in a similar manner to form a Jacobian matrix which estimates the contribution of the errors associated with the input variables to the error of the output variables (Bishop [1996]). Back propagation can also be used to obtain the second derivatives of the error with respect to the weights to form the Hessian

$$\frac{\partial^2 E}{\partial w_{ji}\partial w_{lk}}. \tag{5.22}$$

The Hessian and its inverse plays an important role in neural computing which is detailed in Bishop [1996]. The inverse of the Hessian H of the error with respect to the weights can be approximated using the outer product approximation. If $R \equiv \nabla_w E$, is the gradient of the error function and N is the number of patterns in the data set, then the outer product approximation can be written

$$H_N = \sum_{n=1}^{N} R^n (R^n)^T \tag{5.23}$$

and the Hessian can be built up sequentially using

$$H_{N+1} = H_N + R^{N+1}(R^{N+1})^T. \tag{5.24}$$

Then this matrix identity (Kailath [1980]) can be used to provide the inversion:

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}. \tag{5.25}$$

where I is the identity matrix. Identifying $H_N = A,\ R^{N+1} = B,\ (R^{N+1})^T = C$ we have

$$H_{N+1}^{-1} = H_N^{-1} - \frac{H_N^{-1}R^{N+1}(R^{N+1})^T H_N^{-1}}{1 + (R^{N+1})^T H_N^{-1} R^{N+1}} \tag{5.26}$$

This represents a procedure for evaluating the inverse of a Hessian using a single pass through the data set. The initial matrix, H_0 is chosen to be αH where α is a small quantity. It is important to state that the outer product approximation for use with the sum of squares error function is only likely to be valid for a network trained on the same data set, or one with the same statistical properties as the one used to evaluate the Hessian. For a general network mapping, the second derivative terms will typically not be negligible. The Hessian of the error with respect to the weights can be evaluated exactly for a network of arbitrary feed-forward topology and with an differentiable error function using an algorithm based on back-propagation for the evaluation of first derivatives, detailed on page 157 of Bishop [1996].

For regression problems and for classification problems, the purpose of network training is to model the underlying generator of the data so that the best possible predictions of the target t are made when the trained network is presented with a new input vector x. For associative prediction problems of this kind, it is convenient to decompose the joint probability density in the product of the conditional density of the target data, given the input data and the unconditional density of the input data thus:

$$\mathcal{P}(x, t) = \mathcal{P}(t|x)p(x). \tag{5.27}$$

Many error functions can be motivated from the principle of maximum likelihood. For training data $\{x^n, t^n\}$, the likelihood can be written as

$$\mathcal{L} = \prod_n \mathcal{P}(x^n, t^n) \tag{5.28}$$

$$\mathcal{L} = \prod_n \mathcal{P}(t^n|x^n)\mathcal{P}(x^n) \tag{5.29}$$

under the assumption that each data point (x^n, t^n) is drawn independently from the same distribution. It is generally more convenient to minimise the negative log likelihood than to maximise the likelihood, and these are equivalent since the negative logarithm is a monotonic function. Fitting an ANN by maximum likelihood is known as 'entropy' fitting and is not common (Ripley and Ripley [2001], Joshi et al. [2005]).

$$E = -ln\mathcal{L} = -\sum_n ln\ \mathcal{P}(t^n|x^n) - \sum_n ln\ \mathcal{P}(x^n). \tag{5.30}$$

The second term in equation (5.30) does not depend on the network parameters and is therefore an additive constant which may be omitted from the error function.

These procedures have the advantage that they constrain the weights assigned to the variables during learning to have values falling within the same interval.

5.3.2 Model complexity

The trained network model's ability to generalise is important for making good predictions on new inputs; this requires model complexity to be optimised. As with other statistical models, this requires a trade-off between bias and variance. One way to achieve this in the artificial neural network setting is to compare the performance of networks with a varying number of hidden units. A large data set is advantageous here; as the number of data points grows, more complex models can be supported, so reducing bias whilst ensuring the data constrains the model, simultaneously reducing the variance. In practice, the data available are usually limited. Adding a penalty term to the error function is a particular example of the concept of regularisation (Hastie et al. [2009], Bishop [1996]) and is another way of controlling model complexity. The general form is

$$\tilde{E} = E + \nu\Omega \tag{5.31}$$

where E is the error function of choice, and ν controls the extent to which the penalty term Ω constrains the form of the solution. One of the simplest forms of regularisation is weight decay and consists of the sum of squares of the adaptive parameters (weights) in the network. The minimisation of the error function determines the values for the free parameters in a network, but cannot indicate the optimal number of such parameters, or equivalently the optimal size of the network. As with the trade-off between bias and variance for curve-fitting, the goal of training the data is to produce a network with good generalisation performance on new input vectors, and this is typically not the network which gives the smallest error on the training

data. Networks with too little flexibility will smooth out the structure in the data, corresponding to high bias, whilst very complex networks will over-fit the data corresponding to high variance. In both cases, generalisation is poor. This is where the determination of the regularisation coefficient in equation (5.31) becomes important. A too large value for ν produces a network with large bias, whilst a too small value allows the network to have high variance. Direct minimisation of $\tilde{E}$ leads to $\nu = 0$ which is an over-fitted solution. In the case of weight decay we have

$$\Omega = \sum_i w_i^2 \tag{5.32}$$

with the sum over all weights and biases. When used in curve fitting this form of regularisation is called ridge regression. To produce an over-fitted mapping with regions of large curvature requires large weights. Weight decay encourages weights to be small and avoid the overfitting. Other alternatives include early stopping, when training is stopped at the point where the error on a validation data set stops falling and begins to grow. Curvature driven smoothing is another technique for directly penalising curvature and uses second derivatives. Adding random noise to input vectors has also been shown to lead to improvements in network generalisation (Bishop [1996]). The most common pre-processing is linear rescaling of inputs so that they all lie on a common scale-dividing each variable by the maximum value it could take ensured all the variables fall in the interval zero to one. It is also possible to rescale so that each variable has mean zero and variance one. Rescaling becomes particularly important if weight penalty is used, since all the inputs need to be in the same range to avoid over-penalizing the weights associated with large input values.

5.3.3 Cross validation

Cross-validation is a common mechanism for assessing the optimality of the model; part of the data set is reserved for validating, some for testing and the remainder used for training (Bishop [1996], Dybowski and Gant [2001], Hastie et al. [2009]). When the network has been trained using the training set, then optimisation of the model is made using the validation set. The efficacy of the model in predicting new cases is then reported using the independent test set. Often many partitions are used and the accuracy of various candidate models judged on the average performance of this procedure. The validation set is used to select ν by plotting the training and validation errors against ν. The errors on the training set continue to fall, since that is what the training algorithm requires. It is typical to see that, at a certain value

of ν the error on the validation set ceases to fall and begins to rise. The value of ν where the error is minimum on the validation set is the optimal value and is selected for use in training the network. A similar technique can be used when varying the number of hidden nodes.

5.3.4 The black box

One criticism levelled at neural networks is that they are black box systems by which it is meant that the manner in which artificial neural networks derive an output value from a given feature is not readily comprehensible, making the use of the output from artificial neural networks unacceptable to users. Black box methods such as artificial neural networks, which can be quite useful in purely predictive settings such as pattern recognition, are far less useful for data mining (Hastie et al. [2009]). In a model, desirable characteristics are accuracy and interpretability. Accuracy means the model estimate of the value must be close to the true value, and interpretability means that the input - output relationships that can be extracted from the model are comprehensible to the intended user of the model. There are several types of interpretation: how each input affects the output value, as seen in regression coefficients; input-output relationships as if-then rules, as seen in tree-structured classifiers; how the output value is obtained from the input vector as in the most probable configuration in Bayesian Belief networks (Cooper [1990]). Interpretability is desirable if it uncovers previously unknown but useful input-output summary, or discloses an error in the model. If accurate prediction is what the model is required to produce, and careful, extensive testing and ongoing monitoring provide reassurance of accuracy, the lack of intepretability is less important than if the goal is knowledge discovery. Interpretation of the mass of weights and connections within the network is very difficult when trying to extract previously unknown but useful information from data. This is where saliency (the relative importance of weights, Bishop [1996]) can be used, not only for optimising network structure and performance, but also for giving some intuition about which inputs contribute most to the accuracy of the network predictions. There has been much effort directed at learning rules from trained artificial neural networks (Dybowski and Gant [2001]) for this reason.

5.3.5 Applying the ANNs

The artificial neural network for survival analysis used here takes as inputs the variables found to be significant in the optimal Cox survival analysis, categorised as in Table 6.24 on page 152. The weights were optimised using the maximum likelihood of the function given in Equation 4.17 on page 75. The weight decay and number of hidden nodes was optimised in MATLAB using training sets of 9/10 of the data, selected at random, and a test set of the remaining data. After training, the log likelihood was calculated for the test set. Hidden nodes greater than 3 could not be supported because of lack of data, but 1, 2, and 3 hidden units were tested. The weight decay values 1, 0.1, 0.01, 0.001, 0.0001, 0 were tested with each of the hidden unit configurations and the optimal combination was found to be a weight decay of 0.1 and 3 hidden nodes. A vector of times (scaled by dividing by maximum time to be between 0 and 1) was entered into the optimised network and a vector of values between 0 and 1 returned. These are interpreted as probabilities of survival and were compared the empirical Kaplan- Meier Survivor function and the survivor function obtained from the Cox proportional hazards to assess model fit, as shown in Figure 6.18 on page 160.

The inputs to the artificial neural network for predicting recovery from back pain were treatment allocation plus the same variables used for building the latent class model, listed on page 106, each score divided by the maximum to scale them between 0 and 1. Achievement of a 3-point improvement in RMQ score, represented as a dichotomous variable, was used as the target. The weights were optimised using maximum likelihood using Equation 6.1 on page 125. The network architecture and the weight decay were optimised using 10-fold cross validation scores for percentage error and the log score given in Equation 6.2 on page 118, and those with minimal values retained. Weight decays 0, 0.0001, 0.001, 0.01, 0.1 and 1 were tested along with architectures with 1-5 hidden units. Details are given in Table 6.14 on page 127. A network with 1 hidden unit and a weight decay of 0.1 was optimal. The output of the model is a value between 0 and 1 with values 0.5-1 interpreted as the model predicting recovery for this patient, and as predicting non-recovery for the lower values. The model fit is assessed by percentage accuracy on the fitted network, the log score, the sensitivity and the specificity as in Table 6.15 on page 131.

The receiver operating characteristic (ROC), or ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. The ROC curve illustrates the performance of the ANN as

the value at which predicting recovery / non-recovery is varied. There are four possible outcomes from the classification: a recovering patient can be classified by the model as recovering (a true positive) or not recovering (a false negative) and a non-recovering patient can be classified as recovering (false positive) or non-recovering (true negative). The true positive rate or sensitivity is the proportion of all the recovering patients that are classified as recovering by the model, calculated as true positive / (true positive + false negative). Similarly, the true negative rate or specificity is true negatives / (true negatives + false positives). The ROC curve illustrates the variation in these two quantities as the threshold value used by the model is varied. A line of slope 1 indicates that the model has no discriminatory ability, and that cases are allocated no better than at random.

In medical applications, the relative importance of avoiding false positives and false negatives depends on the specific illness and treatment options under consideration. Clearly, the ideal scenario is perfect classification, but in real-world applications this is unlikely. When used for diagnosis, or other decision support about treatment, correct classification leads to correct (or best) treatment and cure, whilst incorrect treatment can sometimes lead to serious consequences, either from failure to give correct treatment leading to disease progression, or the suffering of serious side-effects without the pay-off of curing the disease. The overall error rate of the classifier is a useful measure of the performance of a classifier and there are others (Hand [2010]).

Chapter 6

Results

This chapter sets out to describe and compare the results obtained by applying the different analyses outlined in Chapter 4 and Chapter 5 to the BeST data and the the Basingstoke and Alton Data. Both data sets were introduced in Chapter 3,

The BeST data from the complex intervention for back pain were analysed using latent class analysis (described in Chapter 4) in an attempt to find classes of patients for whom the response to the intervention differed either positively or negatively. The aim was to facilitate the tailoring of interventions to patients by identifying the subgroups of patients for whom this treatment is particularly effective or particularly ineffective. The manifest (observed) variables used to identify the latent variable were those which the intervention was designed to tackle. The BeST data were also subjected to a classification artificial neural network (ANN, described in Chapter 5), to categorise patients according to whether they experienced significant improvement or not, based on these same input variables. The methods were compared for their ability to identify groups with similar outcomes. The effect of missing data was assessed using multiple imputation to complete the data sets. Comparison was made of the analysis based on the data with the cases with missing variables omitted, and with the missing values multiply imputed.

The Basingstoke and Alton cardiac rehabilitation data were analysed using a Cox proportional hazards model, to identify those covariates which are significant predictors of long-term survival. These data were also subjected to a continuous-time artificial neural network analysis which produced a survival curve. A comparison was made of the two curves' fit to the data and hazard rates.

6.1 Analysis of BeST Back Pain Trial data

The Best Data was analysed by three methods, each using the same explanatory and response variables. Logistic regression was used to predict a minimum RMQ improvement from the explanatory variables. Latent class analysis was used to form classes of patients using the explanatory variables, and the optimal model was used as the explanatory variable in a logistic regression to predict a minimum RMQ improvement. Finally, an ANN was trained using the explanatory variables and used to predict the same minimum RMQ improvement.

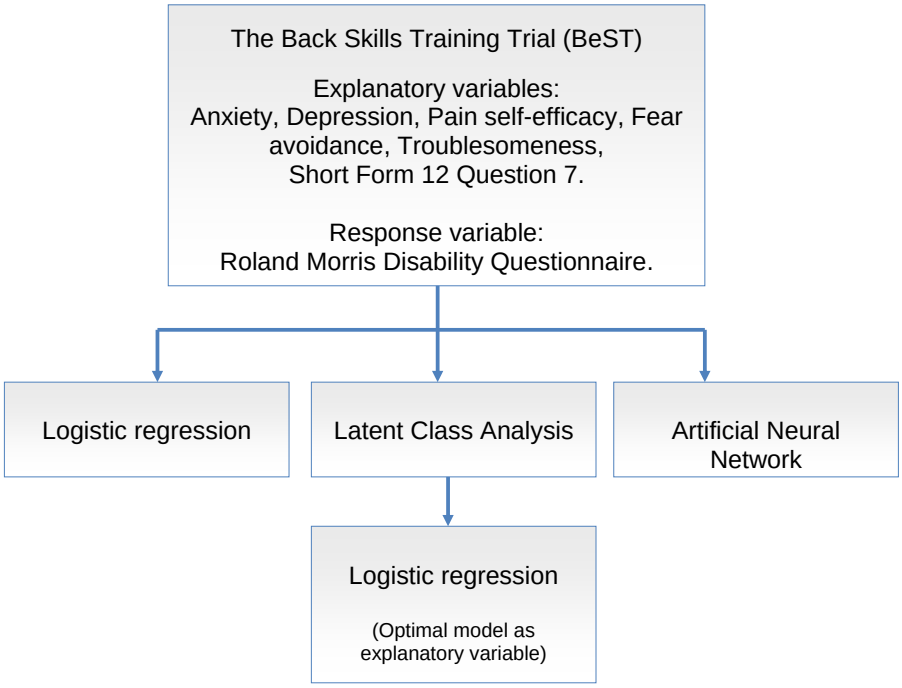


Figure 6.1: Schematic representing the analysis scheme for the BeST data set.

6.2 Latent Class analysis of BeST Back Pain Trial data

6.2.1 Principal findings

Scores which measure the characteristics that the intervention was designed to tackle were successfully used to classify trial participants into groups. For trial participants who received the intervention treatment, there was an association between outcome and class membership in the three-class model. After adjusting for age, sex and work status in a logistic regression model, there was a marginally significant coefficient for the interaction between treatment and membership of class 3 in the three-class model.

6.2.2 Introduction

The BeST Trial demonstrated that a group-based cognitive behavioural approach (CBA), delivered by a range of health professionals, had a sustained effect on chronic non-specific low back pain. This compared favourably with the range of treatments recommended for the early treatment of persistent back pain by the National Institute for Health and Care Excellence (NICE) guidelines (NICE [2009]).
The design, intervention, and main analyses of the BeST Trial were reported in detail by the trial team in (Lamb et al. [2010b]), and summarised in Chapter 3 starting on page 44. Underwood and colleagues undertook a secondary analysis on this data to look for effect moderators (Underwood et al. [2011]) and found that neither troublesomeness nor fear avoidance moderated treatment effect on any of the primary outcomes. The only moderation by baseline variables of the effect of the intervention was on the RMQ outcome: being younger and currently working both moderated treatment effect, resulting in larger improvements as a response to treatment.

The analysis approach was to first find the best latent class model fit, and so to ascertain the optimal number of classes for this data set. Then the relationship between class membership and outcome for each of the intervention and control cohorts was investigated using logistic regression, analysis of deviance and Fishers' exact test. Finally, an adjusted logistic regression was used to control for other variables of interest, namely age, sex and work.

6.2.3 Analysis approach

Latent class analysis was implemented on the full BeST data set using only cases where 12-month follow-up RMQ score was available. Given the context of a cog-

nitive behavioural approach to the treatment of back pain, after discussion with a clinical colleague it was agreed that the most appropriate observed variables for the latent class modelling would be those which the intervention sought to modify. The complete cases were used for the analysis.

The observed items that the intervention tackled are:

- SF12 question 7 i.e. 'During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities, (like visiting with friends, relatives, etc.)?' and responses are 1. All of the time; 2. Most of the time; 3. Some of the time; 4. A little of the time; 5. None of the time. See page 141 of Lamb et al. [2010b] (SF12_7).
- Pain self-efficacy total score (PSE).
- Fear Avoidance Beliefs Questionnaire, items 2-5 relating to physical activity (PA).
- Hospital Anxiety and Depression Scale (HADS)- using the separate scores: HADS depression and HADS anxiety.
- Troublesomeness: moderately; very; extremely.

Each score was calculated according to its questionnaire manual detailed on page 46.

Recall that latent class analysis requires categorical variables (see page 67). The two subscales of anxiety and depression measured by HADS were split according to their standard interpretation, i.e. that scores lower than 8 were considered to indicate a patient was not depressed (anxious), 8 to 10 that they were 'borderline' and over 10 that they were depressed (anxious). The HADS anxiety and depression scales were the most informative in distinguishing between classes.
Similarly, the same cut points were adopted for PA as in the trial, i.e. a score of under 14 was considered to indicate that a patient was not fear avoidant, whist a score of 14 and over that a patient was fear avoidant.
There was no obvious or standard categorisation known for the remaining score, PSE. Some time was spent undertaking sensitivity analysis, and it was decided that medical opinion was a better guide to appropriate discretisation than optimising on the basis of this data set, since any cut points found by the latter method may not generalise to other data sets. It was therefore agreed, after discussion with a

clinical colleague, to split PSE into very low, low and better than low (scores 0-20, 21-30, 31-60) because it made no clinical sense to differentiate between high and very high efficacy patients, but the difference between low and very low may have clinical importance in distinguishing between patient outcomes. SF12_7 was used as it stands, i.e. with a separate category for each of the 5 possible responses. Troublesomeness was also used in its 5 categories, noting that recruitment to the trial required patients to have at least moderately troublesome pain, effectively reducing the categories in use to 3.

A subset was made of patients for whom all the 6 observations detailed above plus baseline and 12-month follow-up RMQ scores were complete. The data losses were as follows in Table 6.1.

Variable	**complete cases out of original 701 (%)**	**cases remaining after sequential removal of missing (%)**	**item level missingness** (%)	**case level missingness** (%)
Fear Avoidance	662 (94.4)	662 (94.4)	34 (4.9)	5 (0.7)
Anxiety	686 (97.9)	648 (92.4)	13 (1.8)	2 (0.3)
Depression	693 (98.9)	645 (92.0)	6 (0.8)	2 (0.3)
SF12 question 7	689 (98.3)	635 (90.6)	NA	12 (1.7)
Pain Self efficacy	676 (96.4)	620 (88.4)	22 (3.1)	3 (0.4)
Troublesomeness	638 (91.0)	567 (80.9)	NA	63 (9.0)
RMQ at baseline	700 (99.9)	567 (80.9)	0 (0.0)	1 (0.1)
RMQ 12 months	498 (71.0)	407 (58.1)	0 (0.0)	203 (29.0)

Table 6.1: Variables: cases complete in individual cases and after sequential removal. SF12_7 and Troublesomeness are single-item scores so there is no distinction between item-level and case-level missingness.

Outcome was defined as a 3-point or greater improvement in RMQ score at 12 months or a score of less than or equal to 3 at baseline, maintained at 12-month follow-up, to include those whose scores may not allow for an improvement of 3-points, but are well (Bombardier et al. [2001], Lauridsen et al. [2006]). This is the definition of 'significant improvement' in RMQ reported throughout this thesis. There were 407 such patients, 181 with a RMQ improvement described above, and 226 without such an improvement. The numbers and proportions of patients in each category of each score was given in Table 3.1 on page 51. A latent class analysis algorithm (R package *poLCA* Linzer and Lewis [2011a]) was run on this complete cases data set, and models with 2, 3, 4, 5 and 6 classes were produced. The models were evaluated by considering the Akaike information criterion (AIC, Akaike [1974]) and Bayesian information criterion (BIC, Schwarz [1978]) as detailed on page 78.

6.2.4 The models

Models with 2 to 6 classes were built and compared. The model with the minimum AIC was the 4-class model while the model with the lowest BIC was the 3-class model. For comparison, the AIC and BIC values of the 5 models is given in Table (6.2).

Number of classes	2	3	4	5	6
AIC	4216.2	4116.8	4104.7*	4109.6	4120.6
BIC	4340.4	4305.2*	4357.2	4426.3	4501.5

Table 6.2: BIC and AIC for latent class models with 2 to 6 classes. Asterisk indicates lowest information criterion.

6.2.5 Patient characteristics for the three and four class models

Typical characteristics of the patients in the three classes were:

- Class 3.1: Patients in this class were generally those whose back pain caused no interference with social activity; were not fear avoidant; had high confidence in managing despite any pain they experienced; were not clinically anxious; were not clinically depressed and whose back pain was moderately troublesome.

- Class 3.2: Patients in this class were generally those whose back pain caused significant interference with social activity; were fear avoidant; had low confidence in managing despite any pain they experienced; were clinically anxious; were clinically depressed and whose back pain was extremely troublesome.

- Class 3.3: Patients in this class were generally those whose back pain caused moderate interference with social activity; were fear avoidant; had high confidence in managing despite any pain they experienced; were borderline or clinically anxious; were borderline or not depressed and whose back pain was moderately to very troublesome.

This list may give the impression that the patient characteristics were clearly defined for each class. However, the estimated class-conditional response probabilities reported by the poLCA software are not 0 or 1 , and in some cases the probabilities in two or more levels of a variable can differ only slightly. For example, in the 4-class model, Class 4 had estimated class-conditional response probabilities 0.16 for not fear avoidant and 0.84 for fear avoidant which are clearly distinct, whilst Class 2 had 0.53 for not fear avoidant and 0.47 for fear avoidant which are roughly equal. Taking all six predictor variables into account, with between 2 and 5 levels in each predictor, the distributions are generally more complex than can be captured in simple descriptions. The colormaps (Figures 6.2, to 6.8) are a useful way to capture the estimated class-conditional response probabilities. The most positive level of any predictor (high confidence, not fear avoidant, not anxious or depressed, etc.) is on the left of the map with the levels ordered towards the most negative level on the right.

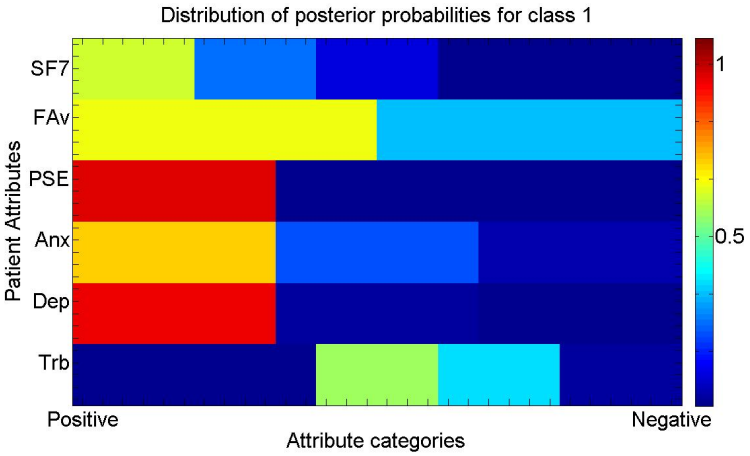


Figure 6.2: Three-class model: Class 1 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

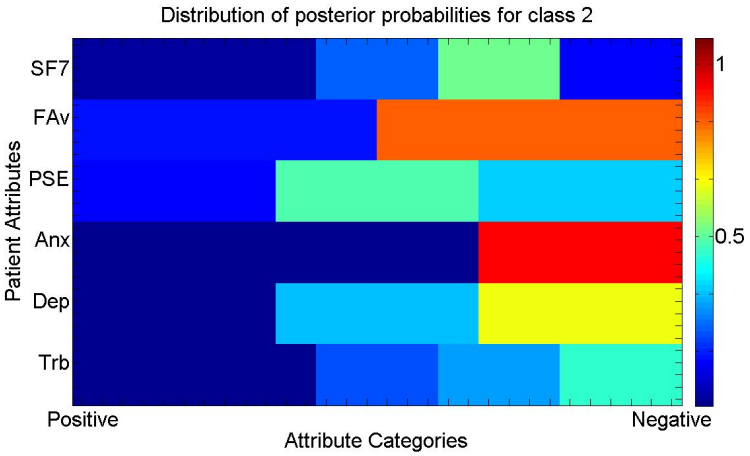


Figure 6.3: Three-class model: Class 2 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

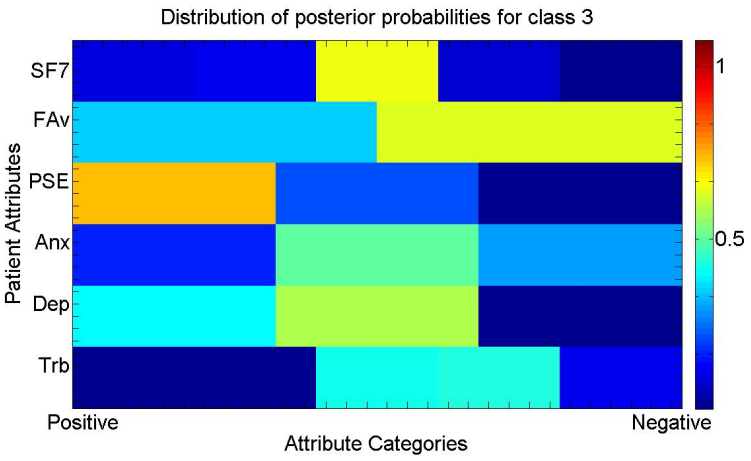


Figure 6.4: Three-class model: Class 3 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

Typical characteristics of the patients in the four classes were:

- Class 4.1: Patients in this class were generally those whose back pain caused little or no interference with social activity; were not fear avoidant; had high confidence in managing despite any pain they experienced; were not clinically anxious; were not clinically depressed and whose back pain was moderately troublesome.

- Class 4.2: Patients in this class were generally those whose back pain caused moderate interference with social activity; could be either fear avoidant or not fear avoidant; had low confidence in managing despite any pain they experienced; were borderline or not clinically anxious; were borderline clinically depressed and whose back pain was very troublesome .

- Class 4.3: Patients in this class were generally those whose back pain caused moderate interference with social activity; were fear avoidant; had high confidence in managing despite any pain they experienced; were borderline to clinically anxious; were borderline depressed and whose back pain was moderately to very troublesome.

- Class 4.4: Patients in this class were generally those whose back pain caused significant interference with social activity; were fear avoidant; had low or very low confidence in managing despite any pain they experienced; were clinically anxious; were clinically depressed and whose back pain was extremely troublesome.

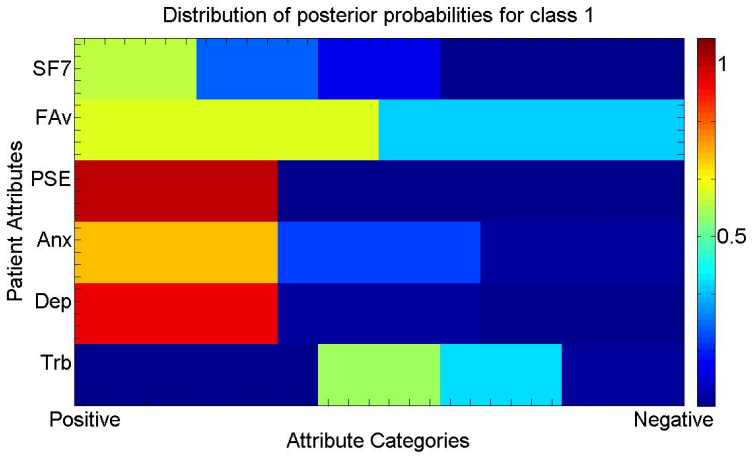


Figure 6.5: Four-class model: Class 1 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

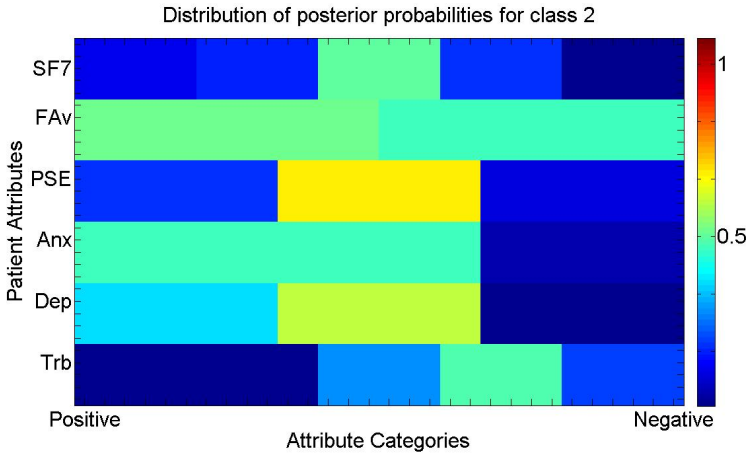


Figure 6.6: Four-class model: Class 2 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

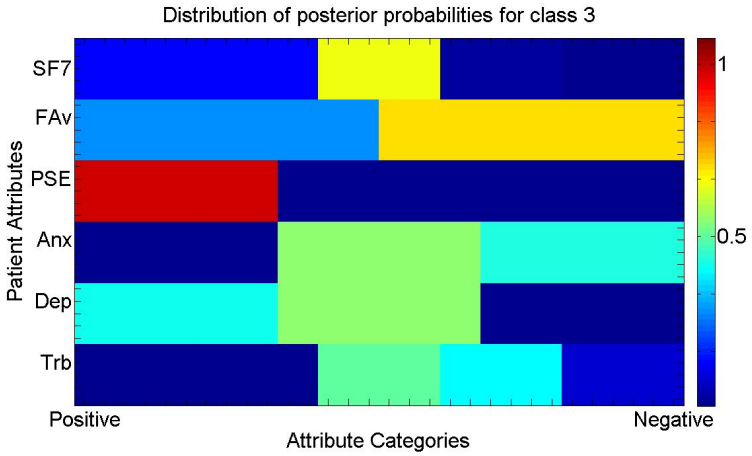


Figure 6.7: Four-class model: Class 3 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

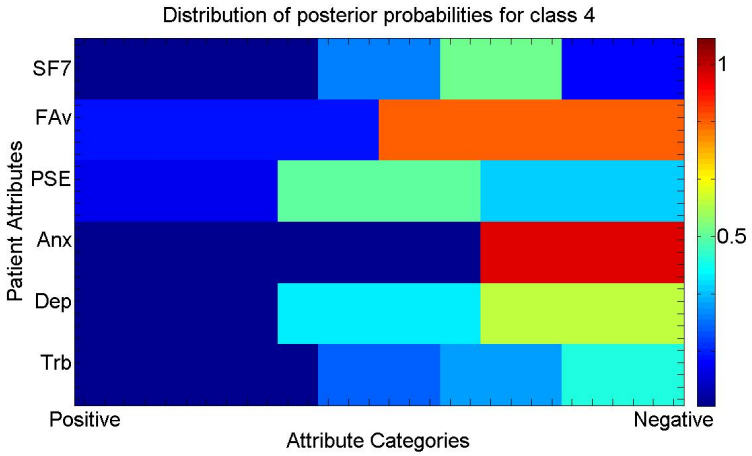


Figure 6.8: Four-class model: Class 4 showing the distribution of the probabilities of responses under each score, with the healthiest patients to the left, and the least health to the right.
The colours near to the red end of the spectrum represent a high probability (near to 1) of having this attribute given the class membership, and those near blue a low probability.

The class membership which was most probable for each individual patient was attached to their record, and associations between outcome and group membership were tested using Fisher's exact test, logistic regression and analysis of deviance. Fisher's exact test, like a χ^2 test, tests for an association between the rows and columns of a table, in this instance, the class and the outcome. Fisher's exact test is valid for all sample sizes. A small p-value suggests an association, and in this analysis a p-value of 5% or lower was used to indicate significance and a p-value of between 5% and 10% was considered marginally significant. Logistic regression predicts the log odds of the outcome as a linear model in the predictors. The p-value given is for the effect of a unit change in the predictor on the log odds significantly different from 0. Note that if the log odds is zero then the odds is 1, and there is no difference between the chance of recovery and the chance of no recovery. The analysis of deviance gives the reduction in residual deviance of adding each predictor variable in turn to the logistic regression model. As with analysis of variance for ordinary least squares regression, it is a measure of how much variability in the outcome is explained by the predictor in question. After attaching the most probable class membership to each case, the patient data were split into two groups according to whether the patients had received the cognitive behavioural approach (CBA) intervention or not, and Fisher's exact test was performed to test for an association between class and outcome. In the case of the Three-class model, Fisher's exact test showed an association between outcome and class (at 5% level) in the case of the CBA (intervention) patients. The 2, 4, 5 and 6-class models did not show an association between outcome and class under Fisher's exact test (at the 5% level) for the CBA (intervention) patients. In the case of the best care only (control group) patients, there was no association between outcome and class for any of the models. The p-values for Fisher's exact test are given in Table 6.3.

Number of classes	2	3	4	5	6
Intervention	0.44	0.05*	0.20	0.32	0.07
Control	1.0	0.70	0.72	0.68	0.23

Table 6.3: Fisher's exact test for association between class membership and outcome. * Significant at 5% level.

The next step was to try to identify which specific classes of patients were more likely to recover and which were less likely to recover. To do this, the models with the best relative goodness of fit were analysed more closely.

Testing the Four-Class Model

The four-class model, having the lowest AIC, was investigated to see if patients in some classes did better with the CBA treatment than patients of the same class who did not receive this treatment. Analysis of deviance was used to test if an interaction term was significant in the logistic regression model.
In the case of the four-class model, numbers of patients in each class were as shown in Table 6.4. Note that Class 2 had fewest patients, just 10% of the total data set, 32 in the intervention group and 10 in the control group.

Class:	1	2	3	4
Patients	210	42	101	54

Table 6.4: Patients in the four-class model. Note that the numbers of patients in this table were from across both arms of the trial, i.e. both those who had the CBA treatment, and those who did not, 407 altogether.

Analysis of deviance showed that treatment was significant in explaining outcome, but the class and interaction between class and treatment were not. This means that receiving treatment given that a patient was in a certain class was not a clear predictor of outcome for the four-class model.
A logistic regression model was used to predict outcome (recovery, as defined on page 108) by treatment and class. The coefficient for treatment was significantly different from zero (p=0.004) for the control group. The coefficients for class membership or the interaction between class and outcome was not significantly different from zero in the regression model.
Since the objective of the analysis was to uncover classes of patients with a common response to the treatment, when the standard contrasts did not give the required level of detail, new, detailed contrasts were constructed, each with one degree of freedom, i.e. being in Class 2 is contrasted with not in Class 2, Class 3 contrasted with not in Class 3, and so on. The contrasts were re-defined to keep the model deviance the same but split the contrasts into one-degree-of-freedom items. In this way, membership of each class was in turn compared to membership of any of the other classes, and the capacity to predict a significant improvement in RMQ (recovery, as defined on page 108) assessed. The analysis of deviance showed that treatment was significant in predicting outcome ($p=6.07 \times 10^{-5}$) but that class and the interaction between class and treatment were not significant.
The patients were split into two groups according to whether they received the CBA intervention or not, and Fisher's exact test was performed to assess if there was any association between class membership and outcome for each intervention group

(results in Table 6.3). Table 6.5 gives numbers in each class in the CBA arm of the trial. Recall that the patients were split 2:1 in favour of the intervention so that splitting patients into groups at random should yield approximately 67% of any class in the treatment arm and 33% in the control arm if there were no effect of class.

Class	**Intervention**			**Control**		
		Recovery			Recovery	
	Number (%)	No	Yes	Number(%)	No	Yes
1	137 (65)	71	66	73 (35)	49	24
2	32 (76)	14	18	10 (24)	8	2
3	68 (67)	28	40	33 (33)	22	11
4	43 (80)	26	17	11 (20)	9	2

Table 6.5: Four-class model: Number and percentage of each class who did and did not recover in each arm of the trial. Recovery is defined in 6.2.3. Recall that the patients were split 2:1 in favour of the intervention.

In all the above analyses, treatment is a predictor of recovery, but neither class membership or receiving treatment given that a patient is in a certain class was a clear predictor of outcome for the four-class model.

Testing the three-Class Model

The three-class model, having an AIC measure almost as low as the four-class model, and a lower BIC was also investigated to see if patients in some classes did better with the CBA treatment than patients of the same class who did not receive this treatment.
In the case of the three-class model, numbers of patients in each class are in Table 6.6.

Class	1	2	3	All
Patients	211	52	144	407

Table 6.6: Patients in the three-class model. Note that the numbers of patients in this table were from across both arms of the trial, i.e. both those who had the CBA treatment, and those who did not.

Analysis of deviance showed that treatment was highly significant ($p=6.07 \times 10^{-5}$) in explaining outcome and also that class was marginally significant (i.e. significant at 10% level, p=0.066) in explaining outcome, but the interaction between class and treatment was not. This means that receiving treatment given that a patient is in a certain class was not a clear predictor of outcome for the three-class model. Logistic regression was used to predict outcomes by treatment and class. The coefficient for treatment in the best care only arm, was significant, but the coefficient for class

membership and for the interaction between class and outcome were not significantly different from zero.
Once again, the contrasts were refined into contrasts each with one degree of freedom, and a logistic regression model fitted. Treatment was significant in the best care only group in the logistic regression model. The analysis of deviance showed that treatment is highly significant in the model. It also revealed marginal significance for membership of Class 2 (at 10% level, p=0.065). It should be noted that Class 2 is the smallest class with 52 patients (12.8% of the whole data set).
The patients were split into two groups according to whether they received the CBA intervention or not, and Fisher's exact test performed to assess if there was any association between class membership and outcome for each intervention group. Table 6.7 gives numbers in each class in the both arms of the trial.

Class	**Intervention**			**Control**		
		Recovery			Recovery	
	Number (%)	No	Yes	Number(%)	No	Yes
1	138 (65)	73	65	73 (35)	49	24
2	41 (79)	25	16	11 (21)	9	2
3	101 (70)	41	60	43 (30)	30	13

Table 6.7: Three-class model: Number and percentage of each class who did and did not recover in each arm of the trial. Recovery is defined in 6.2.3. Recall that the patients were split 2:1 in favour of the intervention.

In all the above analyses, treatment is a predictor of recovery, and membership of class 2 is a predictor of recovery using a 10% significance level but receiving treatment given that a patient is in a certain class was not a clear predictor of outcome for the three-class model either. Class 2 had the fewest patients in it, and whilst there is some indication that membership of Class 2 affects the likelihood of recovery this result could also be an artifact of the low numbers in the class. The proportion of patients in Class 2 in the no-recovery category for both intervention and control patients is much higher than the proportion in the recovery column. The difference in proportions between those who recovered and who did not is far greater in Class 2 than in either of the other two classes, but with so few patients, we cannot rule out that this occurred at random and is not a true effect.

6.2.6 Adjusting for other variables

To investigate the influence of the important variables identified in Underwood et al. [2011], age, sex and work were added to the logistic regression model.
Age was categorised into 5 groups of approximately equal size in the full data set of

701 cases:

Age group	number of patients
under 40	146
40-49	147
50-59	160
60-69	150
70 and over	98
Total	701

Table 6.8: Age categories, showing the number of patients in each category of the full data set of 701 cases.

The question relating to work is: 'are you currently working (either self employed or in paid employment)?'

For all the latent class models from the two-class model up to the six-class model, there was a significant interaction between treatment and work in the analysis of deviance. Treatment was significant in every model. In the two models of interest, the three- and four-class models, the four-class model had a marginally significant interaction between membership of Class 2 and treatment, both in the model with regular contrasts and the model with refined contrasts (in Class 2 compared to not in Class 2). In the three-class model, there was a marginally significant interaction between membership of Class 3 and treatment, again both in the regular contrast model and the refined contrast model. This suggests that, having adjusted for age, sex and work status, work modifies the effect of treatment (see Table 6.11.

These analyses support the findings of Underwood and colleagues, who found that working moderated the treatment effect, resulting in larger improvements as a response to treatment. In addition, for the two models of interest, there is some indication that class membership moderated treatment effect, although the evidence for this is modest.

It is of interest to investigate whether the variables age, sex or work status were represented equally within the classes or whether there is a pattern of the working people, who were marginally more likely to recover, also being the majority constituents of the classes most likely to recover, i.e. Class 3 in the three-class model and Class 2 in the four-class model.

Variables used for identifying classes	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**All**
SF127 1 *	0	0	1	8	9
SF127 2	4	9	4	28	46
SF127 3	21	21	74	15	131
SF127 4	56	7	11	1	75
SF127 5	128	5	11	2	146
Not Anxious	155	21	0	0	176
Borderline	44	20	58	1	123
Clinically Anxious	11	1	43	53	108
Not Depressed	204	14	3	0	257
Borderline	6	27	60	21	114
Clinically depressed	0	1	2	33	36
Not troublesome	3	0	0	0	3
Moderately	120	12	49	13	194
Very	79	21	42	17	159
Extremely	8	9	10	24	51
Fear Avoidant	76	19	71	45	211
Not Fear Avoidant	134	23	30	9	196
Pain Self efficacy Very Low	0	6	0	20	26
Low	0	33	0	29	62
Moderate and above	210	3	101	5	319
Outcome variables	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**All**
RMQ Significant improvement	90	20	51	19	180
No Significant Improvement	120	22	50	35	227
RMQ † mean baseline	6.62	10.87	9.08	14.08	8.47
RMQ † mean 12 months	4.31	8.42	6.68	12.27	6.30
Variables (adjusted model)	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**All**
Age under 40	35	8	25	7	75
40-49	42	9	23	13	87
50-59	48	7	22	19	96
60-69	56	10	24	13	103
70 and over	29	8	7	2	46
Female	113	22	75	36	246
Male	97	20	26	18	161
Working	125	14	54	15	208
Not working	84	28	47	38	197
missing	1	0	0	1	2

Table 6.9: Patient Characteristics in the four-Class latent class model
* SF12 question 7 'How much of the time does your pain interfere with social activities?' 1= all of the time, 2=most of the time, 3=some of the time, 4=a little of the time, 5= none of the time.
† Roland Morris Disability Questionnaire. RMQ score takes integer values 0 to 24, with lower score indicating less disability.

Variables used for identifying classes	**Class 1**	**Class 2**	**Class 3**	**All**
SF127 1 *	0	8	1	9
SF127 2	2	29	15	46
SF127 3	19	11	101	131
SF127 4	60	2	13	75
SF127 5	130	2	14	146
Not Anxious	155	1	20	176
Borderline	43	1	79	123
Clinically Anxious	13	50	45	108
Not Depressed	206	0	51	257
Borderline	5	16	93	114
Clinically depressed	0	36	0	36
Not troublesome	3	0	0	3
Moderately	122	11	61	194
Very	79	16	64	159
Extremely	7	25	19	51
Fear Avoidant	76	44	91	211
Not Fear Avoidant	135	8	53	196
Pain Self efficacy Very Low	2	19	5	26
Low	2	26	34	62
Moderate and above	207	7	105	319
Outcome variables	**Class 1**	**Class 2**	**Class 3**	**All**
RMQ Significant improvement	89	18	73	180
No Significant Improvement	122	34	71	227
RMQ † mean baseline	6.35	13.71	9.69	8.47
RMQ † mean 12 months	4.21	12.02	7.31	6.30
Variables (adjusted model)	**Class 1**	**Class 2**	**Class 3**	**All**
Age under 40	36	8	31	75
40-49	43	13	31	87
50-59	48	17	31	96
60-69	55	12	36	103
70 and over	29	2	15	46
Female	113	36	97	246
Male	98	16	47	161
Working	127	16	65	208
Not working	83	35	79	197
missing	1	1	0	2

Table 6.10: Patient Characteristics in the three-Class latent class model
* SF12 question 7 'How much of the time does your pain interfere with social activities?' 1= all of the time, 2=most of the time, 3=some of the time, 4=a little of the time, 5= none of the time.
† Roland Morris Disability Questionnaire. RMQ score takes integer values 0 to 24, with lower score indicating less disability.

Variable	Class & Treatment model			Age Sex and Work		
	Coeff	Std error	p-value	Coeff	Std error	p-value
Best care only	-0.71	0.24	0.004**	-0.28	0.73	0.70
Cognitive behavioural approach	-0.11	0.49	0.17	-0.42	0.44	0.34
Class 1	-	-	-	0	-	-
Class 2	-0.79	0.82	0.34	-1.25	0.91	0.17
Class 3	-0.12	0.41	0.77	-0.36	0.46	0.44
Treatment * Class 1	-	-	-	0	-	-
Treatment * Class 2	0.46	0.89	0.61	0.99	0.99	0.31
Treatment * Class 3	0.62	0.49	0.21	0.94	0.54	0.08 •
Age under 40	-	-	-	0	-	-
Age 40-49	-	-	-	0.67	0.68	0.32
Age 50-59	-	-	-	0.45	0.66	0.49
Age 60-69	-	-	-	0.07	0.71	0.91
Age 70 and over	-	-	-	-0.48	0.97	0.62
Male	-	-	-	-0.65	0.42	0.12
Female	-	-	-	0	-	-
In Work	-	-	-	-0.64	0.55	0.24
Not in Work	-	-	-	0	-	-
CBA* under Age 40	-	-	-	0	-	-
CBA* Age 40-49	-	-	-	-0.82	0.78	0.29
CBA* Age 50-59	-	-	-	-0.61	0.77	0.42
CBA* Age 60-69	-	-	-	-0.22	0.83	0.79
CBA* Age 70 and over	-	-	-	0.32	1.10	0.77
CBA*Female	-	-	-	0	-	-
CBA* Male	-	-	-	0.38	0.49	0.43
CBA*Not in work	-	-	-	0	-	-
CBA*in work	-	-	-	1.53	0.63	0.02 ⋆

Table 6.11: Logistic Regression Model Coefficients
Significance codes: • significant at $\alpha = 0.1$, ⋆ significant at $\alpha = 0.05$, **significant at $\alpha = 0.01$

These tables reveal that, in the four-class model, Class 2 is very small, 42 patients out of 407 which is just 10.3%. It is likely, therefore, that the marginal association between Class 2 and treament for this model is an artefact of the small number of observations, and may not be a genuine efffect. In the three-class model, 144 patients are in Class 3, 35% of the observations, which is what would be expected if patients were allocated into three groups at random. It is much less likely, therefore, that the marginal association between membership of Class 3 and treatment is an artefact of the distribution of patients between classes.

Scores which measure the characteristics that the intervention was designed to tackle were successfully used to classify trial participants into three and four

Class	Work		No Work		missing		Total
Recovery	Yes	No	Yes	No	Yes	No	
Class 1	59	66	30	54	0	1	210
Class 2	9	5	11	17	0	0	42
Class 3	30	24	22	25	0	0	101
Class 4	8	7	10	28	1	0	54

Table 6.12: Work and recovery in classes for the 4-class model.

Class	Work		No Work		missing		Total
Recovery	Yes	No	Yes	No	Yes	No	
Class 1	67	60	29	54	0	1	211
Class 2	8	8	9	26	1	0	52
Class 3	38	27	35	44	0	0	144

Table 6.13: Work and recovery in classes for the 3-class model.

groups. For trial participants who received the intervention treatment, there was an association between outcome and class membership in the three-class model. After adjusting for age, sex and work status in a logistic regression model, there was a marginally significant interaction between treatment and membership of Class 3 in the three-class model, suggesting that Class 3 patients respond to the treatment differently than Class 2 and Class 1 patients, but with the caveat that the sample size is small. Underwood *et al* concluded that although BeST is one of the larger trials of back pain treatment, it is still too small reliably to detect moderation if it exists (Underwood et al. [2011]).

6.3 Artificial Neural Network analysis of the BeST data

This is a comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioural approach to treatment for non-specific low back pain Barons et al. [2013b].

6.3.1 Principal findings

Using the same 7 variables as for the latent class analysis, a classification ANN and an ordinary logistic regression were more accurate in classifying patients with respect to recovery than a logistic regression using Latent Class membership as the predictor. The best log score performances were by the ANN and the latent class logistic regression, whilst the best sensitivity was shown by the ANN, and for specificity, latent class logistic regression was highest. The best overall performance was the ANN, with the best performance in 3 of the 4 measures and providing both sensitivity and accuracy. The ANN is the best candidate of these three models for decision support for allocating patients to the cognitive behavioural approach to the treatment of back pain.

6.3.2 Introduction

The aim of this study was to compare three methods for using patient characteristics as predictors of a 3-point or greater improvement in the RMQ score between baseline and 12-month follow-up. A classification ANN was compared to two logistic regression models, one using the variables directly, and the other using the classes derived from the latent class analysis as predictors. All three models were subjected to 10-fold cross validation and the averaged accuracy performance was reported.

In each of the three approaches, the same seven variables were used as in the construction of the latent class models above. These were selected on the basis that they measure the attributes which the cognitive behavioural approach was designed to influence. The outcome measure was the RMQ score change by 12-month follow-up. The predictor variables were PSE, SF12_7, HADS (anxiety and depression scales separately) and troublesomeness. The 407 cases with complete data, detailed in Table 6.10, were used for the analyses.

As described in Chapter 5, artificial neural networks are a form of machine learning (Mitchell [1997]). A feed-forward artificial neural network with a single hidden layer was used, and the number of hidden nodes was varied from 1 to 5 to

optimise using the percentage error and the log score comparison metrics described below. The logistic function was used for both activation and output functions. The inputs were scaled $\in [0, 1]$ by dividing each score by the maximum score, and weight decays 0, 0.0001, 0.001, 0.01, and 0.1 were tested. 10-fold cross validation was carried out and the percentage error and the log score calculated to determine the optimal network and for comparison with the other two models.

For comparison with the ANN and the latent class analysis, which are capable of capturing non-linearity, a linear logistic regression was used to predict outcome using all the six predictors listed. This was also used to predict outcome in a 10-fold cross validation, and the log score was calculated for these models.

6.3.3 Comparison metrics

All analyses were carried out in R statistical software (R Development Core Team [2011]). The artificial neural network was built using the the R package *nnet* (Venables and Ripley [2002]) which had a logistic activation function and with a logistic output activation selected. The optimisation of weights was carried out by selecting the entropy option which is maximum conditional likelihood fitting, or equivalently for the case with a binary output, minimising the Kullback-Leibler distance (Venables and Ripley [2002]).

$$E = \sum_p \sum_k \left[t_k^p \ \log \ \frac{t_k^p}{y_k^p} \ + \ (1 - t_k^p) \ \log \ \frac{(1 - t_k^p)}{(1 - y_k^p)} \right] \tag{6.1}$$

where t^p is the target (recovery $\in [0, 1]$), and y^p is the model's output for the pth patient and k is the number of classes. The percentage error - the total percentage of patients assigned to the wrong outcome, is reported, making no distinction between sensitivity errors and specificity errors (false positives or false negatives) and averaged over the 10-fold cross validation. In this application, it is desirable to be able to identify all those for whom the treatment will be effective, even at the risk of allocating some to this programme who were unlikely to benefit, since there are no known or suspected contraindications. The log score is calculated:

$$\sum_p \sum_k -t_k^p \ \log \ p_k^p \tag{6.2}$$

Clearly, whenever the outcome is no recovery ($t_k^p = 0$) then there is zero contribution to this sum, so the log score reported is for the positive outcomes only and is calculated using Equation 6.2. A lower log score indicates greater ability to identify

those who will recover; it will predict a high probability of recovery for a patient who actually recovered. The behaviour of log score is illustrated by plotting probability of recovery from 0.1 to 1 against log score.

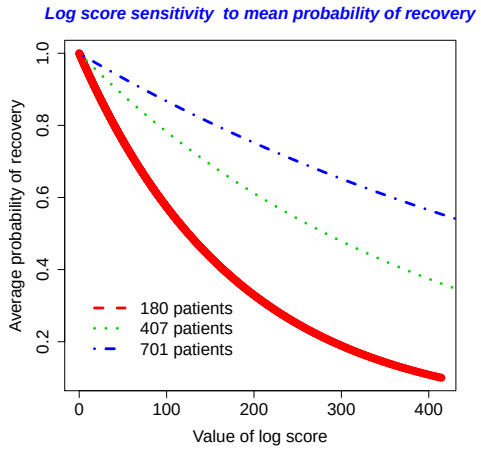


Figure 6.9: As average probability of recovery varies, there is significant change in the value of the log score, for example as average probability of recovery rises from 0.5 to 0.6, the log score reduces from 124.8 to 91.9.

This shows that even a small change in the mean probability of recovery predicted by any of the models, for the recovered patients, influences the log probability score. It follows that the log score is also sensitive to outliers; a single value much lower or higher than the rest will also adjust the log score. This demonstrates that the log score is a useful measure in this context.
The two logistic regression models were subjected to 10-fold cross validation using the R package *DAAG* (Maindonald and Braun [2012]), and the log score was calculated for each.

Artificial neural network

Log score was minimised when there were two hidden nodes for all networks, except the network with weight decay set at 0.1, where the one hidden node configuration had the lowest log score. When weight decay = 0.1, the network with 4 hidden nodes also provides a lower log score than either 3 or 5 hidden nodes. Percentage error (misclassification during cross validation) is minimised in the networks with one hidden node for all values of the weight decay. The distribution of probabilities of recovery assigned by the ANN model (with one hidden node and weight decay 0.1) to the patients who did recover is shown in the histogram (Figure 6.11). A perfect plot would have the entire mass at 1. This plot shows that the bulk of the predicted

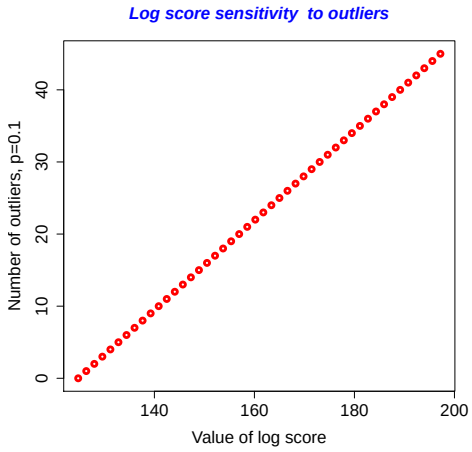


Figure 6.10: This plot shows how the log score varies as the number of patients with probability of recovery much smaller than the remainder rises. For example, if 1 patient had a probability of 0.1 and average probability of recovery for the remaining 179 is 0.5, and then the log score rises from 124.8 to 126.4. If two patients have probabilities of 0.1, the log score is 127.9, if three then 129.6.

probabilities were 0.5 and above, which indicates why the log score is reasonably good at 145.5 and the prediction error is 38.8% (sensitivity 0.58, specificity 0.63).

Hidden nodes		1	2	3	4	5
Weight decay						
0	% error	40.5	41.3	42.3	45.5	43.2
	Log Score	150.0	145.8	154.3	154.6	166.7
0.0001	% error	39.8	42.8	43.9	44.9	45.5
	Log Score	149.9	147.9	155.7	158.4	171.6
0.001	% error	39.8	42.5	42.8	45.2	44.9
	Log Score	150.9	149.6	154.7	157.8	164.2
0.01	% error	40.8	41.5	42.5	43.7	44.9
	Log Score	149.3	144.5	152.4	154.3	161.4
0.1	% error	38.8	41.0	42.0	43.5	43.5
	Log Score	145.5	146.8	149.1	148.5	152.1

Table 6.14: Evaluation of number hidden nodes required for optimal neural network.

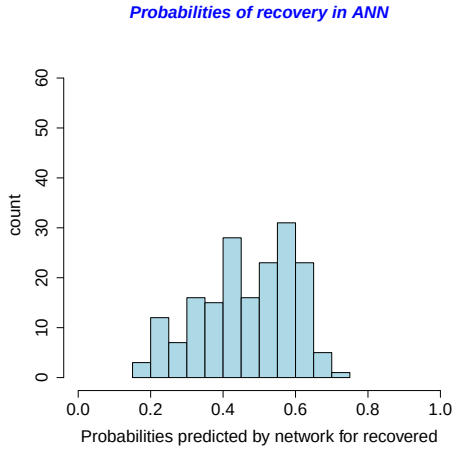


Figure 6.11: Histogram of the probabilities of recovery for cases that did recover, as predicted by the artifical neural network with one hidden node and weight decay 0.1.

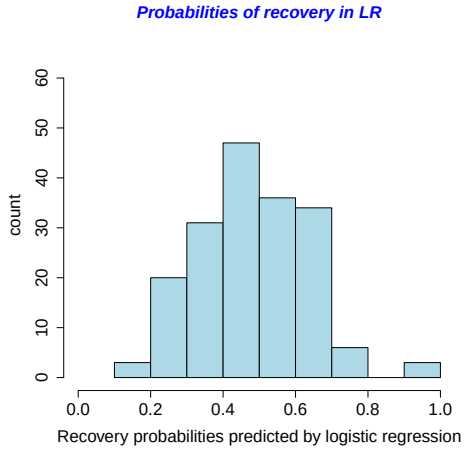


Figure 6.12: Histogram of the probabilities of recovery for cases that did recover, as predicted by the standard logistic regression.

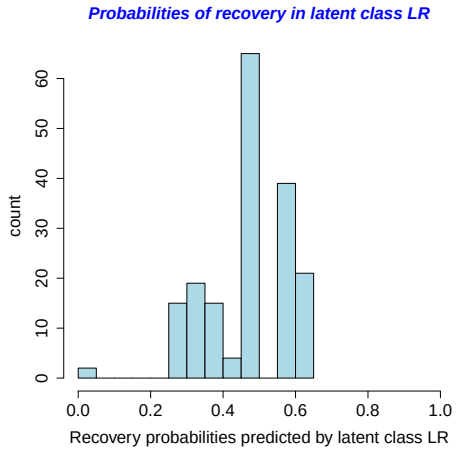


Figure 6.13: Histogram of the probabilities of recovery for cases that did recover, as predicted by the latent class logistic regression.

Here is for ROC curve for the fitted ANN. The diagonal line is the line of no discrimination, and since the ANN line is above this, the model is shown to have some discrimination capability. As with the other measures, the ROC curve also shows that the model has room for improvement. Since there are no known side-effects of attending CBA, it is more important in this application to capture those patients likely to benefit, at the cost of treating some who are unlikely to recover. Without a method for making such a prediction, all patients are likely to be treated and costs will be higher still.

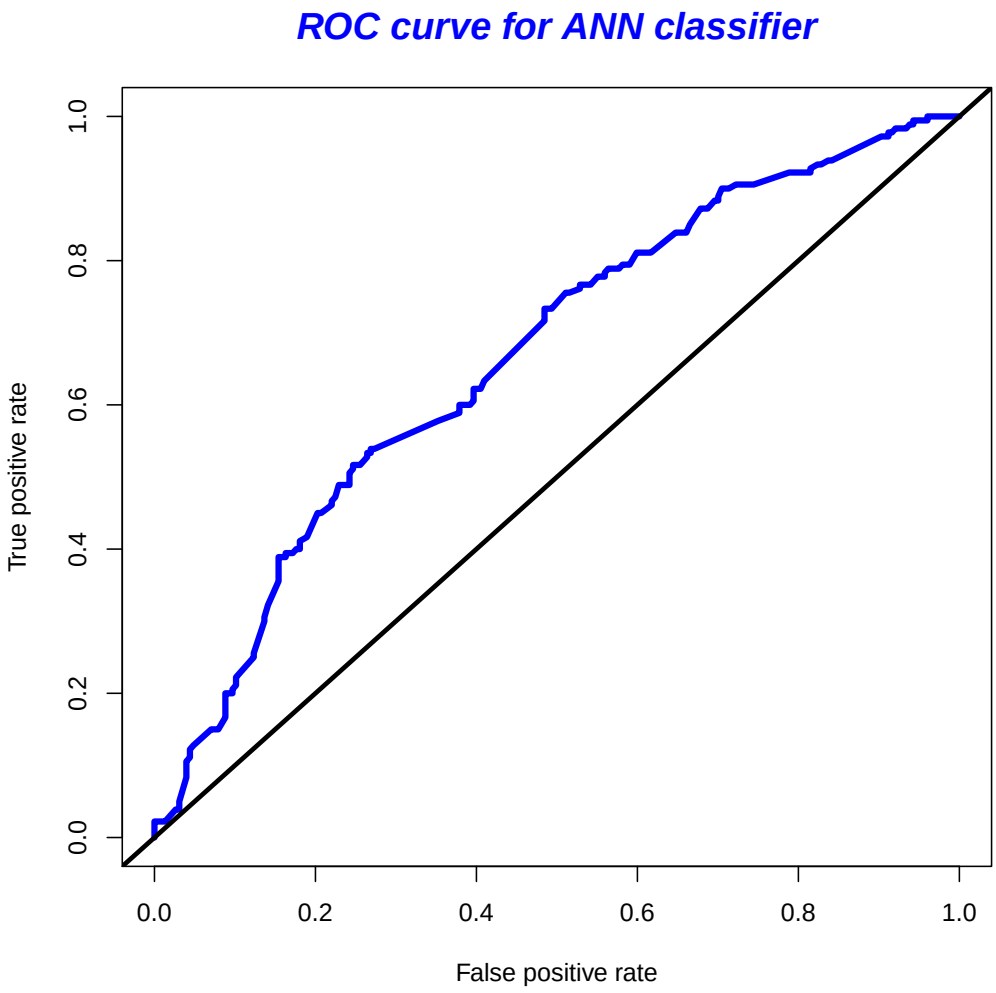


Figure 6.14: ROC curve for illustrating the balance of errors as the threshold at which the boundary between classifications used by the model is varied.

Model	Prediction Error %	Log Score	Sensitivity	Specificity
LCA+ LR (3-class)	42	146.5	0.33	0.82
LR	38.8	170.5	0.46	0.72
ANN (1 hidden unit w-decay 0.1)	38.8	145.5	0.58	0.63

Table 6.15: Comparison of all analysis methods

Logistic Regression

The first logistic regression we consider is the standard logistic regression model using the same explanatory variables as the ANN. 10-fold cross validation gave a prediction error of 38.8%, and a log score of 170.5 (sensitivity 0.46, specificity 0.72). Next the latent class model with three classes was used. The latent class model was built using 6 variables, then the class membership and the treatment were used as predictors in a regression model. The 10-fold cross validation prediction error was 42.0% and the log score was 146.5 (sensitivity 0.33, specificity 0.82).

The distribution of probabilities of recovery assigned by the logistic regression models to the patients who did recover is shown in the histograms (Figures 6.12 and 6.13). The distribution of probabilities assigned to the patients who did recover by the ordinary logistic regression have a few very high values and a higher proportion of the probabilities below 0.5 than in the case of the ANN. This ordinary logistic regression had the highest log score. The latent class logistic regression model had probability distribution for the patients who did recover which is skewed towards the larger values, with the bulk of the probabilities falling in the 0.5 and 0.6 bins. It had fewer very low probabilities than the ordinary logistic regression and also no high probabilities. The log score for this model is only slightly higher than the log score for the best ANN model, indicating that these two models outperform ordinary logistic regression.

The overall accuracy is highest for the ANN and the ordinary logistic regression, with the latent class logistic regression misclassifying 13 more patients than the other two models. The logistic regression is the only linear model, and had the lowest error rate but the highest log score. The other two models allow for non-linearity. The latent class logistic regression model uses 6 of the variables to assign patients to classes, followed by using the class membership and the treatment as predictors of outcome in a logistic regression. This is superior to the ordinary logistic regression in sensitivity but poorer in overall accuracy. The second non-linear model is the ANN using the same 7 variables as inputs. This performs as well as the logistic regression with respect to accuracy, but is superior in the log score. The ANN performs as well

as the latent class logistic regression in terms of log score, but is superior in overall accuracy.

We hypothesise that by using techniques that can account for non-linearities and seeing improvement suggests that linear models do not describe the connection between the cognitive domain and the experience of back pain well. We have shown that the artificial neural network provides the best combination of overall error rate and sensitivity, and would be the best candidate of these three models for decision support for the cognitive behavioural approach to treatment of lower back pain.

6.4 Analysis of the Basingstoke and Alton cardiac rehabilitation cohort data

These data were analysed by two methods. The Cox proportional hazards model, which is a standard survival analysis method, was used and optimised to find the best subset of explanatory variables for all-cause and cardiovascular survival. This subset of explanatory variables was used to fit an ANN and used to predict mortality.

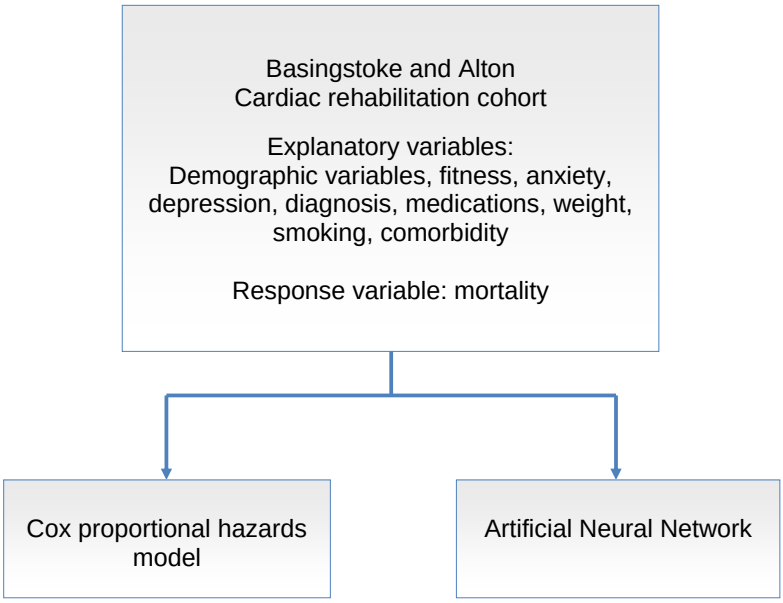


Figure 6.15: Schematic representing the analysis scheme for the BeST data set.

6.5 The Cox Model for Long-Term Survival After a Cardiovascular Event

The long-term mortality of an unselected cohort of patients who have experienced a coronary event or procedure was investigated, to identify factors which were associated with longer term survival.

6.5.1 Principal findings

Predictors of long-term survival (> 5 years) in patients who have had a coronary event or procedure were not identical to the predictors of short-term survival. After adjusting for age, sex and medications, fitness remains a strong predictor of long-term survival from both all-cause and cardiovascular mortality. When changes in fitness during the rehabilitation period were taken into account, an improvement to high fitness is associated with the same benefits as having started with high fitness.

6.5.2 Introduction

Recall that access was granted to data on a cohort of patients recruited between 1st January 1993 and 31st December 2002 through the Basingstoke and Alton (Hampshire, UK) cardiac rehabilitation programme with follow-up to 30th March 2011. The cohort is unselected, includes all NHS referrals, and participants have now been followed for between one day and 18 years and three months providing 11,871 person-years of follow-up. Recruitment to the cohort was undertaken typically 2 to 6 weeks after their index coronary event (Bethell et al. [2009], West et al. [2011]). NHS patients in the area served by the rehabilitation centre were routinely referred to this programme following an acute myocardial infarction, episode of unstable angina or revascularisation. The only other inclusion criterion for the study was that the patients had registered with the cardiac rehabilitation programme. This resulted in an unselected cohort of 2,714 patients. Data collection was undertaken by one person (ST) for each patient, at recruitment and on graduation from the programme, ensuring consistency over time (Turner [2007]). The Office for National Statistics provided data on dates and causes of death. All patients who attended the programme and had baseline fitness measured were included in the analysis. Data collected included whether the programme was completed, diagnosis, co-morbidity, family history, occupation, date of birth, age, sex, smoking history, resting heart rate, cholesterol level, triglycerides level, post code, and, from 1998, height. At both recruitment and on graduation records were made of each patient's weight,

blood pressure, fitness, anxiety and depression (as measured by Hospital Anxiety and Depression Scale (Bjelland et al. [2002],Turner [2007])), current smoking habit and medications (ACE inhibitor, aspirin, beta blocker, statins). There were developments in protocols for secondary prevention medications during the period of the study (Turner [2007]). At the start of the study, around 80% of those attending the Basingstoke and Alton rehabilitation programme were prescribed aspirin, 20% prescribed ACE inhibitor and beta blocker, and almost none prescribed statins. Statin prescriptions climbed dramatically to almost 60% in the two years to 1996, to 85% in the following 4 years, and then more slowly to a steady 95% from 2002 onwards. Beta blocker prescription rose from 20% to 30% between 1993 and 1998 and then to 60% in the 2 years to 2000, remaining between 50% and 60% for the duration of the study. Similarly, ACE inhibitor prescription rose slowly between 1993 and 1998 and then experienced a strong increase, so that by 2002, over 80% of patients were prescribed ACE inhibitors. Aspirin prescription rose steadily from over 80% at the start of the study to almost 100% by 2001. The patients recruited during the last 4 years of enrolment almost all had prescriptions for aspirin and statins, 80% had a prescription for ACE inhibitor and between 40% and 50% had prescriptions for all 4 secondary prevention medications. A plot showing the changes in use of secondary prevention can be found on page 184 of Turner [2007]. It might be expected that these patients had a longer life expectancy post-event than those recruited under the earlier regime.

6.5.3 Calculated variables

Changes in weight, fitness, anxiety and depression were calculated from the entry and exit values for each individual. Since height was not routinely recorded throughout, BMI was available for only 889 patients. Baseline weight was categorised into under 75kg, 75-90kg, and over 90kg and labelled A, B and C respectively for brevity. The index of multiple deprivation was ascertained from post code, occupation was coded under 9 headings (see Table 6.22), and age was categorised into under 50, 50-59, 60-69 and 70 and over. The Modified D'Hoore Co-morbidity Index is designed to assess non-coronary co-morbidity specifically in the out-patient cardiac rehabilitation environment, rather than the acute setting (D'Hoore et al. [1996]), and was calculated for each patient based on the recorded co-morbidities (see table 3.2 on page 58).

Up to August 1995, fitness was measured on a bicycle ergometer with ECG monitoring and measurement of estimated peak workload. After that date, exercise tests were performed on a treadmill, using either the Bruce protocol (Bruce et al.

[1973]) or, for frail or elderly patients, the modified Bruce protocol (Bruce [1973]). Peak exercise tolerance was expressed as the predicted oxygen uptake (VO_2 max) in ml/kg/minute, from the known oxygen cost of bicycling at different workloads (Astrand and Rodahl [1986]). The treadmill test used the same endpoints as for bicycle tests and VO_2 max predicted on the assumption that each one minute of the Bruce protocol uses one MET (metabolic equivalent - or 3.5ml O2/kg/min) and that the first three stages of the modified Bruce protocol each use one MET. VO_2 max<15ml/kg/min was categorised as low fitness, VO_2 max>22ml/kg/min as high fitness with the remainder as medium (established for cardiac rehabilitation by Kavanagh et al. [2002] Kavanagh et al. [2003]).
The researchers who collected the data report that the exercise test protocol for the treadmill tests was either the full Bruce or the modified Bruce protocol, with the latter used for frailer patients as it starts more gently. With the treadmill set at a constant 1.7 mph, the gradient is increased from 10% (full Bruce) or 0% (modified Bruce) every 3 minutes to a maximum gradient of 22%. The metabolic equivalents (METs) are calculated from the number of minutes until the patient had reached 85% of their predicted heart rate maximum for their age or developed symptoms which precluded the test's continuation using the standard conversion that 1 minute of the Bruce protocol used one MET, and the first two stages of the modified Bruce protocol use 1 MET each. (Turner [2007]). Each MET is equivalent to 3.5ml/kg/min so VO_2 max Max is estimated as 3.5(1+x), with x being the number of minutes and the additional 3.5ml/kg/min representing the resting metabolic rate.
Treadmill testing using the Bruce protocol is used for both diagnosis and prognosis in patients at risk of coronary disease. Symptoms of Ischemia prompt a diagnosis of disease. Prognosis is estimated using exercise duration, exercise hypotension, exercise hypertension chronotropic incompetence, heart rate recovery and ventricular ectopy. Exercise duration is a good measure of functional capacity and a longer duration indicates a lower probability of mortality from coronary disease or any other cause, including in healthy subjects and retains its prognostic value after adjusting for age and sex (Miller [2008]). Lower exercise duration can be an indicator of lower fitness or more severe cardiovascular disease.

Depression and anxiety were categorised into none, borderline and depressed or anxious using a Hospital Anxiety and Depression Scale score below 8 to suggest no depression (anxiety), 8-10 to suggest borderline and a score over 10 to suggest clinical depression or clinical anxiety (Bjelland et al. [2002]).

In the main analysis, baseline categories of fitness, depression and anxiety were used as predictors. Since the influence of an improvement in fitness, depression or anxiety categories on survival is of interest, and whether these depend on baseline categories, variables Fitness, Anxiety and Depression were defined so that a scale of 1 to 7 captured each starting category and improvement or deterioration (see Table 6.23). Very few patients started in the highest fitness category and then deteriorated over the course of the programme (5 men) so these were combined with the 'no change' category. There were no patients who deteriorated having started in the mid-fitness category, so this category was omitted. Survival time was defined as the period between the date the participant joined the programme and the date of death.

Statistical analysis

In the survival analyses both all-cause mortality and cardiovascular mortality were considered, with non-cardiovascular deaths treated as censored in the latter analysis. Cox proportional hazards models (Collett [2003], Cox [1972]) were used to model both all-cause and cardiovascular survival, beginning from the subset of variables that were found to be significant predictors (at 5% level) of all-cause mortality after preliminary univariate analyses as the baseline model. A backward stepwise selection algorithm was employed to find the model with minimum AIC (Akaike [1974]), retaining age and sex as the minimum model. The high calibre practice at the rehabilitation programme gives us confidence that there was no loss to follow-up. All analyses were carried out in R statistical software (R Development Core Team [2011]).

Missing data

The Cox proportional hazards model was used on the complete cases. There were 1,529 cases (56.3%) of the 2,714 available which were complete in all 36 variables which were significant in the univariate analysis and became the starting point for the definitive analysis. There were 1,029 cases (38%) which were complete in these variables and also had observations of fitness, depression and anxiety at the end of the programme. To address the issue of missing data in order to assess the credibility of a model built using only the complete cases (Little and Rubin [2002]), hazard ratios from the complete-case analysis were compared with those from an analysis of all data (2,714 cases) after replacement of missing values with imputed data. Multiple imputation was performed in the R package *MICE* (Multivariate Imputation using

Chained Equations, R Development Core Team [2011]), providing 20 completed data sets with values imputed where they were missing. The optimised model was then built with each of these 20 data sets, and pooled estimates of model coefficients and variances calculated using the rules devised by Little and Rubin (Little and Rubin [2002]) for comparison. There were no known reasons to believe that the data were missing other than at random, except in the case of the end fitness data. Reasons for end fitness to be missing included referral for cardiac surgery and poor health, suggesting that time to death for these individuals might differ from those with end fitness observations. A t-test showed that the mean time to death was not the same for those with and without an end fitness measurement ($P<0.001$)
A total of 2,054 patients completed the programme (Turner [2007]); the main reasons for not completing were patient preference, referral for cardiac surgery, poor health or death. Those without complete end fitness data included those who were taken to a different heart rate at the final fitness test, most of whom were on a beta blocker for which the dose had been changed or who were tested at a different time of day, and those who were tested using a different test protocol.

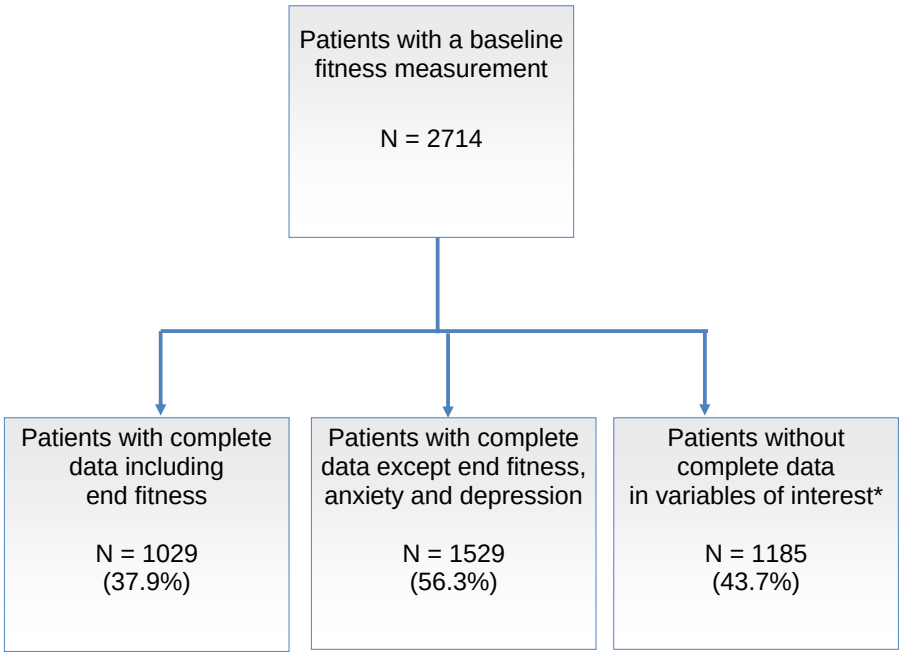


Figure 6.16: Patient recruitment and eligibility. *The number of patients with missing data in individual variables of interest varied, with fitness after the programme missing in 48.5% of cases, and the remaining variables ranging between 0 and 29.4 % . Details of variables and numbers and percentage of missing observations are in Table 6.18 on page 146.

The mean follow-up for those with complete data was 11 years 4 months.
Detailed characteristics of patients are shown in Table 3.3. Women (13.7%) had a higher mean age. The largest age group, for both sexes, was 60-69 years. The most common reason for referral to the programme was acute myocardial infarction (51.1%). Nearly half the women had never smoked compared to just under one third of the men and close to a third of both sexes had recently given up smoking. Around half of all patients had a family history of coronary heart disease and over 70% were free from non-coronary co-morbidity, although a higher proportion of the women had diabetes. The men had a higher mean fitness than the women and the percentage of men in the high-fitness category at recruitment was 3 times that of the women; most of the women were in the low-fitness category. At graduation, 67% of the men and 34% of the women were in the high-fitness category, with one fifth of the patients having improved from the mid-fitness category. Median fitness was improved by a similar amount in both sexes, and high fitness was the largest group at graduation.

There was little evidence of clinical depression in this cohort, and most of those whose scores suggested a borderline category improved by graduation, as did almost all of the few whose scores at recruitment suggested clinical depression. There was more anxiety, although the rates were not high. Again, the majority who began in the borderline category improved, as did a significant proportion of those starting in the clinical anxiety category, but the proportion of women who remained in the clinical anxiety class was 3 times that of the men.

The maximum recorded co-morbidity score was 7 (very high) but only 2.6% of participants had a co-morbidity score of 3 or more (D'Hoore et al. [1996], see Table 3.2 on 58).

Survival

During the course of the study and follow-up of the 1529 participants, 385 died, and of those deaths 192 (49.9%) were from cardiovascular causes.

Two Cox proportional hazards models were constructed from these data, the primary model used baseline only measures of fitness, depression and anxiety, for which there were 1,529 complete cases available. The secondary model used baseline and change fitness, depression and anxiety as described in Table 6.23 for which 1,029 cases were available. Table 6.16 details all the significant predictors in the primary proportional hazards model of all-cause mortality. In this primary model, age was the most

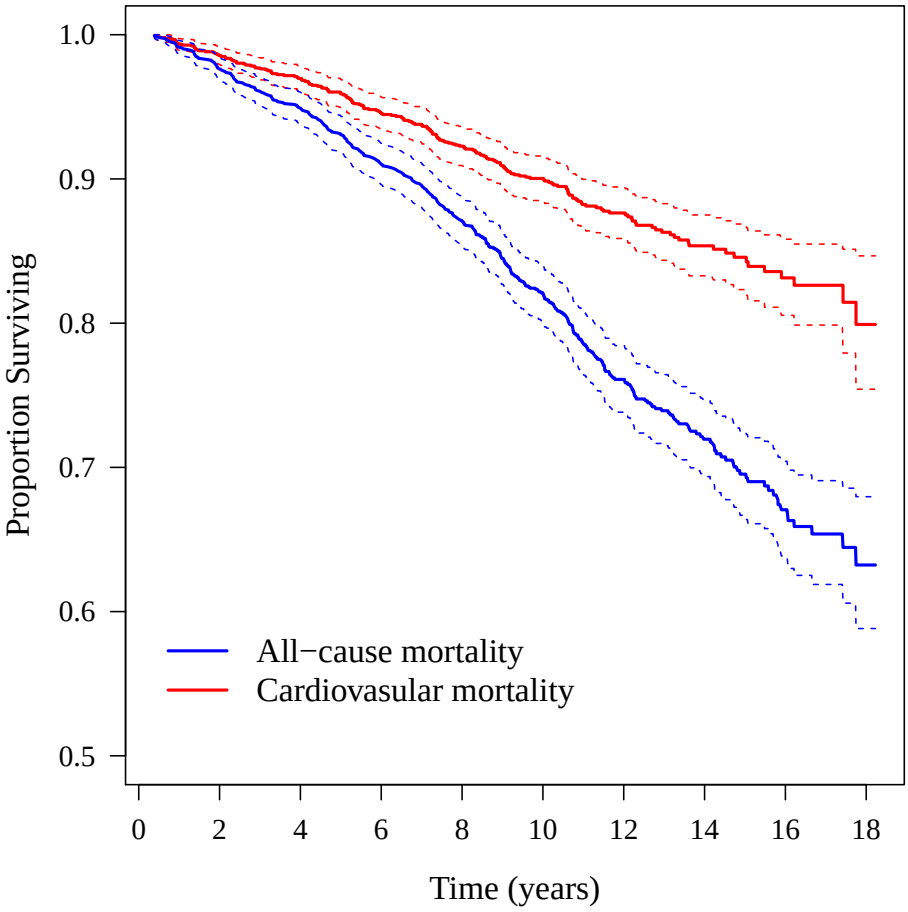


Figure 6.17: Kaplan-Meier survival curves for all cause and cardiovascular mortality for the primary model, the cohort with complete baseline data, 1529 cases. The plot is for the entire observation time and the dotted lines are the 95% confidence intervals.

important predictor of all-cause mortality, with risk increasing with age. In this cohort, the next most important predictor was fitness category at recruitment, with the fittest patients having lowest risk. Risk increased with co-morbidity, and both aspirin and statins prescriptions reduced risk. Myocardial infarction was the diagnosis carrying the greatest risk, along with myocardial infarction with percutaneous coronary intervention, angina and other cardiac diagnoses. Coronary artery bypass graft and angina were much lower risk diagnoses. Female sex carried a lower risk of mortality, as did a lower resting heart rate. Systolic blood pressure was also a risk indicator. ACE inhibitor prescription was associated with higher risk, perhaps

because this prescription is given to the high risk cases.

All-cause survival model					
		complete cases model			Imputed data model
Model Term		Hazard Ratio	Confidence Interval		Pooled Hazard Ratio
			lower .95	upper .95	
Age category	under 50	1	-	-	1
	50-59	2.13	1.17	3.88	2.02
	60-69	3.84	2.15	6.88	3.08
	70+	7.94	4.38	14.39	6.07
Fitness:					
	High baseline	1	-	-	1
	Mid baseline	1.56	1.16	2.09	1.76
	Low baseline	2.60	1.86	3.60	2.76
D'Hoore Co-morbidity score	None	1	-	-	1
	1 (least)	1.23	0.91	1.66	1.17
	2	1.42	1.08	1.86	1.48
	3	1.35	0.68	2.66	1.87
	4 (most)	4.50	2.13	9.50	2.67
Statins	Yes	0.72	0.57	0.89	0.74
	No	1	-	-	1
Aspirin	Yes	0.53	0.35	0.79	0.63
	No	1	-	-	
Diagnostic Category	MI	1	-	-	1
	CABG	0.69	0.54	0.89	0.69
	PCI	0.54	0.32	0.89	0.68
	MI + PCI	0.85	0.44	1.67	0.82
	Angina	0.87	0.54	1.39	0.79
	Other cardiac	0.99	0.44	2.23	0.94
Sex	Male	1	-	-	1
	Female	0.73	0.54	0.98	0.62
Systolic blood pressure before		0.995	0.991	0.999	0.998
Ace inhibitor	Yes	1.26	1.01	1.57	1.19
	No	1	-	-	1
Resting heart rate		1.007	1.000	1.014	1.006

Table 6.16: All-cause survival model, ordered by importance of variables to the model, using baseline only fitness, anxiety and depression categories (1,529 cases, 385 deaths). Pooled hazard ratios are from multiple imputation of missing data. MI is Myocardial Infarction, CABG is coronary artery bypass graft, PCI is percutaneous coronary intervention.

Cardiovascular survival model: baseline fitness					
		complete cases model			Imputed data model
Model Term		Hazard Ratio	Confidence Interval		Pooled Hazard Ratio
			lower .95	upper .95	
Fitness:	High baseline	1	-	-	1
	Mid baseline	1.69	1.11	2.60	2.16
	Low baseline	4.00	2.54	6.30	4.12
Statin	Yes	0.45	0.33	0.61	
	No	1	-	-	
Age category	under 50	1	-	-	1
	50-59	1.59	0.76	3.32	1.53
	60-69	2.68	1.32	5.43	2.42
	70+	4.10	1.98	8.48	3.83
Diagnostic Category	Myocardial Infarction (MI)	1	-	-	1
	Coronary Artery Bypass Graft (CABG)	0.60	0.42	0.85	0.62
	Percutaneous Coronary Intervention (PCI)	0.26	0.09	0.70	0.49
	MI + PCI	1.05	0.42	2.62	0.73
	Angina	0.85	0.46	1.56	0.71
	Other cardiac	0.72	0.23	2.29	1.01
Aspirin	Yes	0.48	0.28	0.80	0.56
	No	1	-	-	1
Sex:	MALE	1	-	-	1
	FEMALE	0.65	0.43	0.99	0.53
Ace inhibitor	Yes	1.42	1.04	1.94	1.37
	No	1	-	-	1

Table 6.17: Optimised Cardiovascular survival model, ordered by importance of variables to the model, using only baseline values for fitness, depression and anxiety (1,529 cases 192 cardiovascular deaths). Pooled hazard ratios are from multiple imputation of missing data.

The cardiovascular mortality model is detailed in Table 6.17. Fitness category at baseline was the strongest predictor of cardiovascular mortality, with higher fitness associated with lower risk. A prescription for statins cut the risk of cardiovascular mortality by more than half in this cohort. Age was the next most significant predictor of mortality, with risk increasing with age as expected. As with all-cause mortality, a diagnosis of myocardial infarction carried the highest associated risk of mortality, with MI+PCI, angina and other cardiac diagnoses equally high. Female sex was associated with lower risk, and a prescription for ACE inhibitor with higher risk.

In the cohort having complete data including end depression, anxiety and fitness (the secondary model), age was still the strongest predictor of risk for all-cause mortality (Table 6.24). The next most important predictor was the combination of fitness category at recruitment, and whether they improved or maintained that fitness, with highest risk attributed to those who began in the low-fitness category. There was no statistically significant difference (assessed at the 5% level) between those who began in the mid-fitness category and improved to high fitness and those who began in the high-fitness category and maintained high fitness. However, those who did not improve sufficiently to move up from the mid-fitness category had significantly higher risk; improvement to a mid-fitness from low-fitness category did not significantly reduce risk, although a significant difference in risk is evident between low and medium fitness for the patients whose category did not change. Having a prescription for statins or having a prescription for aspirin was each associated with a lower risk of mortality. A prescription for ACE inhibitors was associated with a higher risk of mortality and females had lower all-cause mortality.

The cardiovascular mortality model for the cohort having complete data including end depression, anxiety and fitness is detailed in Table 6.25. A prescription for statins was the most powerful predictor of cardiovascular mortality with those having a prescription having one third the risk of those without. Once again, fitness was important, low baseline and failure to improve being powerful predictors of cardiovascular mortality. After fitness, age was important with those over 70 years at higher risk of cardiovascular death. Aspirin was associated with lower risk, as was being female.

Imputed data

The hazard ratios derived from the pooled imputed data (shown in Tables 6.16 and 6.17), were very similar to those from the complete cases model, both in size and direction, and fall within the confidence intervals given for the model estimates. This suggests that the analysis that used only the complete cases did not produce substantailly different results on account of removing the cases with incomplete data. There were 889 cases which were complete in baseline BMI as well as baseline fitness, depression and anxiety as in the primary model, and both all-cause and cardiovascular mortality models were derived for these, to test the importance of BMI. In neither case did BMI remain in the optimised model.

Main findings

Attaining a fitness level of VO_2 max>22ml/kg/min (defined here as high fitness) in the early months following a cardiac event or procedure is associated with improved long-term survival in those who have experienced a coronary event or procedure. (A useful comparison is that a young, untrained male would typically have a fitness level of VO_2 max 35-40 ml/kg/min and a cycling athlete VO_2 max 80 ml/kg/min.) High fitness at recruitment to the rehabilitation programme was likely to reflect high fitness before a coronary event or procedure, but there was no statistically significant difference between patients who improved from moderate fitness at recruitment to high fitness at graduation and those who maintained high fitness from recruitment, for those who completed the programme. Secondary prevention medications were also strongly associated with improved long-term survival in both all-cause and cardiovascular mortality. In particular, a prescription for statins or a prescription for aspirin were associated with lower risk of both all-cause and cardiovascular mortality. A prescription for ACE inhibitors was associated with higher mortality, but those on ACE inhibitors were the high risk cases. Patients having a CABG surgery and PCI have a significantly higher long-term survival from cardiovascular mortality than do patients with a myocardial infarction or angina.

Fitness for life confers significant potential benefits for those who may go on to experience a coronary event or procedure. Promotion of fitness after a coronary event or procedure, even for those already moderately fit, had potential for improved life expectancy.

The key role of medications in the early weeks after a cardiac event or procedure on reducing long term mortality both from cardiovascular causes and all-causes in patients experiencing a cardiac event or procedure had been demonstrated previously (Unal et al. [2004]) and in this model. We have no follow-up information on the adherence or changes to medication after graduation from the programme, so do not know how long patients continued with their medication. Given this, the strong effect of secondary preventative medication is striking.

The protective effect of female sex extends both to cardiovascular causes of death and to all-causes in cardiovascular patients, contrary to some other findings (Dallongevillle et al. [2010]). Weight does not appear to affect all-cause or cardiovascular mortality in cardiovascular patients, which is consistent with other studies (Romero-Corral et al. [2006], Shahian et al. [2012]). A prescription for ACE in-

hibitors at entry to the programme were closely related to a high risk level and were associated with increased risk of cardiovascular mortality. When fitness is measured and included in the survival analysis, BMI ceases to be a predictor of all-cause or cardiovascular mortality.

Numbers and percentages of the individual variables that were missing in the 2714 cases which met the inclusion criterion of having a baseline fitness measurement were shown in Table 6.18.

Variable	description	Missing	%
id	Unique identifier	1	0.04
Illness date	Date of the index event	0	0
Entry date	Date entered the programme	0	0
Graduate date	Date graduated from the programme	0	0
Death date	Date of death	0	0
Diagnosis ‡	A number from 1 to 8 where 1 is myocardial infarction (MI), 2 is coronary artery bypass grafting (CABG), 3 is percutaneous transluminal coronary angioplasty(PCI)with or without stenting, 4 is angina pectoris (AP), 5 is valve surgery (VS), 6 is other cardiac conditions e.g. cardiomyopathy, ischaemic and non-ischaemic heart failure (OC), 7 is non-cardiac conditions (NC) and 8 is myocardial infarction with percutaneous transluminal coronary angioplasty as a single episode of care, (MI+PCI).	0	0
Family History ‡	Family history, yes / no	0	0
Age ‡	Age in years	0	0
age category ‡	Age in categories: 1 is under 50, 2 is 50 to 59, 3 is 60 to 69 and 4 is 70	0	0
Sex ‡	Sex	0	0
Cholesterol	Cholesterol measurement at recruitment	537	19.8
Triglycerides	Triglycerides measurement at recruitment	787	29.0
Diabetes ‡	patient has diabetes, yes / no	0	0
Comorbidity	List of comorbidities in free text	0	0
Height	Height in Metres	1161	42.8
Weight ‡	Weight in kilogrammes at recruitment	20	0.7
Weight after ‡	Weight in kilogrammes at graduation	799	29.4
weight category before ‡	A is defined as under 75kg, 75-90kg as overweight, and over 90kg as C.	20	0.7

Table 6.18: Missing Data (continued next page)
‡indicates this variable was a significant predictor of mortality in a single-variable model.

Variable	description	Missing	%
sbp before ‡	Systolic blood pressure at recruitment	15	0.6
dbp before	Diastolic blood pressure at recruitment	26	1.0
sbp after ‡	Systolic blood pressure at graduation	742	27.3
dbp after	Diastolic blood pressure at recruitment	763	28.1
VO_2 before ‡	Fitness at recruitment in VO_2 max	0	0
VO_2 after ‡	Fitness at graduation in VO_2 max	1316	48.5
VO_2 category before ‡	Fitness category at entry	0	0
VO_2 category after ‡	Fitness category at exit	1316	48.5
anxiety before	Anxiety measured at recruitment using the hospital anxiety and depression . scale (HADS). A score between 8 and 10 indicates borderline anxiety, whilst over 10 suggests clinical anxiety.	89	3.3
anxiety after ‡	Categorisation by HADS at graduation	748	27.6
depression before ‡	Categorisation by HADS at recruitment	89	3.3
depression after ‡	Categorisation by HADS at graduation	749	27.6
over all health before	Patient's perception of overall health at entry - one of six domains from the Dartmouth Coop / Wonca charts used to assess functional health and quality of life. Only patients joining after April 1996 were assessed in this way	773	28.5
over all health after	1 (excellent) to 5(poor) patient's perception of overall health at graduation	1275	47.0
life before	patient perception of life in general at recruitment	773	28.5
life after	at graduation	1275	47.0
feelings before	patient perception of feelings at recruitment	773	28.5
feelings after	at graduation	1275	47.0
painful tension before	patient perception of painful tension at recruitment	775	28.6
painful tension after	painful tension at graduation	1276	47.0
physical fitness before	patient perception of physical fitness at recruitment	775	28.6
physical fitness after	physical fitness at graduation	1274	46.9
social support before	patient perception of social support available to them at recruitment	775	28.6
social support after	at graduation	1274	46.9

Table 6.19: Missing Data continued.
‡indicates this variable was a significant predictor of mortality in a single-variable model.

Variable	description	Missing	%
risk category before ‡	Risk category at recruitment (high, medium, low)	101	3.7
risk category after ‡	Risk category at exit	730	26.9
Smoking history ‡	coded 0 to 4 where 0 is never smoked, 1 is not smoked for 10 years or more, 2 is not smoked for between 1 and 10 years, 3 is recent quitter, and 4 is current smoker.	3	0.1
aspirin before ‡	Prescription for aspirin at recruitment, yes / no	1	0.04
aspirin after ‡	Prescription for aspirin at graduation	671	24.7
ace before ‡	Prescription for ACE inhibitor at recruitment	1	0.04
ace after ‡	Prescription for ACE inhibitor at graduation	671	24.7
bb before ‡	Prescription for beta blockers at recruitment	1	0.04
bb after ‡	Prescription for beta blockers at graduation	673	24.8
statin before ‡	Prescription for statins at recruitment	9	0.3
statin after ‡	Prescription for statins at graduation	678	25.0
full secondary prevention before	Prescription for aspirin and ACE inhibitors and beta blockers and statins at recruitment	1	0.04
full secondary prevention after ‡	Prescription for aspirin and ACE inhibitors and beta blockers and statins at graduation	1	0.04
resting heart rate ‡	resting heart rate at entry	7	0.3
hrateafter ‡	Heart rate after exercise	681	25.1
exercise sessions	number of exercise sessions attended to graduation or drop-out	105	3.9
imd2004score ‡	Index of multiple deprivation derived from post code	114	4.2
combined total comorbidity ‡	D'Hoore comorbidity score	0	0
occupation code ‡	Occupational Code 1-9: Managers & senior officials, Professional occupations, Associate professional, Administrative & secretarial, Skilled trade, Personal service, Sales & customer, Process, plant & machines, Elementary occupations	293	10.8
completer category‡	1 is completed the programme, 2 is started but did not complete, 3 is never started.	0	0

Table 6.20: Missing Data continued.
‡indicates this variable was a significant predictor of mortality in a single-variable model.

For the 1,029 cases with complete end fitness, anxiety & depression data see Table 6.21.

Variable	Male		Female		Total	
Number	895	(86.9%)	134	(13.1%)	1,029	(100%)
Mean years of follow-up (sd)	11.6	(3.9)	11.2	(3.8)	11.5	(3.9)
Mean age in years (sd)	61.1	(9.3)	63.1	(9.0)	61.3	(9.3)
	N	%	N	%	N	%
Age group under 50 years	104	11.6	14	10.5	118	11.5
Age group 50-59 years	275	30.7	29	21.6	304	29.5
Age group 60-69 years	336	37.6	52	38.8	388	37.7
Age group 70 years and over	180	20.1	39	29.1	219	21.3
Diagnostic Category						
Myocardial Infarction (MI)	456	50.9	73	54.5	529	51.4
Coronary Artery Bypass Graft (CABG)	269	30.1	34	25.4	303	29.4
Percutaneous Coronary Intervention (PCI)	81	9.1	12	9.0	93	9.0
MI + PCI	36	4.0	3	2.2	52	5.1
Angina	41	4.6	11	8.2	13	1.3
Other cardiac	12	1.3	1	0.7	39	3.8
Smoking history						
Never smoked	249	27.8	63	47.0	312	30.3
Not for 10 years+	285	31.9	19	14.2	304	29.6
Not for 1-10 years	34	3.8	4	2.9	38	3.7
Recent quitter	273	30.5	40	29.9	313	30.4
Current smoker	54	6.0	8	6.0	62	6.0
D'Hoore Co-morbidity score						
None	663	74.1	89	66.4	752	73.1
1 (least)	111	12.4	14	10.5	125	12.1
2	102	11.4	27	20.2	129	12.5
3	12	1.3	3	2.2	15	1.5
4 (most)	7	0.8	1	0.7	8	0.8
Diagnosis of diabetes	92	10.3	22	16.4	114	11.1
Family history of CHD	424	47.4	67	50.0	491	47.7
Weight at baseline						
A under 75kg	286	32.0	93	69.4	379	36.8
B 75-90kg	418	46.7	25	18.7	443	43.1
C over 90kg	191	21.3	16	11.9	207	20.1

Table 6.21: Baseline values for patients at recruitment to the programme: secondary analysis - 1,029 patients

Variable	Male		Female		Total	
	N	%	N	%	N	%
ACE inhibitor No	479	53.5	54	40.3	533	51.8
ACE inhibitor Yes	416	46.5	80	59.7	496	48.2
Aspirin No	31	3.5	10	7.5	41	4.0
Aspirin Yes	864	96.5	124	92.5	988	96.0
Statin No	338	37.8	43	32.1	381	37.0
Statin Yes	557	62.2	91	67.9	649	63.0
Beta blockers No	573	64.0	80	59.7	653	63.5
Beta blockers Yes	322	36.0	54	40.3	376	36.5
Occupation						
Managers & senior officials	152	17.0	11	8.2	163	15.8
Professional Occupations	98	10.9	7	5.2	105	10.2
Associate Professional	105	11.7	17	12.7	122	11.8
Administrative & secretarial	75	8.4	41	30.6	116	11.3
Skilled trade	250	27.9	7	5.2	257	25.0
Personal service	14	1.6	16	11.9	30	2.9
Sales and customer	18	2.0	10	7.5	28	2.7
Process, plant & machines	110	12.3	10	7.5	120	11.7
Elementary occupations	73	8.2	15	11.2	88	8.6

Table 6.22: baseline values of patients at recruitment to the programme continued: secondary analysis - 1,029 patients

Variable	Male		Female		Total	
	N	%	N	%	N	%
Fitness						
High baseline, no change	397	44.3	18	13.4	415	40.3
High baseline, deteriorate	5	0.6	0	0	5	0.5
Mid baseline, improve	203	22.7	28	20.9	231	22.5
Mid baseline, no change	151	16.9	25	18.7	176	17.1
Mid baseline, deteriorate	0	0	0	0	0	0
Low baseline, improve	86	9.6	20	14.9	106	10.3
Low baseline, no change	53	5.9	43	32.1	96	9.3
Depression						
Not depressed, no change	780	87.2	107	79.9	887	86.2
Not depressed, deteriorate	14	1.6	2	1.5	16	1.5
Borderline, improve	62	6.9	15	11.2	77	7.5
Borderline, no change	8	0.9	2	1.5	9	0.9
Borderline, deteriorate	2	0.2	0	0	3	0.3
Depressed, improve	26	2.9	7	5.2	33	3.2
Depressed, no change	3	0.3	1	0.7	4	0.4
Anxiety						
Not anxious, no change	607	67.8	73	54.4	887	86.2
Not anxious, deteriorate	40	4.5	10	7.5	16	1.5
Borderline, improve	99	11.1	17	12.7	77	7.5
Borderline, no change	45	5.0	10	7.5	9	0.9
Borderline, deteriorate	9	1.0	0	0	3	0.3
Anxious, improve	68	7.6	11	8.2	33	3.2
Anxious, no change	27	3.0	13	9.7	4	0.4
Median final estimated						
VO_2 ml/kg / min	25.3		18.1		24.5	
(10th, 90th percentiles)	(16.3,	35.0)	(10.4,	26.8)	(15.2,	35.0)
Median change from baseline in						
VO_2 ml/kg / min	3.3		2.6		3.2	
(10th, 90th percentiles)	(0.5,	8.1)	(0.0,	7.95)	(0.3,	8.1)

Table 6.23: Change from baseline for patients at graduation from the programme - 1,029 patients

Table of models with baseline and end fitness, depression and anxiety

All-cause survival model					
		complete cases model			**Imputed data model**
Model Term		Hazard Ratio	Confidence Interval		Pooled Hazard Ratio
			lower .95	upper .95	
Age category	under 50	1	-	-	1
	50-59	1.66	0.86	3.22	2.12
	60-69	2.48	1.31	4.69	3.05
	70+	5.54	2.90	10.61	5.80
Fitness:	High baseline, no change	1	-	-	1
	Mid baseline, improve	1.29	0.86	1.93	1.52
	Mid baseline, no change	2.32	1.59	3.38	2.39
	Low baseline, improve	2.84	1.87	4.31	2.84
	Low baseline, no change	3.78	2.40	5.94	3.80
Aspirin	Yes	0.38	0.24	0.59	0.63
	No	1	-	-	1
Ace inhibitor	Yes	1.45	1.12	1.88	1.29
	No	1	-	-	1
Sex	Male	1	-	-	1
	Female	0.64	0.44	0.92	0.60
Statins	Yes	0.74	0.57	0.97	0.72
	No	1	-	-	1

Table 6.24: Optimised all-cause survival model for secondary analysis, ordered by importance of variables to the model. Pooled hazard ratios are from multiple imputation of missing data.

Cardiovascular survival model					
		complete cases model			**Imputed data model**
Model Term		Hazard Ratio	Confidence Interval		Pooled Hazard Ratio
			lower .95	upper .95	
Fitness:	High baseline, no change	1	-	-	1
	Mid baseline, improve	1.08	0.58	2.00	1.61
	Mid baseline, no change	2.19	1.26	3.80	2.86
	Low baseline, improve	3.40	1.86	6.20	3.18
	Low baseline, no change	5.10	2.67	9.76	5.23
Statin	Yes	0.43	0.29	0.63	0.52
	No	1	-	-	1
Age category	under 50	1	-	-	1
	50-59	1.16	0.52	2.60	1.61
	60-69	1.83	0.84	3.98	2.47
	70+	3.31	1.47	7.43	3.82
Aspirin	Yes	0.39	0.22	0.71	0.60
	No	1	-	-	1
Sex:	Male	1	-	-	1
	Female	0.50	0.29	0.87	0.51
Ace inhibitor	Yes	1.58	1.08	2.30	1.39
	No	1	-	-	1
Diagnostic Category	Myocardial Infarction (MI)	1	-	-	1
	Coronary Artery Bypass Graft (CABG)	0.63	0.41	0.98	0.65
	Percutaneous Coronary Intervention (PCI)	0.20	0.04	0.83	0.47
	MI + PCI	1.18	0.36	3.86	0.77
	Angina	0.96	0.49	1.89	0.73
	Other cardiac	1.02	0.25	4.19	0.93

Table 6.25: Optimised Cardiovascular survival model for secondary analysis, ordered by importance of variables to the model. Pooled hazard ratios are from multiple imputation of missing data.

6.6 Artificial Neural Network for Modelling Long-Term Survival After a Cardiovascular Event

Artificial neural networks are a novel approach to modelling survival after a cardiac event. A Cox proportional hazards modelling approach has been used above with the data from the Basingstoke and Alton cardiac rehabilitation centre. These results show the comparison of the two approaches.

6.6.1 Principal findings

Both the Cox proportion hazards model and the ANN model produce survival estimates which were a good fit to the Kaplan-Meier estimate of survival from the data. The Cox model is constrained by its proportional hazards assumptions, and so the hazard over time is always proportional to the baseline hazard. The ANN is free from this constraint, and whilst for some covariates (e.g. statins, fitness) a proportional hazard is modelled, for others (e.g. ACE inhibitors, age) the hazard over time varies differently between categories of the covariate.
The comparison of hazard rates produced by the Cox and ANN models has shown the ability of the ANN to model hazard rates that vary over time in ways not constrained by a proportional hazards assumption and has produced some revealing variations in modelling this data set, which provide a basis for hypothesis generation and potential for tailoring interventions more closely to individual patients.

6.6.2 Introduction

The 6 variables found to be significant in the fitted Cox model (i.e. age, sex, fitness and medications ACE inhibitor, aspirin and statins) were used as inputs in a feed-forward artificial neural network with a single hidden layer comprising three hidden nodes (multi-layer perceptron). The network was regularised using weight decay with a parameter of 0.5. Optimisation was by minimisation of the log likelihood, and carried out in MATLAB® using scripts adapted from those developed by Johnston and Reeves (Reeves and Johnston [2008]) and offered in the book Intelligent and Adaptive Systems in Medicine edited by Keith Burnham and Olivier Haas. The interface between MATLAB® and the R statistical software (R Development Core Team [2011]) was via the R package *R.matlab*. The hazard rate for the ANN model was calculated using the equation 4.13. The Cox model was fitted using R statistical software (R Development Core Team [2011]) and the hazard rate calculated by multiplying the baseline hazard by the relevant coefficients using the fact that the exponential func-

tion obeys the basic exponentiation identity, $e^{x+y} = e^x \times e^y$ so that when calculating the hazard rate from the baseline hazard rate $h_i(t) = e^(\beta_1 x_{1i} + \beta_2 x_{2i}) h_o(t)$ can be calculated by the baseline hazard multiplied by the relevant exponentiated coefficients, both drawn from the fitted Cox model thus: $h_i(t) = e^{\beta_1 x_{1i}} \times e^{\beta_2 x_{2i}} h_o(t)$, See page 73.

Two aspects of the model were explored here, the fit to data and the differences in the model hazard rates.
Fitness was in 5 categories labelled 1,3,4,6,7 which represent the 5 combinations of before and after rehabilitation fitness categories seen in the data set; the combinations of baseline category and improvement represented by category 2 and 5 were not in the data set (see Table 6.23 on page 151). Age was in 4 categories (under 50 Years, 50-59, 60-69 and 70 and over) labelled 1 to 4 respectively, sex was labelled 1 for male and 2 for female, and the medications were 1 if prescribed, 0 otherwise. Altogether there were 320 possible combinations of these 6 variable categories. Each of the 320 was represented by a 6-figure number and 159 of the possibilities were to exist in the data set. These are shown in Tables 6.26, 6.27, & 6.28 starting on page 156, with details of the number of deaths from all causes observed in each variable category combination. Variable category combinations are indicated by a 6-figure number giving the categories of the variables in the sequence: age, sex, fitness, and medications ACE inhibitor, aspirin and statins, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins.

Variable Combination	Alive	Dead	Variable Combination	Alive	Dead
111000	1	0	211111	48	2
111010	16	1	213000	1	0
111011	22	1	213001	0	1
111110	4	0	213010	19	3
111111	23	2	213011	12	1
113000	1	0	213101	0	1
113010	13	1	213110	7	1
113011	3	0	213111	14	3
113101	1	0	214000	1	3
113110	2	0	214010	10	3
113111	4	0	214011	2	3
114010	2	0	214110	2	3
114011	0	1	214111	7	1
114110	2	1	216010	2	0
116101	0	1	216110	0	2
116110	1	1	216111	7	3
121000	1	0	217010	0	1
121011	2	0	217110	0	1
121111	2	0	217111	0	1
123001	1	0	221010	1	0
123011	1	0	221011	1	0
123110	1	0	221110	1	0
123111	1	0	221111	1	0
124010	1	0	223010	1	0
124011	1	0	223011	3	2
124110	1	0	223100	1	0
124111	0	1	223110	0	1
126110	0	1	223111	1	1
211000	1	0	224001	0	1
211001	1	1	224010	3	0
211010	35	3	224110	1	0
211011	52	4	224111	3	0
211110	13	0	226000	1	0

Table 6.26: Combinations of variables with numbers of patients in the data set alive or dead. The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins.

Variable Combination	Alive	Dead	Variable Combination	Alive	Dead
226011	1	0	321110	1	0
226111	1	0	321111	4	0
227011	2	0	323000	1	0
227111	2	0	323010	2	0
311001	1	0	323011	1	0
311010*	22	7	323110	2	0
311011*	29	9	323111	2	0
311101	0	0	324000	0	1
311110	3	4	324010	1	0
311111	52	4	324011	1	0
313001	1	0	324110	1	0
313010*	15	10	324111	2	0
313011	14	0	326011	4	0
313110	5	3	326110	0	2
313111	28	4	326111	5	1
314000	2	0	327010	1	1
314010*	18	8	327011	7	0
314011	13	4	327101	0	1
314100	0	1	327110	2	1
314101	1	0	327111	3	4
314110	1	6	411010	4	1
314111	16	5	411011	7	4
316000	0	0	411110	2	1
316010	1	5	411111	16	4
316011	6	1	413000	0	1
316101	1	1	413010	4	3
316110	4	2	413011	2	6
316111	4	2	413110	4	0
317010	1	1	413111	10	3
317011	4	4	414010	1	8
317110	0	1	414011	8	6
317111	6	2	414110	1	5
321011	1	0	414111	4	4

Table 6.27: Combinations of variables with numbers of patients in the data set alive or dead (continued). The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins.
* The combinations with an asterisk are shown in the survival plots

Variable Combination	Alive	Dead	Variable Combination	Alive	Dead
416000	0	1	423011	2	1
416010	3	3	423111	2	1
416011	3	4	424010	1	1
416100	0	1	424011	1	0
416101	1	1	424111	1	3
416110	1	4	426010	0	1
416111*	11	8	426110	0	1
417010	0	5	426111	0	1
417011	3	3	427000	0	1
417100	0	1	427010	0	2
417110	0	5	427101	0	1
417111*	7	7	427110	0	4
421110	1	0	427111	8	3
421111	1	1			

Table 6.28: Combinations of variables with numbers of patients in the data set alive or dead (continued). The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins.
* The combinations with an asterisk are shown in the survival plots.

6.6.3 Model fit to data

Both the Cox and ANN models produce survival curves which were comparable to the Kaplan-Meier estimate of the survivor function obtained from the data.
The Kaplan-Meier estimates of the survivor function are shown in Figure 6.18 for the variable category combinations in the data set that have at least 7 deaths and so provide sufficient information to make a useful comparison between the Kaplan-Meier curve and the model survival curves. These variable category combinations have between 14 and 38 cases observed in the data and are asterisked in the Tables 6.26, 6.27, & 6.28. These cases all fell in the 60+ age brackets, were all male and all have a prescription for aspirin. They have a variety of fitness categories and other medication regimes. Both the Cox proportional hazards model and ANN model estimates of the survivor curve for those same variable category combinations are also shown. The confidence intervals on each model's survival curves shown are 95% confidence intervals. Both models clearly produce survival curves that were a good fit to the Kaplan Meier estimate from data, and their confidence intervals overlap. The ability of the ANN to produce a more flexible model is clearly seen in each plot.

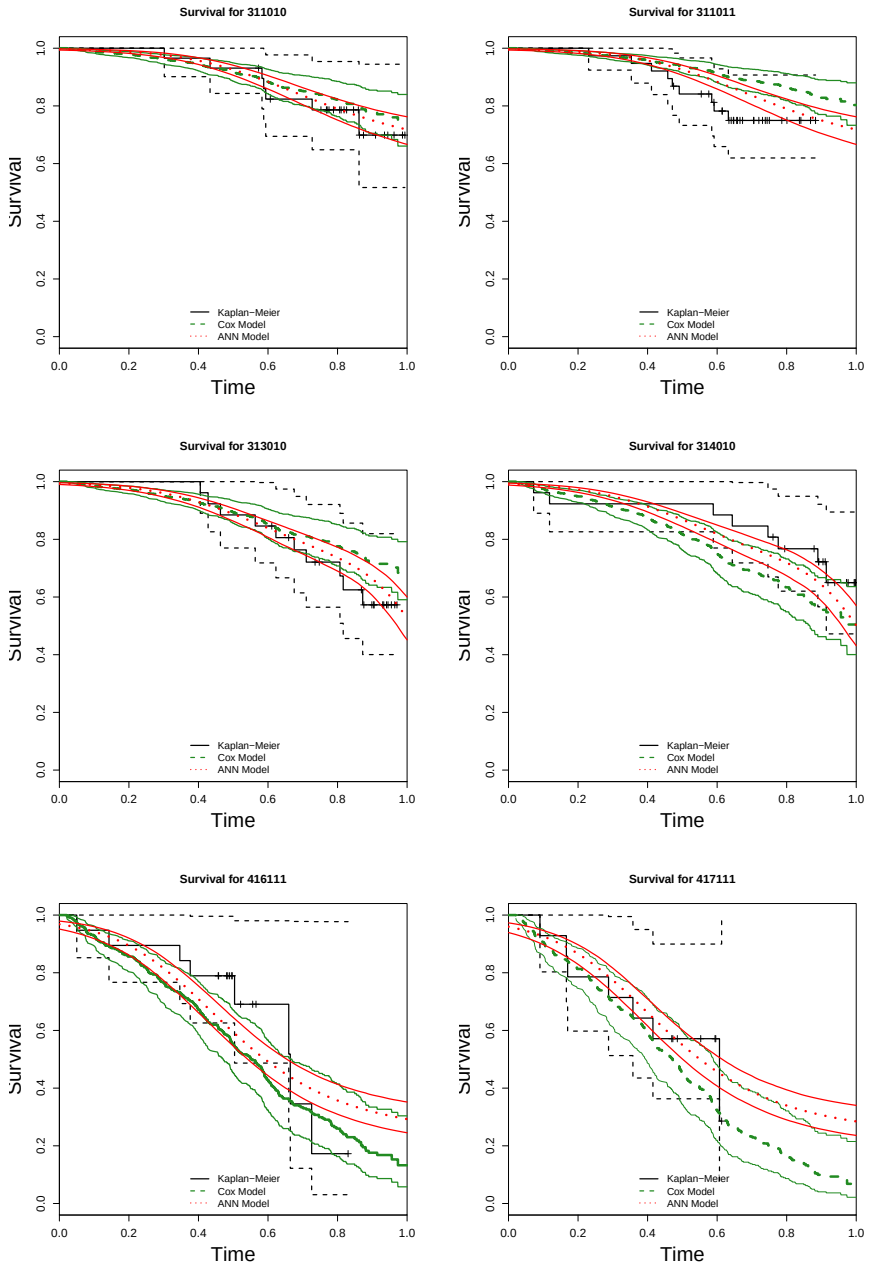


Figure 6.18: Survival plots showing the Kaplan-Meier estimate from the data, the Cox proportional hazards model and the ANN model for data subsets with 7 or more deaths, marked with an asterisk in the Tables 1, 2, and 3. 95% confidence intervals for each model are also shown.

6.6.4 Model predictions compared

For the 6 cases where there is the most data (marked with an asterisk in the Tables 6.26, 6.27, & 6.28), the contrasts between the hazard rates for the levels of each variable were plotted in turn, so that the hazard rate for males is plotted against females, each medication against not taking that medication, fitness levels and age categories against each other. The item being varied is indicated in the subset label by an asterisk in the place of its category level, so that 3*1011 indicates the hazard rate plot comparing males and females who were aged 60-69, had no prescription for ACE inhibitor, and did have a prescription for both aspirin and statins.

The contrasts between male and female are shown in in Figures 6.19 and 6.20 on page 162. The hazard rates on the right of each figure show the Cox proportional hazards model, and on the left the hazard rate given by the ANN model. Note that the hazard rates for the Cox proportional hazards model were produced by multiplying the baseline hazard rate by their relevant exponentiated coefficients in the optimised Cox model, and so were constrained to be proportional to each other in line with the model assumptions. This is seen in the case of the male / female contrasts, in that female hazard rates follow the same course over time and were always lower than the male hazard. The ability of the ANN model to capture non-linearity means that the hazard rates were not constrained to be proportional and can vary independently from each other. In the male / female contrasts, the hazard rates for the 60-69 age group were the same for males and females at early times, whereas at later times, the male risk was seen to rise sharply. In the 70+ age group, the female hazard rate was consistently lower than the male rate, and reading these together it should be noted that, with a mean follow-up time of 11.5 years (max 18 years 3 months), the cases recorded as being in the 60-69 age group at baseline would have mostly have reached the 70+ age group by the end of follow-up.

Contrasting the hazard rates produced by the Cox proportional hazards and ANN models for statins compared with no statins (Figures 6.21 and 6.22 on page 164), a very similar picture emerges. Both cases with and without a prescription for statins have a similar hazard rates at early times, and the hazard rate for the no statins cases rises steeply at later times for the 60-69 age group (when they will be 70+). Once again, the behaviour of the hazard rate curves for the 70+ age group is different. It is the same regardless of statins status and is much steeper than the younger age group. The ANN hazard plot suggests that the model switches from an exponential curve to a logistic curve between the two age categories when considering statins.

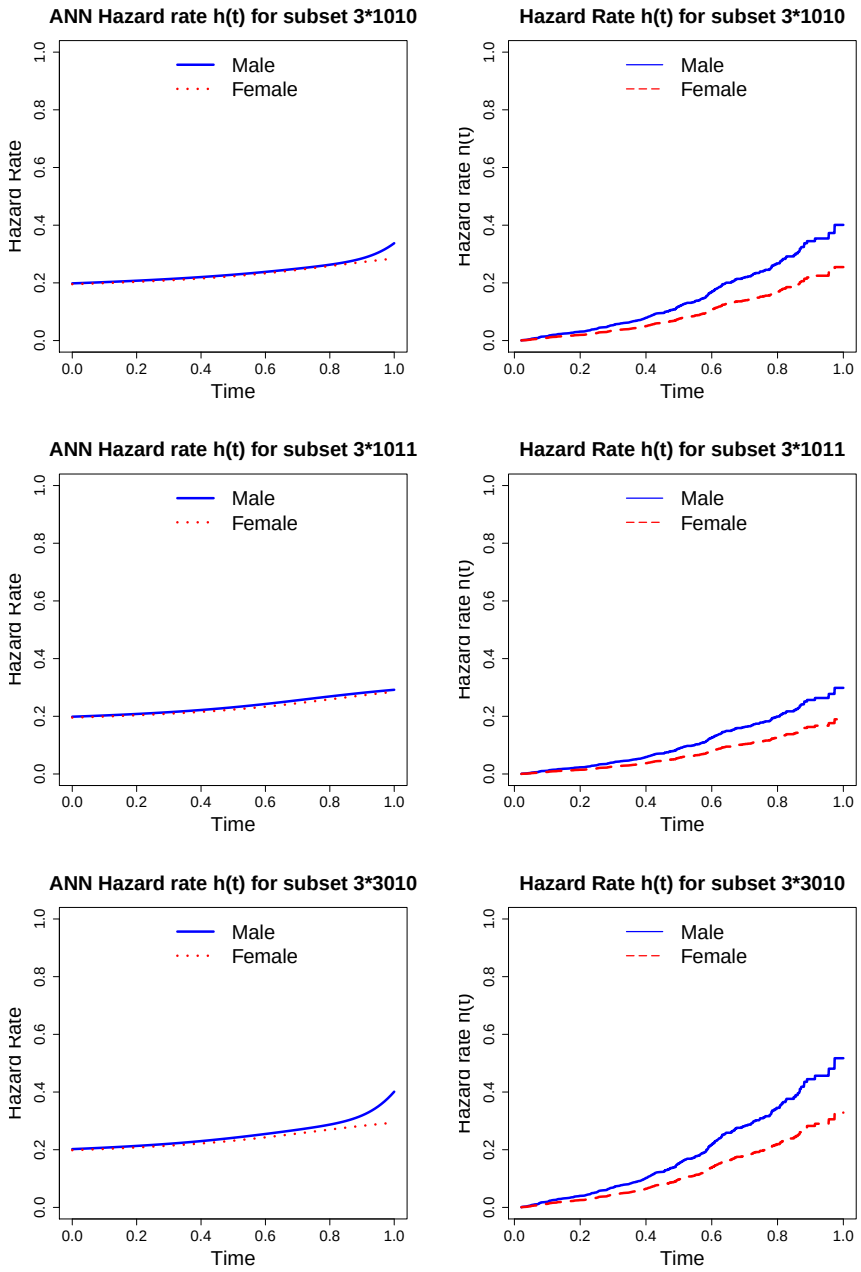


Figure 6.19: Hazard rate Males and Females. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13. The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins. * asterisked position is the variable being compared, e.g. male and female.

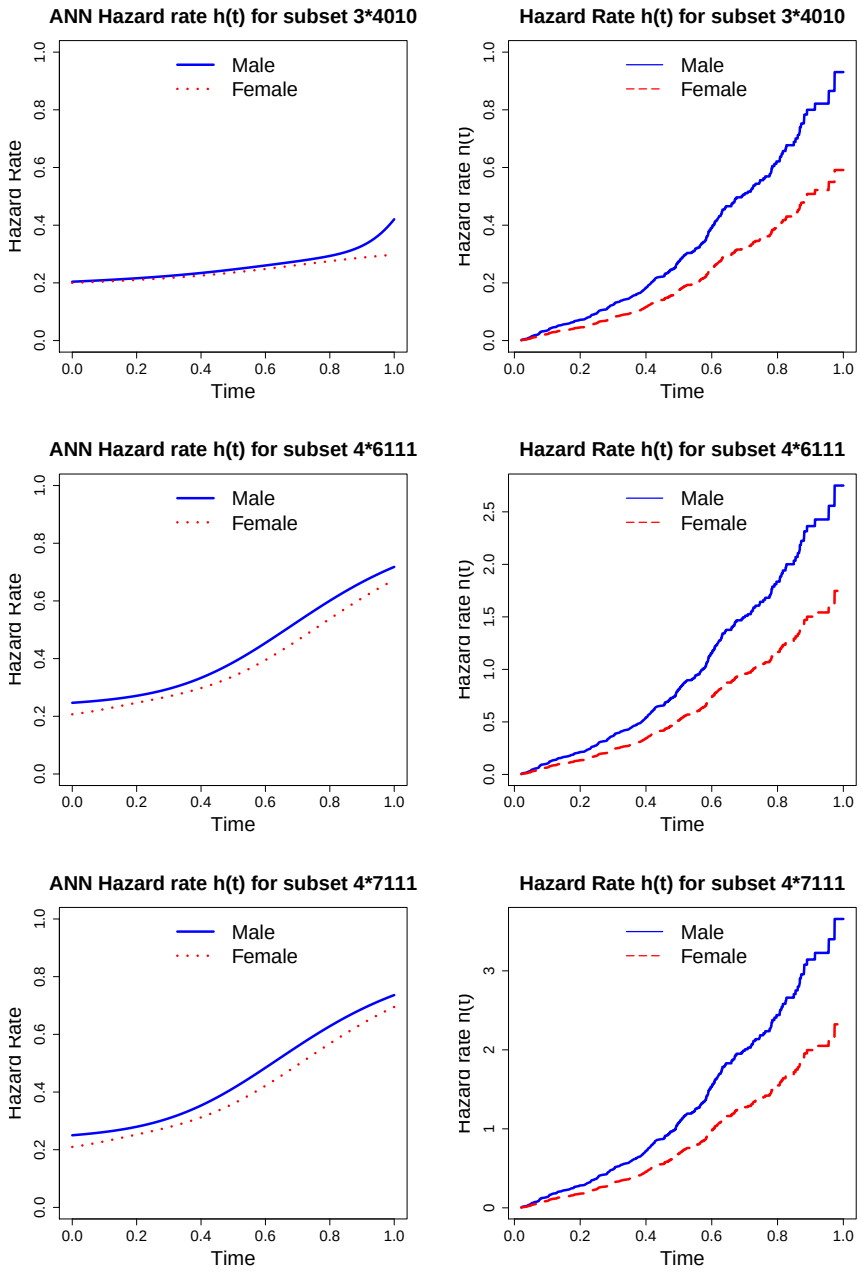


Figure 6.20: Hazard rate Males and Females. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13. The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 41777 represents age category 4 (70+), male, low fitness at start and end of rehabilitation, a prescription for ACE inhibitor, a prescription for aspirins and also for statins. * asterisked position is the variable being compared, e.g. male and female.

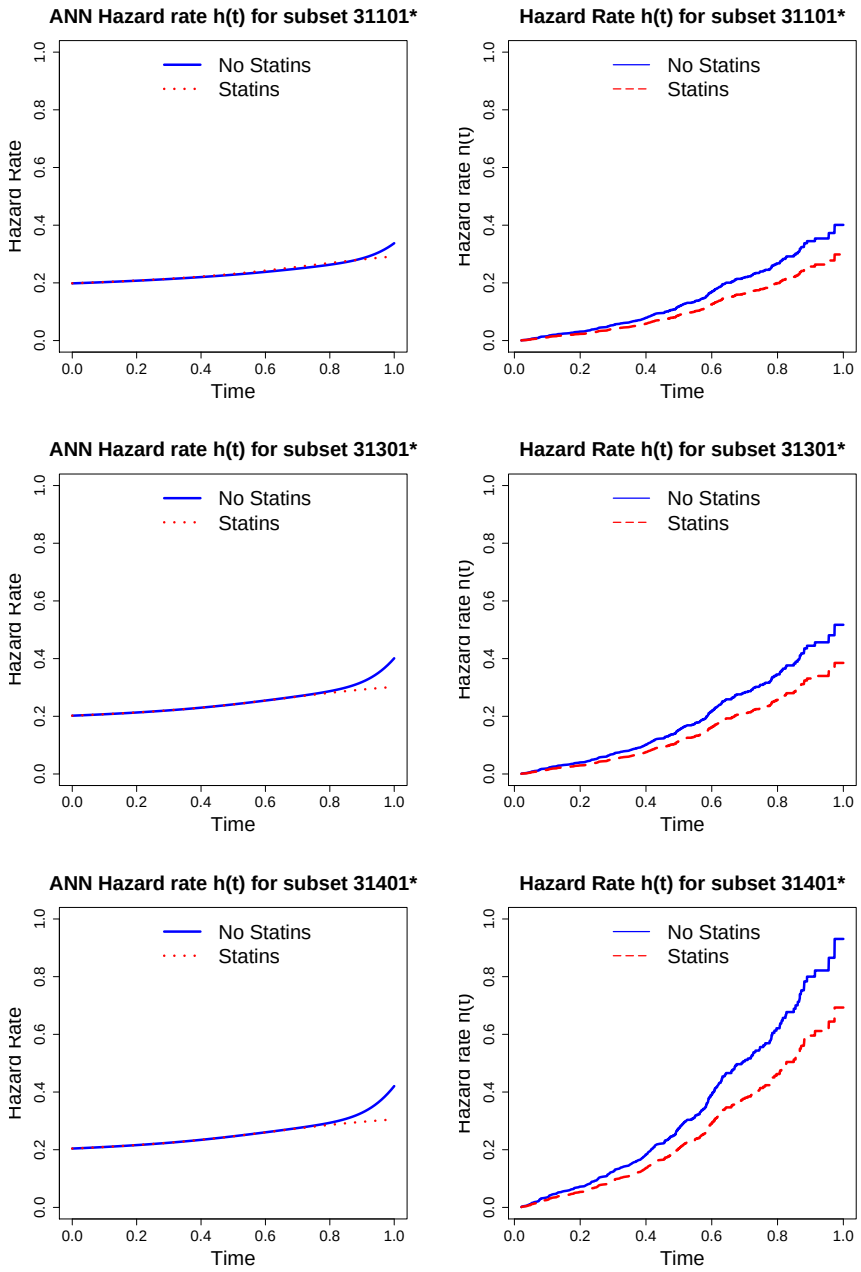


Figure 6.21: Hazard rate statins and no statins. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13. The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 311011 represents age category 3 (60-69), male, high fitness at start and end of rehabilitation, no prescription for ACE inhibitor, a prescription for aspirins and also for statins. * asterisked position is the variable being compared, e.g. male and female.

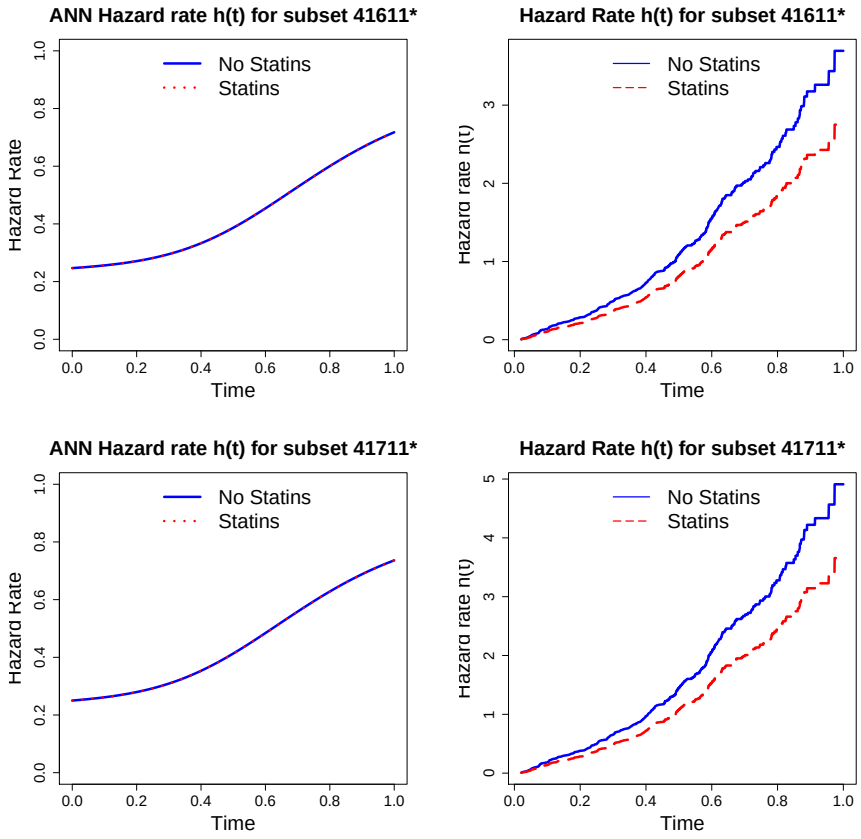


Figure 6.22: Hazard rate statins and no statins. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13. The combination label is made up from the values of the categorical variables in the sequence age, sex, fitness, ACE inhibitor, aspirin, statin, so that 41777 represents age category 4 (70+), male, low fitness at start and end of rehabilitation, a prescription for ACE inhibitor, a prescription for aspirins and also for statins. * asterisked position is the variable being compared, e.g. male and female.

The power of the ANN model to model non-linearity is clearly seen in comparing the plots of hazard rates for cases with and without a prescription for ACE inhibitors (Figures 6.23 and 6.24 on page 167). In the Cox model, the coefficients for ACE inhibitors are larger than 1, indicating that a prescription of ACE inhibitors was associated with higher risk (probably because ACE inhibitors were prescribed for the high risk cases, such as heart failure). This feature is also captured in the ANN model hazard rate curves, where a prescription for ACE inhibitor produces a higher hazard rate. The only difference between cases indicated by 311*10 and 311*11 is that the latter had a prescription for statins. The hazard rate is significantly modified, making the hazard rate for ACE and No ACE the same for early times, with hazard rate for No ACE increasing in later times. In the cases 311*10, 313*10 and 314*10, the changing fitness levels make no difference to the fact that having a prescription for ACE inhibitor gives a much steeper and higher hazard rate. However, in the 70+ age group 416*11 and 417*11, the hazard rate is very similar regardless of ACE prescription status, and is more similar to the higher ACE prescription curve for the younger age category, suggesting that increased age negates the lower risk of no ACE prescription.

In the contrast between aspirin and no aspirin prescription (Figures 6.25 and 6.26 on page 169) the hazard rate of the ANN model for those with an aspirin prescription is much lower than those without an aspirin prescription for the 60-69 age group, but the difference disappears for the 70+ age group, despite that group also having a prescription for statins, which affects the time course of the hazard rate in the younger, fitter group 3110*1, by removing the upward turn at later times for those with an aspirin prescription (red dashed line). It is noteworthy that for the 70+ age group with the lowest fitness, there is a small but clearly discernible benefit for the hazard rate with a prescription for aspirin, not seen in the fitter 70+ cases.

Age (Figures 6.27 and 6.28 on 171) is an important predictor of risk in the Cox proportional hazards model, and the hazard rate plots for this model show the proportional hazards between age classes, with the 70+ age group having the largest multiplicative factor and so a much steeper increase in hazard rate over time. The ANN model also captures this feature, producing a proportional-like hazard for the 3 younger age groups, which were more similar to each other than in the Cox model, and a steeper hazard curve over time for the cases in the 70+ age category at the beginning of the observations. This pattern was repeated for all the sample variable category combinations, with a slight modification in the case of the lowest fitness categories, where the 60-69 age group showed a steep increase in hazard rate at later times, when it would be entering the 70+ age category.

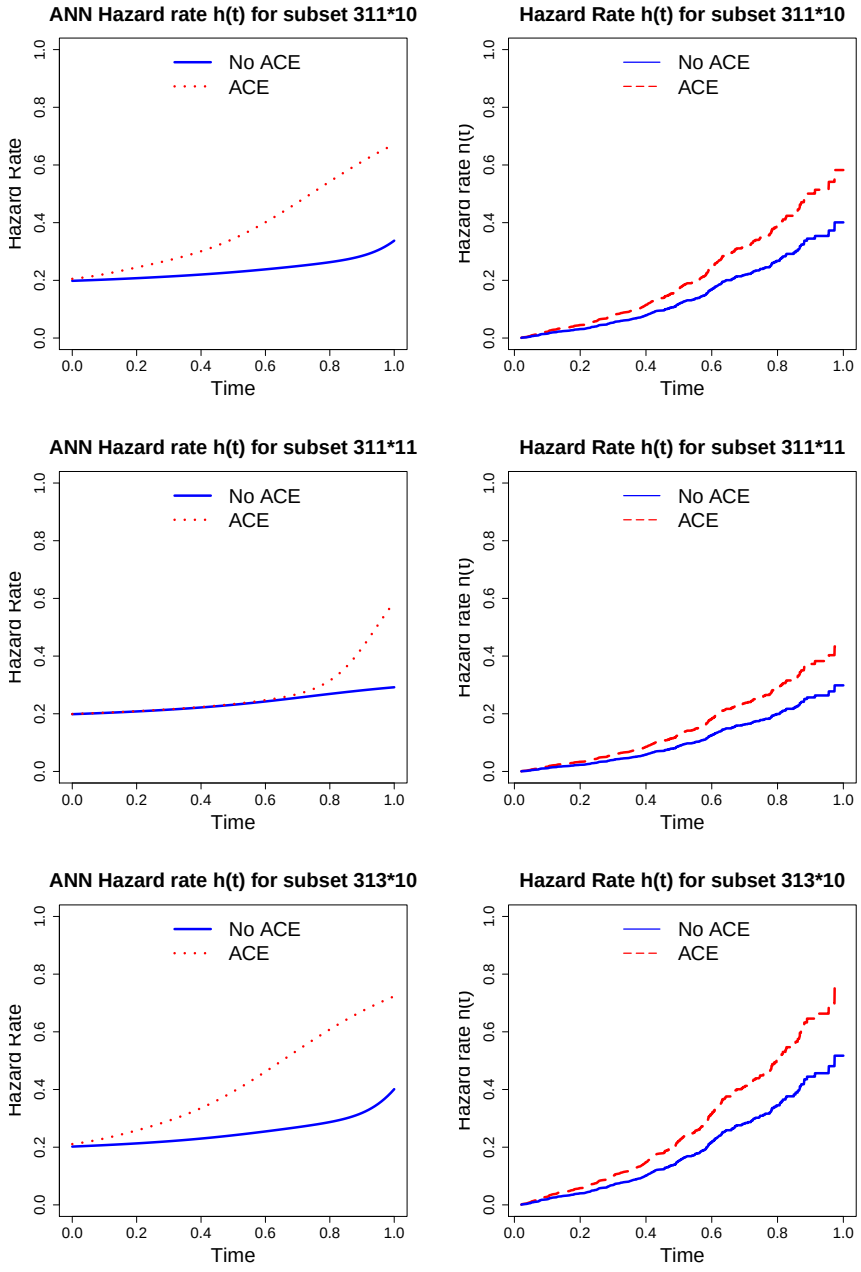


Figure 6.23: Hazard rate ACE inhibitor and No ACE inhibitor. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13.

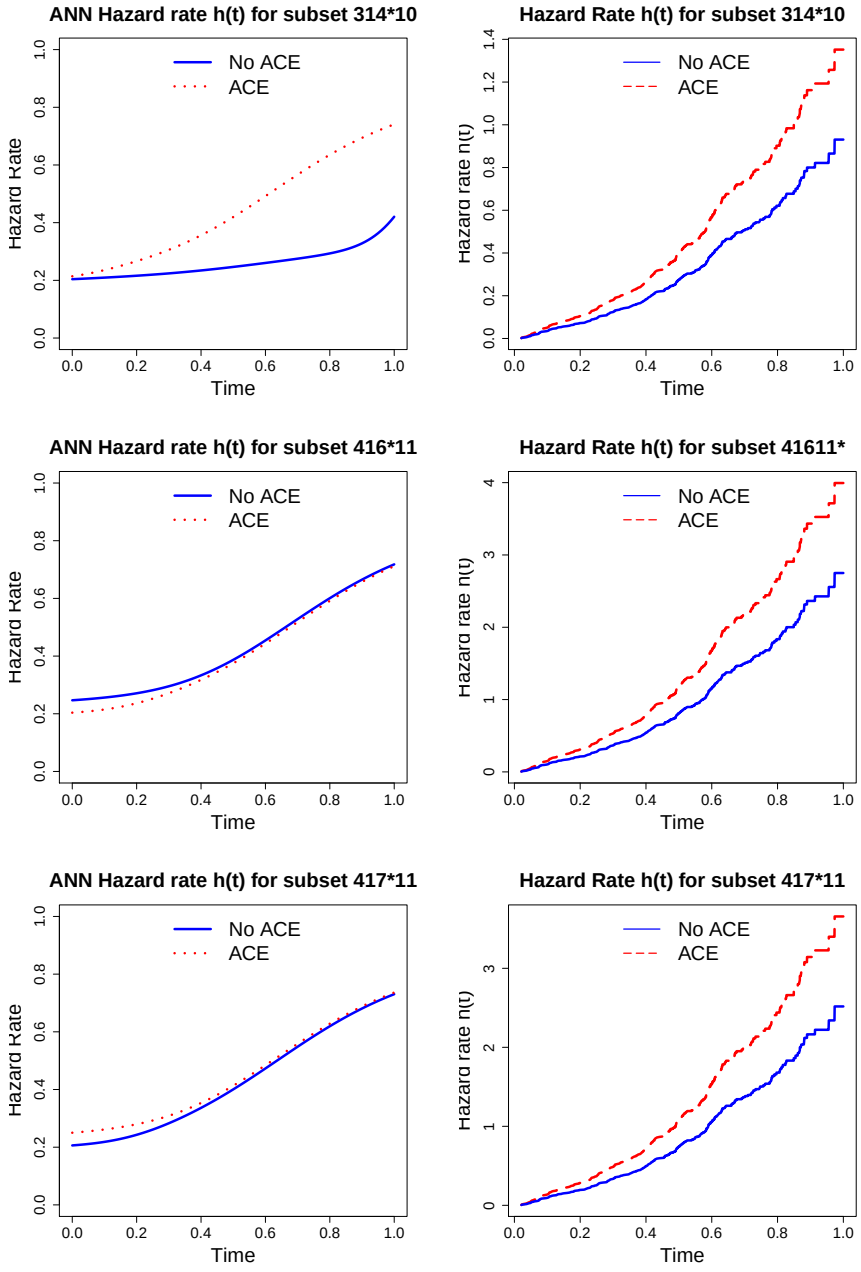


Figure 6.24: Hazard rate ACE inhibitor and No ACE inhibitor. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13.

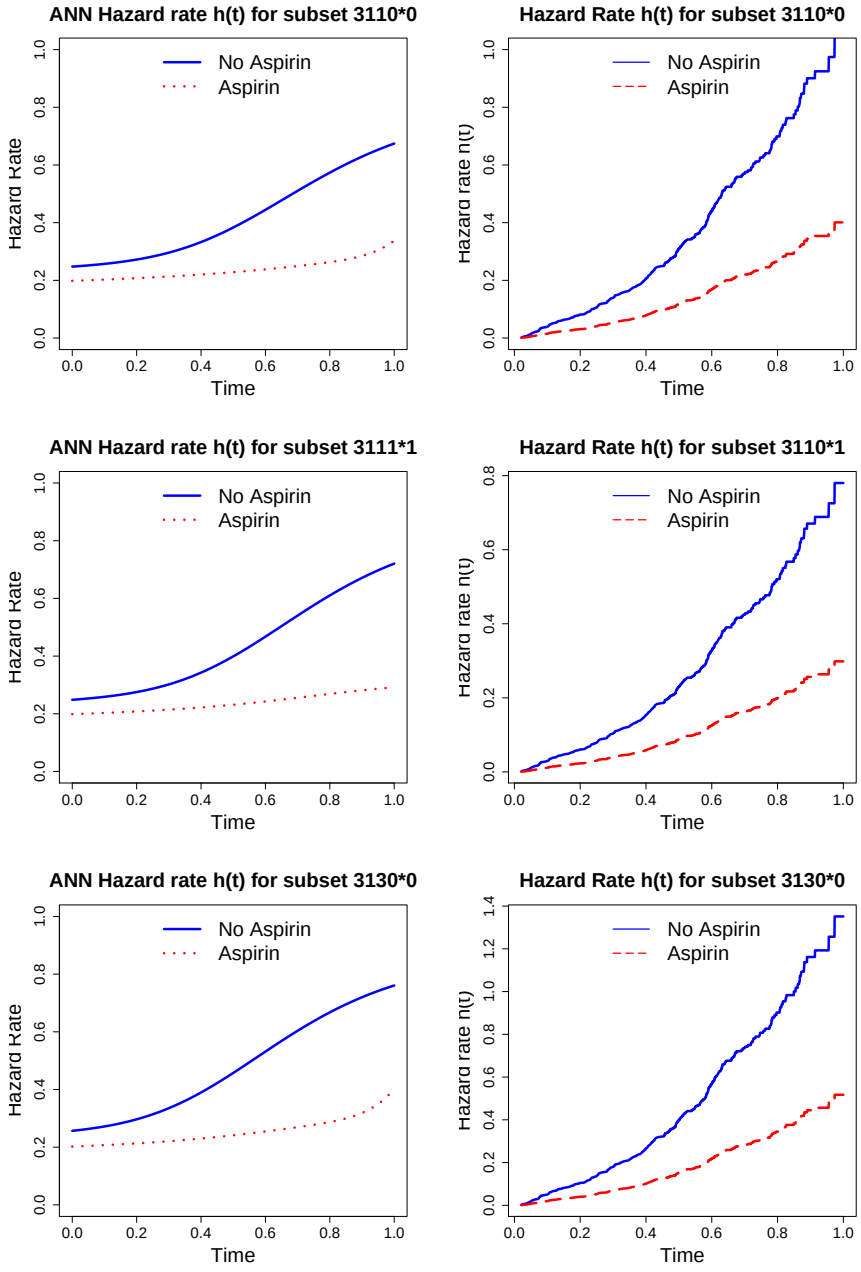


Figure 6.25: Hazard rate aspirin and no aspirin. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13.

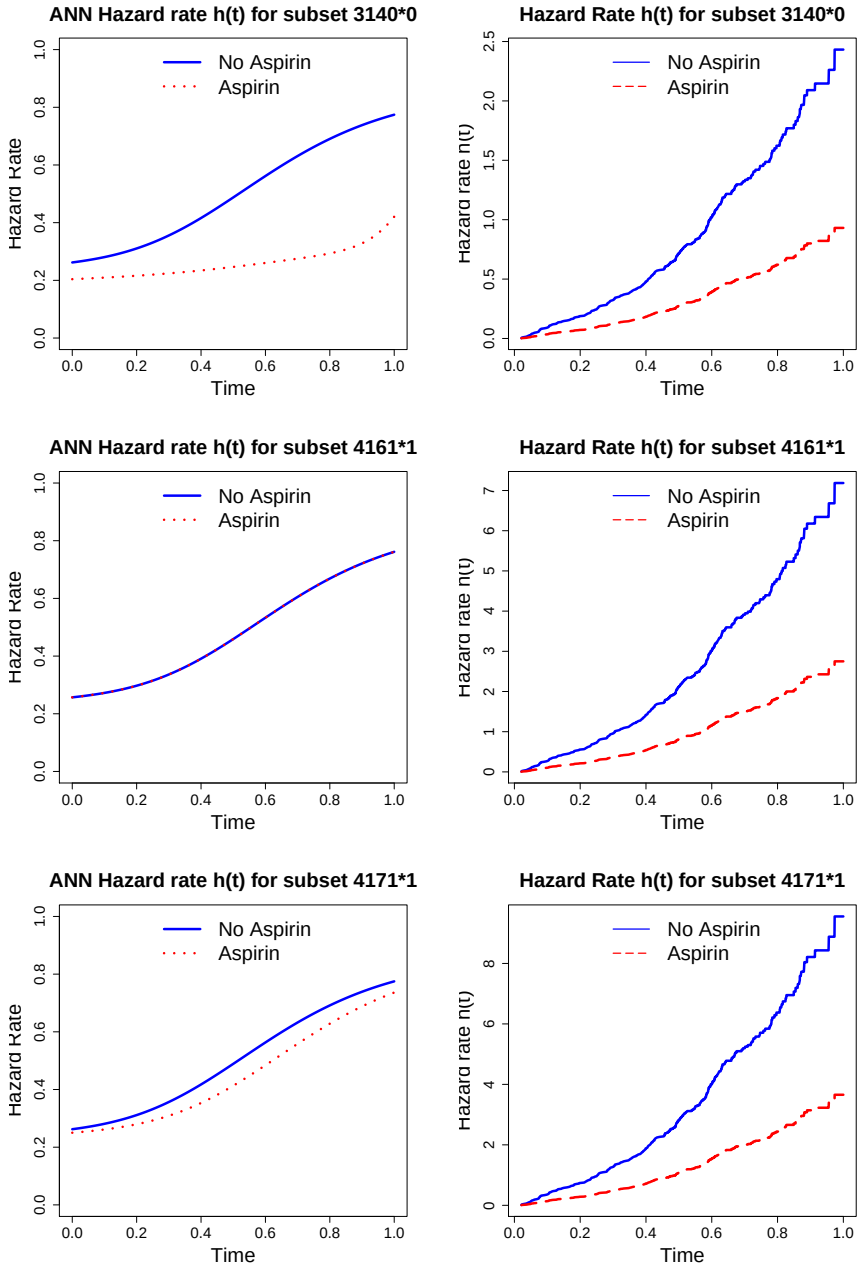


Figure 6.26: Hazard rate aspirin and no aspirin. The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13.

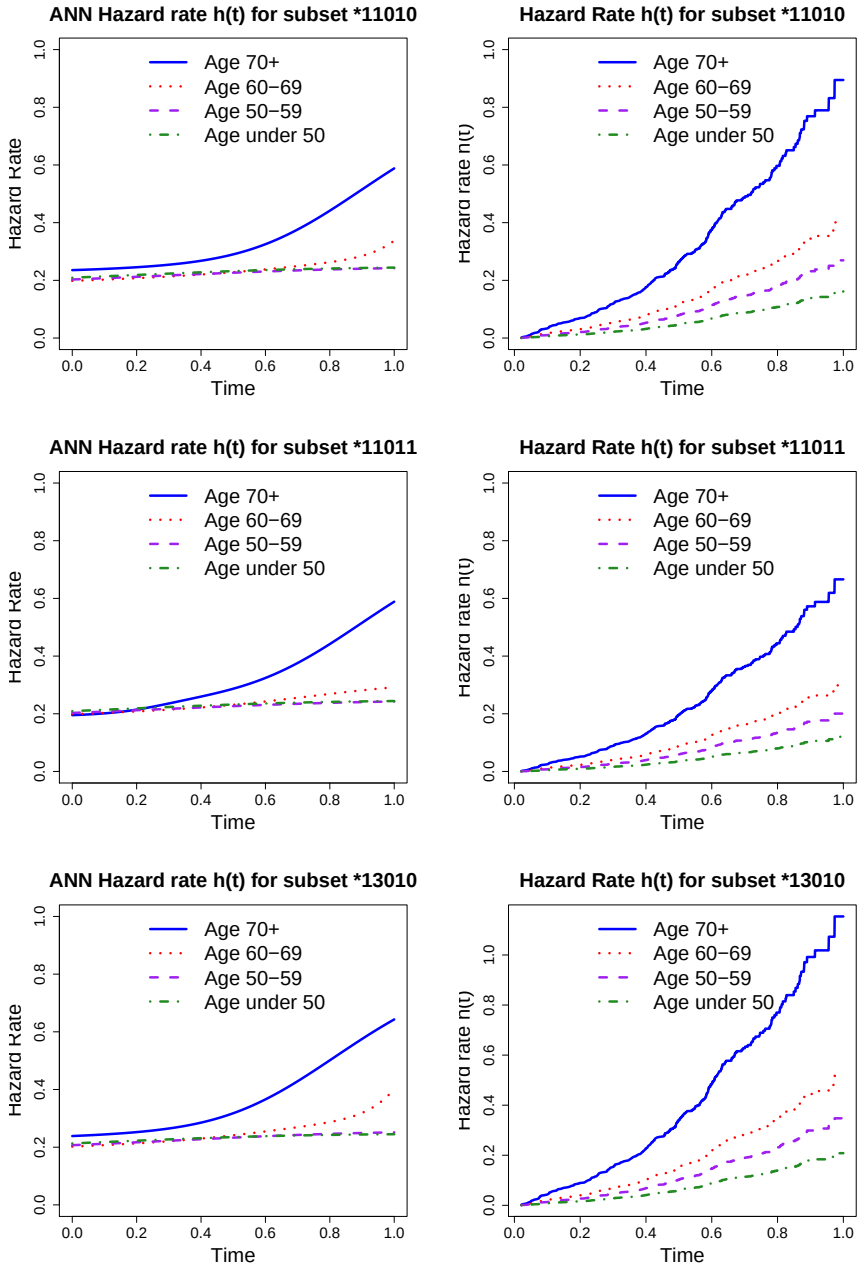


Figure 6.27: Hazard rate in different age groups (age at recruitment). The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13.

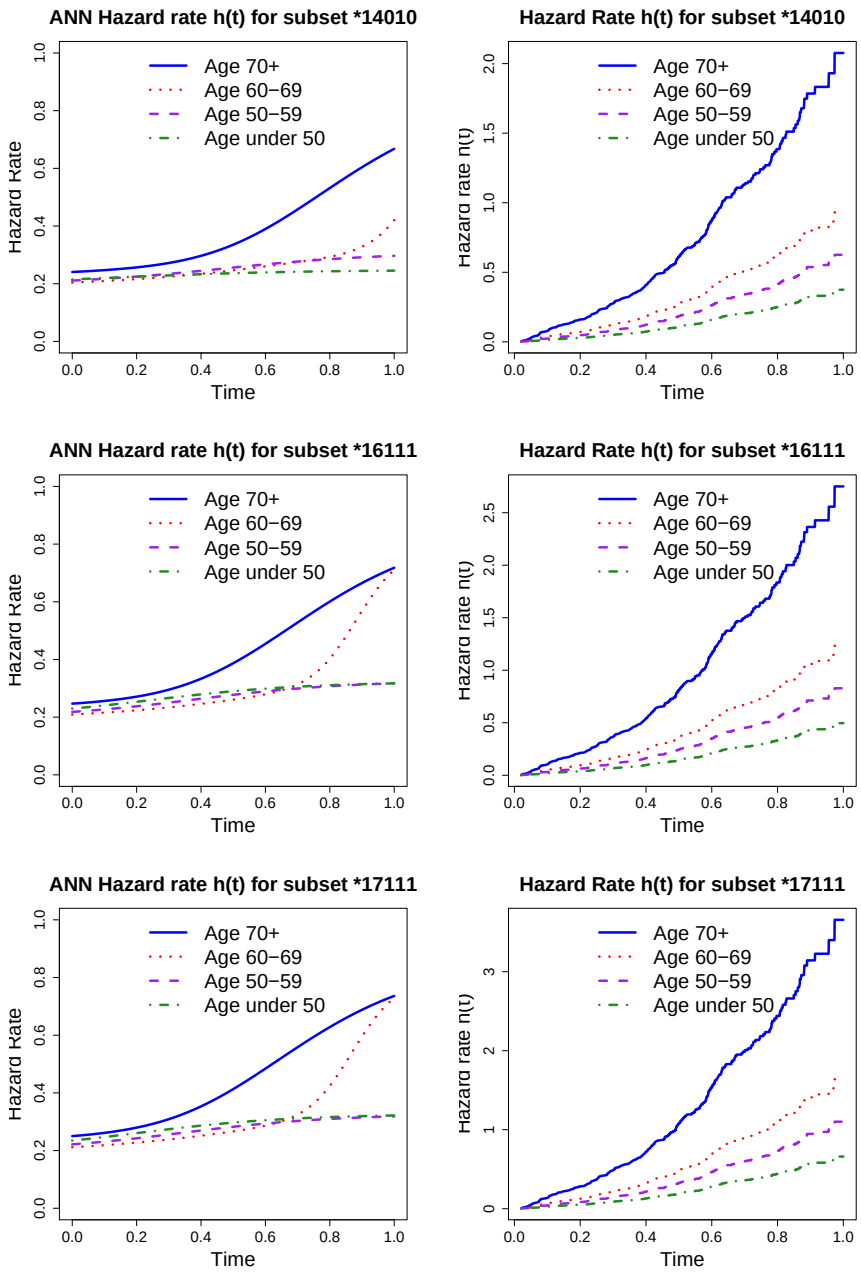


Figure 6.28: Hazard rate in different age groups (age at recruitment). The Cox proportional hazards model is shown on the right of the figure, and the ANN model on the left. The hazard rate for the Cox model is calculated by multiplying the baseline hazard by the relevant exponentiated model coefficients, and the ANN hazard rate is calculated using equation 4.13. Fix the headings in two plots *16111 and *17111

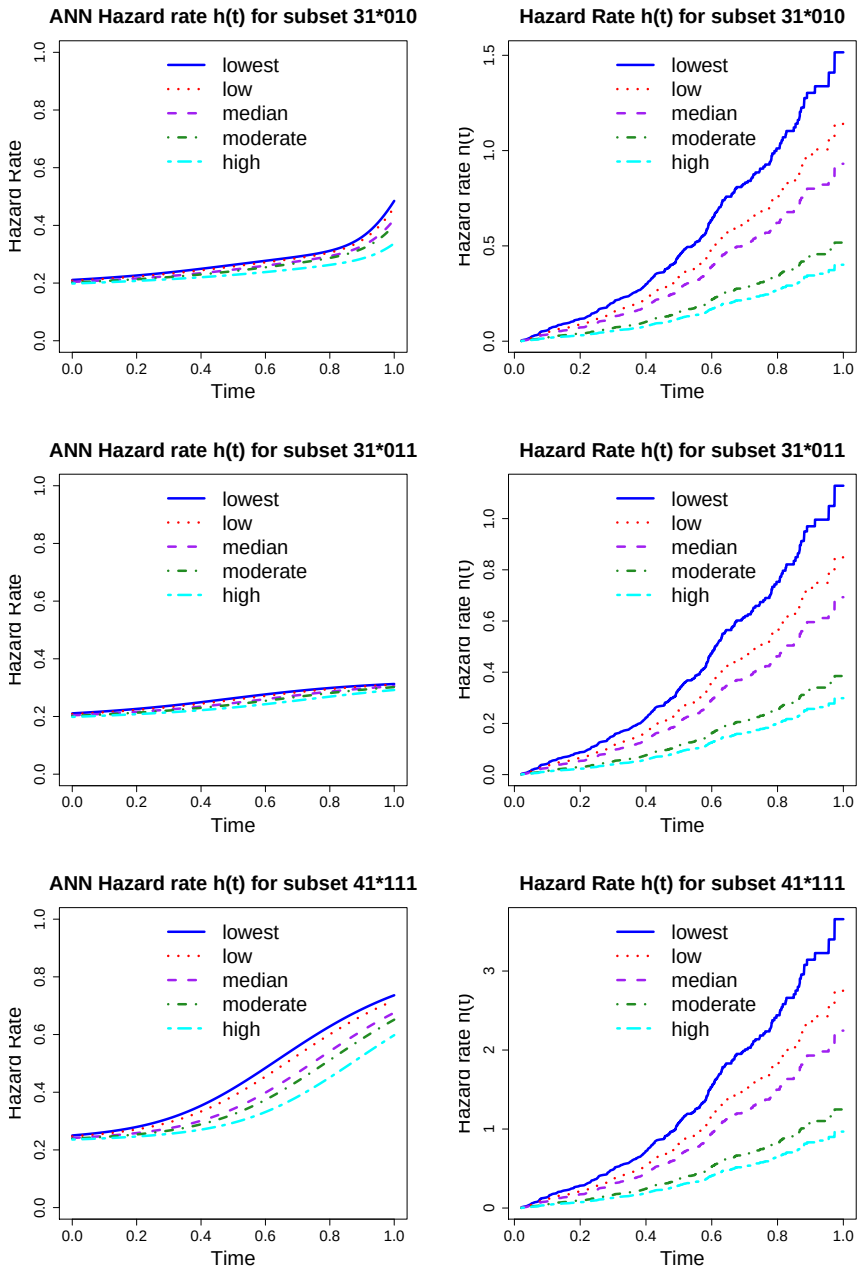


Figure 6.29: Hazard rate in different fitness categories. Change the labels to 13467 . Fitness is in 5 categories labelled 1,3,4,6,7 which represent the 5 combinations of before and after rehabilitation fitness categories seen in the data set: high fitness at start of rehabilitation maintained (1), mid fitness at start improved to high fitness (3), mid fitness at start maintained (4), low fitness at start improved (6), low fitness at start and end (7).

Fitness (Figure 6.29 on page 173) shows a proportional hazard rate for both the Cox model, where it is constrained to be proportional, and in the ANN model where it is not. There were, however, still clear differences between the models. The sharp rise in the hazard rate at later times seen for cases 31*010 is not seen in 31*011, where the only difference is a prescription of statins for the latter. The oldest age group also had a steeper rise in hazard rate over time than the 60-69 age group in every fitness group. The spread in the 70+ age group is is also greater than the 60-69 age group.

The comparison of hazard rates produced by the Cox and ANN models has shown the ability of the ANN to model hazard rates that vary over time in ways not constrained by a proportional hazards assumption and has produced some revealing variations in modelling this data set, which provide a basis for hypothesis generation and potential for tailoring interventions more closely to individual patients.

A full discussion of the results presented in this chapter is given in Chapter 7.

Chapter 7

Discussion

This chapter begins with a discussion of complexity science and health care as a complex system. This is followed by the results in the context of the literature which is divided into the same four subdivisions in the results and literature review, namely LCA for the BeST data, ANN for the BeST data, long-term survival after a cardiac event using a Cox model and ANN for survival modelling. Following this discussion, each of the research aims stated in the introduction have been revisited and evaluated. The chapter closes with a discussion of further research directions.

7.1 Complexity Science

The New England Complex Systems Institute introduces Complexity Science thus: 'Complex Systems is a new field of science studying how parts of a system give rise to the collective behaviors of the system, and how the system interacts with its environment. Social systems formed (in part) out of people, the brain formed out of neurons, molecules formed out of atoms, the weather formed out of air flows are all examples of complex systems. The field of complex systems cuts across all traditional disciplines of science, as well as engineering, management, and medicine. It focuses on certain questions about parts, wholes and relationships. These questions are relevant to all traditional fields.' (The New England Complex Systems Institute [2013]).

7.2 Complexity in health care

It is well-established that health care and health itself are complex systems (Plsek and Greenhalgh [2001], Wilson and Holt [2001], Plsek and Wilson [2001], Fraser and

Greenhalgh [2001], Topolski [2009]). In a series of 4 articles on Complexity Science and its relevance to health care in 2001, the British Medical Journal defined complex adaptive systems: 'A complex adaptive system is a collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent's actions changes the context for other agents. Examples include the immune system, a colony of termites, the financial markets and just about any collection of humans (for example, a family, a committee, or a primary health care team).' (Plsek and Greenhalgh [2001]).
There follows an overview of complexity in various aspects of health and health care and how ideas from Complexity Science have been used to address the challenges in the health care system. At the end of this chapter is a statement of the aims of this research.

7.2.1 The Complexity challenge

Conventional models of the universe and its subsystems as machines with simple rules and absolutely predictable and controllable outcomes, have been the framework for understanding medicine and health care as evidenced by the orientation of medicine around organ-based disciplines and physiological processes and organisation around linear, hierarchical relationships and rules (Plexus [2003]). The shift from a bio-mechanical view of the body followed the realisation that no part is constant, independent or predictable; complex systems typically have fuzzy boundaries, membership of sub-systems can change and agents can simultaneously be members of several systems (work, family, hobby, communities, for example). Conventional, reductionist scientific thinking assumes that we shall eventually figure it all out and resolve the unresolved issues, whilst Complexity theory accommodates uncertainty and inherent tension between different parts of the system (Plsek and Greenhalgh [2001]). Complex systems often exhibit non-linear behaviour and sensitive dependence on initial conditions, such that small differences in the initial variables leads to large differences in outcomes. This makes complex systems fundamentally unpredictable over time. Despite this lack of predictability, it is often possible to make generally true and practically useful statements about a complex system. Attractors are patterns of behaviour within a complex system. A health care example is the tendency of psychotherapy patients to accept counsellor's advice when it is framed in ways that enhance their core sense of autonomy, integrity and ideals. Order, innovation and progress can emerge within a complex system without them being imposed from outside, e.g termite colonies, driving conventions and behaviour patterns in repeated meetings. Whilst a surgical theatre team performing a routine procedure

might think in mechanistic terms and fall into agreed roles, few situations in modern health care have such a high degree of certainty and agreement about procedure and, for those, rigid protocols may not be appropriate. It is claimed that mechanistic thinking is no longer appropriate in health care and that flexible responses based on autonomy and creativity should be embraced (Plsek and Greenhalgh [2001]).

7.2.2 Complexity and clinical care

The complexity of both biological and social systems are said to be the reason few, if any, human illnesses can be said to have a single cause or cure. Glycaemic control in a diabetes patient is an example of this inherent complexity. The management of the condition based on simple cause-and effect understanding of insulin and blood glucose levels led to more hypoglycaemia episodes than an approach which emphasized the patients' own intimate knowledge of their own profiles and body rhythms and their experimentation with practical responses to variations (Wilson and Holt [2001]). Another complex systems approach to modelling glucose levels in diabetes found a difference in how the levels were correlated at different time points between people with diabetes and those without (Khovanova et al. [2012]). Where lifestyle changes are needed to improve health, the patients' readiness to change is a key element of success, and it is suggested this readiness comes when the system is far from equilibrium, and a new attractor can be accommodated (Wilson and Holt [2001]). The wider context of other advice which the patient may receive or seek, whether from a social network, internet source, alternative therapy etc. has been explored and, for example, there is evidence that social networking is changing people's health-related experience through the emergence of on-line expert patient groups and the discovery of community around particular health conditions, which exhibit both complementarity and competition with the patient-doctor encounter (Griffiths et al. [2012]).

7.2.3 Complexity in leadership and management

Leadership and management, including in health care, depends largely on productive interaction. However, the organisation and management of health care is enacted as if a well functioning organisation were akin to a large, well-oiled machine. This leads to the imposition of separate budgets and targets for primary, secondary and social care which promotes an internal focus for each part (Plsek and Wilson [2001]). One example is the target to administer thrombolytic drugs to eligible patients within 60 minutes of the onset of a myocardial infarction (heart attack) which is confounded

with a target for administration within 30 minutes of arrival at hospital, with ambulance response times and also any delay by the patients in calling for help, so that the patient may miss out on full intended benefit despite the individual targets being met. It is suggested that Complexity-based organisational thinking leads to whole-system goals and pooled resources which generate relationships among the stakeholders which provoke creative ideas organisation. A complex adaptive systems framework could improve outcomes in transitions between health care settings, including hospital discharge, which was found to be associated with high re-admission and adverse event rates in the newly-discharged (Geary and Schumacher [2012]). Where the organisation as a machine paradigm is in place, initiating change is seen as akin to a replacement part being fitted, and resistance to change tackled with the hammer of strong leadership, sanctions, or strict budget controls (Plsek and Wilson [2001], Johnson et al. [2010]). Instead, it was suggested, a complex systems approach considers resistance to be the pull of a different attractor, such as a desire to focus on an under-served patient group. Understanding the context and attitude in which change is attempted can offer strategies to adapt without loss of focus on what is done well. Variation within a complex organisation arises out of interaction between different factors. The strict elimination of all variation is an appeal to the machine metaphor and will have the byproduct of stifling innovation. In the nursing home context, complexity-type management led to better patient outcomes (Anderson et al. [2003]). In the US health care system, multi-scale complex systems analysis suggests that the overall system can be dramatically improved by establishing two separate but linked health care systems with distinct organisational forms: one a high-efficiency system performing large-scale repetitive tasks such as screening tests, inoculations, and generic health care, and the other a high-complexity system treating complex medical problems of individual patients (Bar-Yam [2006]).

7.2.4 Complexity and medical education

As the government and managers seek to deliver a health care system that is ever safer, constantly up to date, and focuses on patients' changing needs, medical education needs to deliver not only competence (knowledge, skills and attitude) but also capability (the ability of an individual to adapt to change, generate new knowledge, and continue to improve their performance). Capability cannot be taught or passively assimilated, but requires existing competencies to be adapted and tuned to new circumstances (Plsek and Wilson [2001]). In the modern information age, an expert is someone who knows how to access knowledge efficiently and can form conceptual links between seemingly unrelated areas. Learning how things are inter-

connected is often more useful than learning about the pieces for applying learning to new contexts, but this is not a strength of traditional curricula. Checklist-driven approaches to clinical care become useful only after the problem has been thoroughly understood; the acquisition of understanding requires intuition and imagination (Weiss et al. [2011]). Education that makes use of insights from complex systems e.g. through storytelling and group problem-based learning helps build on these distinctly human capabilities (Plsek and Wilson [2001]).

7.2.5 Health as a complex system

Health is traditionally defined in western medicine as absence of disease, and life defined in terms of death (Topolski [2009]). As the nineteenth century biomedical model of health replaced theories of humours and spirits, it was itself expanded by Engel's 20th century biopsychosocial model. The experience of health results from many different interconnected factors, changes in which can have a non-linear effect on the whole, leading to the conclusion that Engel's model can be further improved using the concepts from Complex Systems science. This new model asserts that living systems are open to their environment (e.g.weather, Parsons et al. [2011]), and borrows the language of statistical physics in describing ageing (Topolski [2009]) and illness (Burton et al. [2010]) as increase in entropy and decrease in complexity. This model describes a log-normal distribution of maximum health potential with a minimal sum of health for survival. The theoretical maximum health potential is modified by the influence of surroundings on the six paired parameters which describe health, vis. physical and environmental, emotional and social and cognitive and semiotic which sum to describe the experience of health and well-being. This Complexity science based model unites the concepts of identifiable pathologies and subjective illness in one conceptual model.

7.2.6 The complex individual

Griffiths et al. [2010] state that 'This notion of the individual has much in common with that of the complex system: open systems with emergent properties and transformational potential. The emergent property of a complex system is the quality that indicates its nature as a whole, and cannot necessarily be predicted or explained by detailed examination of the parts of the system. There is constant interaction within a complex system and between the system and its environment, and so constant change. The nature of this change might be small adjustments or adaptations, but the system remains qualitatively similar; however, change might

be transformational and, if this occurs, the system becomes qualitatively different. The nature or quality of the system is the emergent property of the system at that time. Complexity science is concerned with describing and explaining the patterns of change of a system, both the transformations and the adjustment and adaptation between transformations. This social theory on the nature of the individual and complex systems suggested that we could consider individuals as complex systems with emergent properties (what we have referred to above as how a person is overall, her or his qualitative state) that are constantly changing, and that the pattern of change might be important. An individual's pattern of adjustment and adaptation, and any transformations, could result from all he or she does and all that happens to him or her over time. '

This description attempts to capture some of the reasons why individuals do not respond uniformly to stimuli (Thorogood et al. [2006]). For example, the stigma of depression intersects with responses to depression: some people get more depressed, but for others stigma makes them get up and prove themselves (Boardman et al. [2011]). Similarly, alcohol has both positive and negative health effects: same exposure, different outcomes. In a linear model, the coefficients of these may be both positive and negative and so tend toward zero, despite the effect being significant. Human development is also a complex process (McGonigle-Chalmers and Kusel [2012], McGonigle-Chalmers et al. [2008]) and human interactions are key to health and disease (Danon et al. [2012], Dong et al. [2012], Centola [2011], House [2011], Centola [2010], Christakis and Fowler [2008], Fowler and Christakis [2008], Centola [2007], Christakis and Fowler [2007], Centola et al. [2005]).

On every scale, health and health care are complex systems and yet most of the quantitative analysis relating health care to outcomes uses linear models. A 2010 review collated more than 100 articles about complex adaptive systems thinking in health care and related sectors (Health-Foundation [2010]) and noted a lack of empirical research.

Having established that health care is a complex system which has non-linear interactions between its constituent parts, the overarching aim of this research was to explore whether using non-linear models can better model health care and so capture the information contained in health care data sets of complex interventions.

7.3 Latent Class Analysis of BeST Back Pain Trial Data

Many studies have used classification techniques to distinguish between categories of patients with low back pain (LBP), and the treatment-based classification schemes have been described in the literature review. Tailoring on the basis of classification in some studies lead to a reduction in pain, whilst others showed no evidence that tailoring treatment on the basis of classification was effective. In the vast majority of cases, the classification was made based on clinicians' observations of increased and decreased pain under varying postural conditions of the patient, and only a few on patient-reported measures. There were a few schemes which used statistical methods such as the k-means algorithm to form clusters based on means of various measures. The k-means method requires that a distance between two measurement of something is meaningfully defined. In the case of a sum score like HADS, it is not clear how meaningful the mean of the score is, or how to take a mean of the categorical version of the variable (e.g., clinically depressed, borderline, normal), so k-means was not considered suitable for this investigation. In contrast, latent class analysis (LCA) is a model-based approach (as described on page 64) which does not require such a distance metric. Two papers were found that described the use of LCA in relation to back pain. In one, the use of LCA for the classification of LBP time courses was found to produce descriptive classes, but these were not linked to treatments or outcomes (Dunn et al. [2006]). The other LCA model linked patient characteristics to quality of life (36-Item Short Form Health Survey, SF-36, Version 1), rather than pain and disability changes (Beales et al. [2012]). These last two studies are the most similar to the BeST LCA study detailed here in that they used some of the same measures, but they were not attempting to associate outcome with a treatment based on these patient characteristics.

The need to find classification schemes which can direct treatment and improve both pain and disability, elucidated in Fairbank et al. [2011], has yet to be fully satisfied. An important contribution has been made in the BeST LCA study presented in this thesis, in using the patient characteristics that the treatment was designed to tackle to derive classes of patients, and relating these classes to pain and disability outcomes. The results starting on page 105 show that it is possible to identify distinct classes of patients using LBP interference with social activities, pain self-efficacy, fear avoidance beliefs relating to physical activity, depression, anxiety and perceived troublesomeness. In a cognitive behavioural approach (CBA) to tackling these characteristics in people with non-specific LBP, there is some evidence that at least a predefined minimum level of reduction in pain and disability is more likely to be

enjoyed by one class than the others.

Additionally, in a model adjusted for class membership, age, gender and employment, an association between outcome and work status was found which was also found independently by a separate secondary analysis of the same data looking for potential effect moderators using the same outcome measures (Underwood et al. [2011]). Underwood's work tested three hypotheses that were defined at the design stage of the trial: that treatment benefit would be greater in those with more troublesome back pain; that the treatment effect would be greater in those with high levels of fear avoidance; that the treatment benefit would be greater in those with subacute low back pain (≤ 3 months) than those with chronic low back pain. For those analyses, the Bonferroni correction was applied and the significance assessed at $\alpha = 0.025$. In addition, they investigated baseline demographic data to identify potential predictors or moderators of outcome, using 2-sided tests with statistical significance assessed at $\alpha = 0.05$. The findings were that univariate analyses showed that age, employment, benefits, and the MVK disability score were predictors of change from baseline of RMQ, MVK disability and MVK pain. It was also discovered that troublesomeness, duration, baseline RMQ, and baseline MVK pain predicted outcome in some but not all three measures. A stepwise selection model was tested and found significant interactions only when using RMQ as the outcome. Specifically, being younger and employed were effect modifiers. Doubt is expressed in Underwood's paper whether this is a true moderation effect.

In the BeST LCA presented here, although treatment predicts recovery in the 4-class model, no association can be found between class membership and recovery or receiving treatment given membership of any of the classes in the logistic regression analyses of the unadjusted model or in the analysis of deviance. In the model adjusted for age, gender and employment, however, there was a marginally significant coefficient for the interaction between membership of Class 2 and treatment with the CBA intervention in the logistic regression models, suggesting that after adjusting for age, gender and work status, membership of Class 2 is associated with recovery for those receiving CBA. In this 4-class model, Class 2 is populated with patients who were moderately affected on all dimensions (details on page 112). The significance of Class 2 was not replicated in the analysis of deviance for the adjusted model, which showed treatment to be highly significant, and the interaction of treatment and work to be significant. Looking at Table 6.9 (page 120), it is clear that both Class 2 and Class 4 have at least twice as many members not working as working, and that Class 1 has 50% more in the working category than not working, so the link between class and work is evident. In Table 6.12 it is clear that of those in work, approximately

half recovered and half did not (bearing in mind that allocation to treatment is 2:1 in favour of the CBA intervention), whilst in the not-working group, the majority did not recover. This is seen particularly in classes 1 and 4, where approximately twice as many did not recover as did recover. Whilst the significance of results relating to Class 2 and Class 4 might be artifacts of low numbers in those classes, the same cannot be said for Class 1 or Class 3 suggesting that the results with respect to work are robust, and are confirming results obtained by alternative methods in Underwood et al. [2011].

In the 3-class model, treatment is a predictor of recovery throughout. Additionally, membership of Class 2 is marginally significant in analysis of deviance in the unadjusted model, indicating the possibility of a different response to treatment for the patients in Class 2. However, Class 2 of the three-class model is also the smallest class with just 52 patients (12.8% of the data set) and therefore any results for this class might be highly dependent on the precise mix of characteristics of the individuals allocated to Class 2, and may, therefore, not replicate in a different data set. Recall that in this data set, those in Class 2 of the 3-class model are the patients with the severest symptoms (details on page 109). The indication that membership of Class 2 affects the likelihood of recovery may be related to the severity of the symptoms individuals are experiencing, or an artefact of the low numbers in the class. The lack of significance for the interaction term indicates that receiving treatment given a patient belongs to a certain class is not a clear predictor of outcome. Fisher's exact test showed an asociation between class membership and recovery for those in the treatment arm of the trial, which was not mirrored for those in the best care only (control) arm of the trial, suggesting a different response to treatment between the two arms of the trial which is class-dependent. In the adjusted model, the interaction between treatment and work is significant in the analysis of deviance, both for the regular and for the refined contrasts. Furthermore, in the refined contrast model, membership of Class 2 was again marginally significant, although the interaction of Class 2 and treatment was not. Once again, in the adjusted model the interaction of treatment and work is highly significant. The same patterns can be seen in Table 6.13 as in Table 6.12, namely that whilst among the people in work, recovery was evenly distributed between recovery and no recovery, the majority of people out of work did not recover, with Class 2 having three times as many who did not recover as did. This could explain why the analysis uncovered work status and Class 2 membership both to be important. Whilst there is some indication that membership of Class 2 affects the likelihood of recovery, caution must be exercised

for reasons stated above.

Whilst the number and characteristics of classes yielding the models with minimum AIC and BIC scores was robust, during the analysis it became clear that the regression coefficients in the linear models relating class membership to outcome were sensitive to changes in a few patient cases, and therefore the conclusions drawn from these analyses may be fragile. It is clear that dividing 407 patient cases into 3 or 4 classes will yield classes which may be small and therefore in danger of having atypical profiles. There is no detail given in the BeST trial report (Lamb et al. [2010b]) as to why there is so much of the follow-up RMQ missing. If the original 701 patients all had a follow-up RMQ score, the results from these analyses would be more reliable.

An important contribution has been made in identifying classes of patients with LBP based on the patient characteristics that the CBA treatment was designed to tackle. These classes have been related to outcomes in adjusted and unadjusted models. There are also some hints that using the patient characteristics that the intervention was designed to tackle has identified classes within the patient population that have differing responses to CBA. A previously identified association between outcome and work status, which was found independently using alternative methods, has been also identified by these analyses. This work has the potential to identify those for whom a cognitive behavioural approach (CBA) is effective or ineffective and so guide the choice of treatments to those most suitable for individuals.

7.4 Artificial Neural Network Analysis of the BeST Data

The use of ANNs for predicting the outcome of treatment for back pain, using baseline measures of the patient characteristics the intervention is designed to tackle, has explored a gap in the literature. For patients offered a cognitive behavioural approach to back pain, this technique would be able to provide a good estimate of whether a 3-point improvement on the RMQ is likely to be achieved for an individual, given their baseline characteristics. If a similar prediction were available for the other back pain treatment options, then these could be combined into a useful tool for allocating a given patient to the most appropriate treatment based on their baseline characteristics. Having established that back pain has a non-linear relationship with some covariates such as frequency of physical activity in univariate analysis, using techniques that can account for non-linearities is a reasonable approach. The

fact that these non-linear models were shown to display improved performance over linear models suggests that linear models do not describe the connection between the cognitive domain and the experience of back pain well. The values of the sensitivity and specificity for the ANN are in line with those produced by Ripley *et al.* (Ripley et al. [2004]).

Given prevalence of LBP, numbers in this trial were small, especially after we require the data to be complete in the explanatory and the response variable, RMQ improvement. It would be useful to be able to replicate the technique on other and especially larger data sets to eliminate the possibility of an atypical result based on a small data set. Perhaps more data would discount the benefits of non-linear modelling we see here, or perhaps confirm it, but much more data would be needed before strong claims can be made. It is also possible that using the measures of the mechanism of treatment are not the best predictors and future work could test the other variables available in this data set and others to see if these have greater predictive power in linear or non-linear models. Nevertheless, in this type of health care application, which uses scores from self-reported questionnaires, it could be argued that the data will always be complex and that the levels of performance we see here are unlikely to be surpassed simply by collecting more data.

We have shown that the artificial neural network provides the best combination of overall error rate and sensitivity, and would be the best candidate of the three models tested for decision support for the cognitive behavioural approach to treatment of lower back pain.

7.5 The Cox Model for Long-Term Survival After a Cardiovascular Event

This analysis of the Basingstoke and Alton data set contributes to the body of research seeking to identify risk factors for long-term mortality in coronary patients. The majority of the trials included in systematic reviews have followed up participants for two years or less, whilst this observational study has followed up for up to 18 years. This length of follow-up is unusually long for studies on patients experiencing a cardiac event or procedure (Heran et al. [2011]). Other studies that did follow up for more than 5 years, limited participants to men under the age of 65 without diabetes or significant comorbidity, whilst this study includes men and women of all ages and all does not exclude diabetes or other comorbidities. Such limitations have compromised the generalisability of the systematic reviews (Turner [2007]). Most studies did not include fitness or anxiety and depression as possible

explanatory variables, whilst this study included both, and also change in fitness during the rehabilitation programme. Differences between cardiovascular mortality and all-cause mortality were not considered in the majority of these studies, and there was little consideration of secondary prevention medications as predictors of survival.

The three long-term studies detailed on page 31 are different from this study in important ways: they had only male participants under 65 years old, only those who had suffered MI, and excluded those with significant comorbidity. In all cases, there was an exercise dose that was fixed, and fitness was not measured, in contrast with the Basingstoke and Alton data set where fitness was measured at entry and exit, and the programme for any individual lasted as long as it took for a given level of exercise capacity to be reached. In addition Dorn *et al.* excluded those failing to complete the initial 6-week low-level exercise programme, further reducing the similarity between their study cohort and patients typically seen in cardiac rehabilitation (Dorn et al. [1999]).

The strengths of this study are the large, unselected cohort with an average 11.5 years of follow-up for those with complete observations. The data were collected in routine clinical practice and collected by one person, Dr Sally Turner, giving confidence in the quality of the data and relevance to NHS practice. The problem of potential of bias in the model due to the missing data has been dealt with using an established method (namely multiple imputation and pooling) and the robustness of the model confirmed. These patients, including a proportion women that is typical for studies in this area, already have coronary heart disease, and there is very little literature for this class of patient. The benefits of a single NHS centre study with consistent staffing and programme is that there is consistent delivery of the programme between patients. Over the course of the study, two different methods of estimating fitness (VO_2max) were used, but only patients whose fitness at the beginning and end are measured in the same way are included. One weakness of the Basingstoke and Alton data set is that physical fitness was assessed indirectly using predicted, not measured, oxygen uptake. However, these method of fitness assessment are typical of the pragmatic methods used in such NHS settings.
Another limitation was the use of the same fitness categories for all patients, and this is likely to underestimate the beneficial effects of fitness for women (Kavanagh et al. [2002], Kavanagh et al. [2003]). Patients who never attended, or attended at least once but did not have a baseline measure of fitness are included in the population of patients typically experiencing a cardiac event or procedure, and excluding

these cases from the analysis diminishes the generalisability of the results. Some of these individuals will have been the most frail patients, and excluding them also explains the initial period of no deaths seen in the Kaplan Meier plot (Figure 6.17). However, our study showed that fitness level in the early weeks after a cardiac event or procedure was equally important for cohorts with a wider range of initial fitness, a wider age range, significant co-morbidities and that included women.

In the survival analysis of the Basingstoke and Alton data set presented in this thesis, fitness category was a significant predictor of all-cause mortality and cardiovascular mortality, whether the baseline measurement was used alone or whether it was used alongside fitness category improvement. This is a much more appropriate variable than simple exercise dose since it is well known that individuals' dose response for exercise varies widely (Scharhag-Rosenberger et al. [2010], Savage et al. [2009]). It has previously been found that improvement in fitness during a CR program was associated with decreased mortality in those with low fitness alone (Martin et al. [2011], Vanhees et al. [1995]). The change in peak VO_2max during the training period of cardiac rehabilitation was found to predict cardiovascular mortality and lower resting heart was linked to cardiovascular survival (Fox et al. [2007], Fox et al. [2008]).

BMI was not available in the a large part of the Basingstoke and Alton data set, so weight was categorised in such a way as to represent normal, overweight and obese categories in an average-height person in a typical UK group with this gender mix. Whilst alone weight category was a significant predictor of mortality, in the presence of other variables, including fitness category, it ceased to be significant. Moreover, when the 889 cases which did have complete data in baseline BMI (and other variables of interest) were modelled, BMI was not retained in the model as a significant predictor of either all-cause or cardiovascular mortality, but fitness was significantly related to mortality. This finding was consistent with other literature, for example, the GOSPEL study, and intensive intervention that decreased CV mortality plus nonfatal myocardial infarction (MI) and stroke, included various lifestyle targets alongside its fixed-dose exercise prescription, including healthy Mediterranean diet and BMI ≤ 25. There were other studies linking BMI with survival, and whilst some considered BMI and exercise the majority did not consider both BMI and fitness (Li et al. [2006], Willett et al. [1995]). The health effects of BMI adjustment (weight loss) were found to be complex, possibly composed of oppositely acting processes, and in need more research (Sorensen et al. [2005]). In a

systematic review of cohort studies, the authors concluded that better outcomes for cardiovascular and total mortality seen in the overweight and mildly obese groups could not be explained by adjustment for confounding factors. They suggested these findings could be explained by the lack of discriminatory power of BMI to differentiate between body fat and lean mass (Romero-Corral et al. [2006]).
The all-cause mortality rate of fit, obese men was found to be not significantly different from that of fit, lean men and that unfit men with low waist girths had greater risk of all-cause mortality than did fit men with high waist girths (Lee et al. [1999]). Others found that BMI, as a predictor of all-cause mortality risk in women, may be misleading unless cardiorespiratory fitness (CRF) is also considered. (Farrell et al. [2002], Kodama et al. [2009]). It is clear that the relationship between BMI, fitness and survival is not well-established.

The Basingstoke and Alton data set study found that fitness and improvement in fitness within the early months after a coronary event or procedure predict long term survival of patients. One fifth of the patients in this study who completed the programme and therefore had an end fitness measurement, moved from mid-fitness to high fitness and these had as good a long-term outcome as those who had begun in the highest fitness category, and a better prognosis than those who began in the mid-fitness category and did not improve. There was also a group who moved from the low fitness to the mid-fitness category and had improved survival over those who did not. The Basingstoke and Alton CR programme differed from other programmes because the exercise training continued for each individual until they were able to perform a physical exercise session without a break, i.e. until they attained a measure of stamina, then their fitness was measured, making no assumptions about dose-response.

Recall that lower VO_2 max estimated from a treadmill test can indicate lower fitness or more severe cardiovascular disease and a higher risk in the AACVPR (see pafe 136. Those patients whose fitness category is low and who do not improve may be those whose cardiovascular disease is more severe. Exercise regimes aimed at improving fitness category should also take account of their disease severity.

The largest age group for both genders in this study was 60-69 years, with 536 (35.1%) of the patients with complete baseline data aged over 65, who would have been excluded from other studies on age grounds , and similarly for the many participants with comorbidities. And yet these patients are typical of those present-

ing for CR after a cardiac event or procedure, and it is valuable to have information on risk factors and especially strategies for modifying risk factors which apply to them.

In the Basingstoke and Alton cohort, secondary prevention medications contribute significantly to improved long-term survival in both all-cause and cardiovascular mortality, as expected for short-term mortality (Liakopoulos et al. [2012],Vale et al. [2011a], Perez et al. [2009]), with the exception of ACE inhibitors which were associated with lower life expectancy. As discussed previously, ACE inhibitors are prescribed to those in high risk categories (Skinner and Cooper [2011]), and this association is probably a reflection of underlying poor health; there was a correlation in the Basingstoke and Alton data set between a prescription for ACE inhibitors and high risk category. There is no information on how long the medications were used, suggesting that the effect is very strong.

In the Basingstoke and Alton data set, diagnosis is a significant predictor of cardiovascular mortality and of all-cause mortality when fitness at baseline only is used as an explanatory variable. When fitness plus change in fitness category is used, diagnosis is not a significant predictor of mortality suggesting that change in fitness category is correlated with diagnosis. In the other models, patients having coronary artery bypass surgery and percutaneous coronary intervention have a significantly higher long-term survival from cardiovascular mortality than do patients with a myocardial infarction or angina. This accords with Buckley *et al.* where prognosis to death or cardiac outcomes for patients with angina alone was similar to those with previous acute MI or revascularisation, while health status was poorer (Buckley and Murphy [2009]).

Smoking is known to be a factor in the development of heart disease, so it might be expected to be a predictor of mortality in those who have had a cardiac event or procedure (Ketonen et al. [2008], Hardoon et al. [2008]). In the Basingstoke and Alton data set, current smoking habit data showed nearly half the women had never smoked compared to just under one third of the men and close to a third of both genders had recently given up smoking. Whilst current smoking habit was a predictor of mortality when used alone, it ceased to be a significant part of the model in the presence of other variables. This may be because current smoking habit is a 'snapshot' at a point in time and does not give all the information about long-term behaviour. Initial cessation of smoking is quite common in CHD patients. However,

approximately 50% of those who stop smoking begin smoking again within 1 year after the event (Barth et al. [2004]). Another possibility is that smoking is correlated with fitness, and having controlled for fitness, smoking is no longer an independent predictor of mortality. Recent epidemiologic studies show evidence for a link between smoking behaviour and depression (Schmitz et al. [2003]).

There was little evidence of clinical depression in the Basingstoke and Alton cohort, and most of those whose HADS scores suggested either depressed or a borderline category improved by graduation. Depression and anxiety were each predictors of long-term mortality in univariate models, but neither was significant in the multivariate model. Depression is known to be associated with short-term mortality. Dickens *et al.* investigated the association of depression with mortality following MI in 588 subjects and followed up their cases for up to 8 years. Depression was not associated with cardiac mortality, whether depression was detected immediately before MI, 12 months after MI or at both time points. They concluded that the association between depression and post-MI mortality is complex, possibly being limited to depression immediately after MI (Dickens et al. [2007]). A recent meta-analysis found that the presence of depressive symptoms after myocardial infarction was not uncommon and was associated with a 2-fold to 2.5-fold increased risk of impaired cardiovascular outcome within 2 years (van Melle et al. [2004]). They concluded that the potential mechanisms linking depression and impaired cardiovascular prognosis were still poorly understood. First, unhealthy behavior of depressed MI patients (diminished compliance, smoking, unhealthy diet, inactivity) was found to be important. Second, evidence was found to be growing that physiological mechanisms were involved. Finally, physicians also prescribed significantly less thrombolysis, aspirin, ACE inhibitors, and BB to MI patients with comorbid depression. Therefore, deficits in quality of medical care may explain, in part, the excess mortality experienced by patients with depression after MI. Another meta-analysis found results on the long-term impact of clinical depression to be rather limited and further research to be necessary to obtain more stable effects on the impact of depressive disorders on mortality in CHD patients (Barth et al. [2004]). The ENRICHD RCT sought to investigate the effect of treating depression and low perceived social support on mortality and recurrent MI after a first MI (Berkman et al. [2003]). 1,084 women and 1,397 men were randomised between usual care and individual and group cognitive behavioural therapy during the first 6 months after MI. The intervention improved depression and social isolation, but did not increase event-free survival at mean 29 months follow-up. The relative improvement in the psychosocial intervention was

less than expected because of substantial improvement in the usual care group. It is known that the natural history of depression following acute coronary syndrome (ACS) is variable (Haas [2006]). It is not clear, therefore, whether the improvement in depression and anxiety seen in the Basingstoke and Alton cohort is part of the natural history of the illness or due to treatment, exercise or the support of the group at the exercise and education sessions. Nevertheless the improvement is welcome, and with much longer follow-up, confirms that depression is not associated with long-term mortality. Another study found that depression predicts failure to complete cardiac rehabilitation (Casey et al. [2008]), and this was also evident in the Basingstoke and Alton cohort (Turner [2007])

With 13.7% of patients female, the Basingstoke and Alton cohort was fairly typical of the patients seen in CR. As discussed above, the majority of trials have included only men, leading some researchers to conclude that coronary heart disease in women is under diagnosed, under treated, and under-researched (Mikhail [2005]). Women account for less than 30% of the participants in most studies and trials in cardiology making it difficult to draw conclusive evidence on managing cardiovascular disease in women. Despite differences between the sexes in risk factors, presentation, and response to treatment, women continue to receive similar treatments to men on the basis of trials that include mainly male participants (Mikhail [2005]). One review article concluded that the pathogenesis of CHD was very similar for men and women. Yet, diabetes, HDL and triglycerides levels were found to have a greater impact on CHD risk in women compared to men. In addition, there were indications that risk factors such as smoking, family history and inflammation characterized as C-reactive protein, had a more negative influence on CHD in women than in men. On the other hand the evidence showing that lipoprotein(a) was a cardiovascular risk factor seems to be stronger in men than in women. The majority of cardiovascular risk factors showed no important differences between the genders (Roeters van Lennep et al. [2002]). The EUROASPIRE III survey (in 22 European countries in 2006-2007) identified 8,966 consecutive patients (25.3% women) who had had a coronary event or revascularisation before the age of 80. The results showed that despite similarities in medication exposure, women were less likely than men to achieve BP, LDL-cholesterol and HbA1c targets after a coronary event. This gap did not appear to narrow between 1994 and 2007 (Dallongevillle et al. [2010]). In an investigation into the effect of doctor and patioent gender, Adams *et al.* found doctors appeared less affected by patient gender but both male and especially female doctors took more account of male patients' age, and considered more age-related

disease possibilities for men than women. Findings highlighted the need for better integration of knowledge about female presentations within accepted CHD risk models, and did not support the contention that women receive better-quality care from female doctors (Adams et al. [2008]).

A postcode-based index of multiple deprivation was recorded for the Basingstoke and Alton cohort. Health may be expected to be generally better in an affluent area such as this than in a deprived geographical area, so these results may not be fully generalisable to other regions. A study analysing recent socioeconomic trends in coronary heart disease mortality in England found both death rates and the number of deaths were lowest in the most affluent quintile and the pace of fall was also faster than in the most deprived quintile. Overall, about half of the decrease in death rates was attributable to improvements in uptake of medical and surgical treatments. The contribution of these to the deaths averted was very similar across all quintiles. Risk factor changes accounted for approximately a third fewer deaths in 2007 than occurred in 2000, but were responsible for a smaller proportion of deaths prevented in the most affluent quintile compared with the most deprived. However, the benefits of improvements in blood pressure, cholesterol, smoking, and physical activity were partly negated by rises in body mass index and diabetes, particularly in more deprived quintiles (Bajekal et al. [2012]). Occupation was recorded for the Basingstoke and Alton cohort, but was not a significant predictor of mortality in the multivariate model. In a prospective occupational cohort study of 17,186 male civil servants aged 40-69 years between 1967 and 1970 in the UK (the Whitehall study), employment grade was used as a proxy for socioeconomic position. Men with low employment grades generally had less favourable risk profiles than those with high grades. Systolic blood pressure, glucose, and proportions of current smokers and non-insulin dependent diabetes patients were higher in the low employment grade group than in the high grade group. However, total cholesterol concentrations were slightly greater in the high grade group than in the low grade group. During the 15-year follow-up, 1,262 men died from coronary heart disease and low employment grade and all risk factors were associated with high mortality (Kivimäki et al. [2008]).

Promotion of fitness after a coronary event or procedure may extend life expectancy. Monitoring of patients at risk of a coronary event and intervention where appropriate to prevent myocardial infarction may improve life expectancy.

7.6 Artificial Neural Network for Modelling Long-Term Survival After a Cardiovascular Event

The ANN and Cox models both produced survival probabilities for mortality in cardiovascular patients consistent to with the Kaplan-Meier estimate and with each other (page 160), in line with the findings of Joshi and Reeves (Joshi et al. [2005]) and Ripley *et al.* (Ripley et al. [2004]) in their studies on cancer. The use of Kaplan-Meier curves for comparison is standard practice and there are recent proposal to correct for case-mix where this is appropriate (MacKenzie et al. [2012]).

Whilst Ripley gave exemplar hazard functions for each of the 7 models studied, neither Ripley nor Joshi showed or discussed the hazard rates produced by their respective ANN models in relation to their data. In the new work presented here, using the models' ability to predict a fixed length of time has been avoided as a basis of comparison. Cox models are not designed to estimate the probability of survival at a fixed time, they are intended to show the dependence of the survivor curve on the explanatory features (Ripley and Ripley [2001]). When used for prediction they are able to predict the whole survivor curves and it is not surprising that they are less well able to predict a single point on that curve than the methods designed to predict just one point. Furthermore, censoring biases will always favour the ANN as it estimates the probability of survival to a fixed time conditional on the patient still being under follow-up, and not the unconditional probability estimated by survival analysis or by a clinician. Ripley *et al.* use the deviance as a measure of fit (the sum of minus twice the logarithms of the predicted probability of the event over all the patients in the training set) which provides a more sensitive measure of fit than the success rate, especially in the survival analysis models where the exact time of death is used.

The comparison between the Cox proportional hazards and the ANN hazards is given on pages 162 to 173. The Cox model is constrained by its proportional hazards assumption, but the ANN is free to vary over time and between levels of the covariates, allowing the development of hypotheses which can be used in further investigations. The contrast between male and female hazards in the ANN model is negligible, when other variables are held constant, in all but the two groups with the oldest age and lowest fitness. Here, there is a small increase in hazard for the males, but the hazards are proportional, as with the Cox model. In addition, the hazard increases more quickly over time for these elderly, unfit patients. This suggests that the effect of gender itself on the hazard is proportional but modified by age or by fitness or both. Statins appears to have little or no effect on the hazard. There is a

slight upturn at late times for the patients without a Statins prescription, but this is a slight difference. Consistently with the hazard split by age, the shape of the hazard is different for the older, less fit patients, increasing more sharply over time, but in this case the the curves lie on top of each other, suggesting that hazard is not different for those with and without a prescription for Statins. The same cannot be said for those with and without an ACE inhibitor prescription, except in the case of the older, less fit patients. In the younger patients, the hazard is increased markedly for those having an ACE inhibitor prescription, and increases more quickly at later times. The exception is the patients also having a prescription for Statins; compared to those who do not, and with all other variables held constant, the hazard is the same for those with and without ACE inhibitor prescriptions, except at very late times when the hazard for those with a prescription have a rapidly increasing hazard. This suggests that the effect of a prescription for ACE inhibitors on hazard is modified by a prescription for Statins, including for older patients. It may also be modified by age or or both. Aspirin also shows a clear difference in the hazard as modelled by the ANN, between those who have and those who do not have a prescription. In this case, having a prescription lowers the risk, and for the slightly younger, fitter patients, the risk remains largely unchanged over time. For those without a prescription for Aspirin, the hazard profile is very similar to that of the oldest, least fit patients, for whom Aspirin has little or no modifying effect on the hazard. Having a prescription for Statins does not alter the hazard for having or not having a prescription for Aspirin, suggesting that Statins and Aspirin act independently. The effect of Aspirin on the hazard may be modified by age or fitness or both. When the hazard for each of the 4 age categories is investigated, the ANN models shows those who are over 70 years of age at baseline to have a higher hazard than the other 3 categories, which are approximately the same, regardless of Statins. Fitness does not seem to have an effect on hazard, except for the least fit categories where at late times, those age 60-69 years at baseline have a sharp increase in hazard. This suggests that there is something akin to a step change in hazard for those who are least fit when they reach around 70 years of age. Finally, fitness seems to have a proportional relationship with hazard in the ANN model, with the differences in hazard being greatest for the over 70s, although by no means large. This suggests there is an interaction between age and fitness.

It seems to be an open question whether the effects of fitness differ measurably for the over 70s, although the ability maintain fitness gains in the over 70s does seem limited (Oerkild et al. [2012]). According to this ANN model, the hazard for the older and least fit patients is qualitatively different than than of the other patients,

and either age or fitness or both seem to affect the effect of medications and gender. Moreover, it seems that age and fitness affect each other. This warrants additional investigation if interventions suitable for patients of all demographic groups are to be developed.

The hazards under the ANN model, when all other covariate are kept constant, are proportional for gender, statins and fitness, but not for age, ACE inhibitors or aspirin. This suggests that using a model able to capture non-linear relationships has revealed nuances that the linear model cannot. These hypotheses have been drawn using subsets of the data where there is a reasonable number of patients. It happens that all such subsets have male patients, since this cohort was overwhelmingly male. This contrasts with the previous work; Joshi & Reeves had 1,160 females to 786 (41%) males in malignant melanoma and Ripley all female (breast cancer). Furthermore, Joshi used different cut-points for the male and female patient age categories, determined by on Martingale residuals of the Cox model.

An ANN approach has successfully been extended to a non-cancer data set and hazard rates produced by the model have been examined and hypotheses generated.

7.7 Concordance with Research Aims and Objectives

The research aims were set out on page 13, and in this section each objective has been revisited and the extent to which it has been achieved has been evaluated. Discussion on how the results would need to be extended to the point where policy would be informed has also been described.

7.7.1 Can latent class analysis identify sub-sets of patients within a cohort of patients recruited to a clinical trial for non-specific low back pain?

Using a probabilistic model, classes of patients have been identified within the BeST cohort. From a number of competing models, it was possible to identify two models with substantially better fit than the others. The models have proved to be robust against minor re-specification of cases. This contributes to the existing work seeking to understand the different responses of patients to the same treatment by dividing

them into groups. This is a first step in the attempt to take a population-level result and apply it to groups of individuals sharing relevant characteristics.

7.7.2 Is it possible to use linear models and the classes identified by the latent class analysis to tailor interventions for non-specific low back pain to patients?

Having identified sub-populations (classes) within the LBP population in the BeST trial, the next question was 'did the patients in different classes respond differently to the CBA treatment?'. This was tackled by using logistic regression to predict outcomes using class membership and treatment allocation as predictors. A second logistic regression was used to predict outcomes using class membership, age, sex and work status, since other work suggested these might be important. It proved difficult to establish with any certainty that the classes of patients responded differently to the CBA complex intervention. Whilst an association between class membership and outcome for those who received the complex intervention was evident, it was not possible to discern whether the association for any given class was positive or negative (whether a particular class was more likely or less likely to achieve the specified improvement). It was not possible, either, to identify which specific class or classes responded differently to the complex intervention in terms of outcome. This is not unusual in complex systems. As already discussed to the introduction, complex systems often exhibit non-linear behaviour and sensitive dependence on initial conditions making them unpredictable over time.

7.7.3 How do the performances of the linear and non-linear models compare for the prediction of patient outcomes following a cognitive behavioural approach intervention for non-specific low back pain?

The linear logistic regression based on class membership alone was compared to a logistic regression using the same variables as the LCA used to model classes and these were both compared to an ANN which also used the same variables for prediction. Over all, the non-linear ANN performed better in terms of accuracy of prediction, sensitivity and log score than either of the linear models. These models could each be used to inform a clinician's advice to a patient about the likely efficacy for them of the BeST CBA intervention. In such an application, the ANN is likely to be more useful, but ANNs lack interpretability. If the clinician wished, or was required, to justify their advice, the lack of interpretability would make this difficult.

7.7.4 What is the potential for these linear and non-linear models to inform policy on the treatment of non-specific low back pain?

The BeST trial showed that the CBA intervention was effective on average, and the LCA work has attempted to answer the question 'for whom is it effective?'. If a model is to be used to to inform policy or aid clinical decision making, it needs to be accurate and easy to understand. As they stand at present, the accuracy of the models introduced here is around 60%. Whether this is good enough depends on the alternatives and the cost of a misclassification. In the narrow application of allocating patients to CBA treatment for their back pain, it could be argued that the alternative is to allocate all patients to CBA treatment, since it was effective on average. In that case, any of these models might provide an improvement, although the non-linear model had the highest accuracy of the three. In an ideal scenario, there would be a single tool which identified the particular treatment, out of the array of options available, which would be optimal for the patient at hand. The STarT tool (Hill et al. [2011]) has made a step in this direction by identifying broad types of treatment best suited to patients who fall into 3 categories for risk of chronicity. STarT is available on the web and around 75 clinicians, in a variety of countries and settings, have indicated that they are using the tool with a variety of treatment approaches (STarT [accessed 18/05/13]).

Accuracy of the models could be improved with more data, particularly the question about the associations between class membership and outcome. One approach would be to combine data from different data sets. There is a project underway at the University of Warwick to collect anonymised data of therapist-delivered intervention for low back pain, called Low Back Pain Repository project. There is precedent for this kind of approach: researchers combined data from 2 large trials (making 40,000 cases) and were able to define 28 sub-groups and contribute valuable insight on benefits and risks of aspirin after ischaemic stroke for small subgroups of patients for whom the individual trials did not contain enough cases to arrive at firm conclusions (Chen et al. [2000]). The difficulties of different studies defining different outcome measures is being addressed by the Core Outcome Measures in Effectiveness Trials (COMET) initiative launched in 2010. COMET facilitates the development of agreed standardised minimum sets of outcomes to be measured and reported in all clinical trials, audits of practice or other forms of research for a specific condition. Outcomes in a particular study are not restricted to those in the core outcome set, but there is an expectation that the core outcomes will also be

collected and reported to allow the results of trials and other studies to be compared, contrasted and combined as appropriate (Williamson et al. [2012]).

7.7.5 Using a linear model, what predicts long-term survival after a cardiac event or procedure, at population level?

Using a Cox model, a linear model for survival, it was discovered that fitness category, secondary prevention medications (statin, aspirin and ACE inhibitor), diagnosis, age and sex were associated with cardiovascular mortality after a cardiac event or procedure at population level. All cause mortality was associated with these same variables, and in addition, co-morbidity and systolic blood pressure. It was also found that after adjusting for fitness, BMI was not associated with either cardiovascular or all-cause mortality after a cardiac event or procedure. Since the Cox model is a linear model, its parameters have a straightforward interpretation as hazard ratios, which can be used to predict the likely survival experience of an individual one the value of their variables is known.

7.7.6 Can a non-linear artificial neural network be used to model long-term survival after a cardiac event or procedure?

The particular form of ANN selected for use in this study had previously been used on a Cancer data set, and was selected for ease of comparison. As with the cancer data set, this ANN was able to model survival. and the concordance of the survival probability distribution produced with the Kaplan-Meier estimate was as good as that of the Cox model

7.7.7 How do linear and non-linear models compare in modelling the survival and hazard of a population who have experienced a a cardiac event or procedure?

The concordance of the survival probability distribution produced by the ANN with the Kaplan-Meier estimate was as good as that of the Cox model,
When the Cox and ANN survival models produced hazard rates for subsets of individuals who all shared the same values of explanatory variables, the Cox model was constrained by its proportional hazards assumption, but the ANN had no such constraint. Where the hazards modelled were those for groups of patients whose explanatory variables differed only in one variable, hypotheses about the effects of different explanatory variables and interactions between them could be formed, as

described above. This is the first time an ANN survival model has been used in this way.

7.7.8 What is the potential for this research to inform policy for those who have experienced a a cardiac event or procedure?

Cardiovascular disease is a well-studied disease. Nevertheless, its context is constantly changing, with new forms of clinical management and definitions, such as the redefinition of myocardial infarction in the year 2000 and in 2012 (Thygesen et al. [2012]). The hypotheses generated using the ANN could be tested, given sufficient cases, and the remarks made above about the pooling of data hold for this case, too. There is a cardiac surgery register (Bridgewater et al. [2010]), but its focus is quality and short term survival and not all the patients in cardiac rehabilitation will have had surgery.
Hip to waist ratio and BMI are not independent of fitness and whenever the prognostic power of the former is suspected, cardiorespiratory fitness should always be measured to discriminate between these indicators. Caution should be exercised in attributing causation to BMI and advice to adjust BMI to reduce reduce mortality in cardiovascular patients is not supported by this research. However, the promotion of fitness to reduce mortality in Cardiovascular patients is well supported by this research.

7.8 Future research

The research presented here has some limitations which provide opportunity for further research. Whilst the use of non-linear models did provide additional insight, their efficacy in tailoring interventions to patients has not yet reached the point of informing policy or useful application in clinical practice.

7.8.1 BeST back pain trial

The BeST study, with 701 participants, is described as a large-scale randomised controlled trial (Lamb et al. [2010a]). Whilst this study size was appropriate for the demonstration of long-term effectiveness and cost effectiveness of CBA intervention at population level, it has (especially with only 407 with complete observations) constrained the usefulness of the division of the patient population into classes using LCA. In this respect, the ANN was a more successful non-linear model, since its superior ability to predict recovery makes it a candidate for use in aiding clinical

decisions about the treatment most likely to be effective for a given individual. However, with an accuracy of 61% its usefulness is limited. Larger data sets would enable the assessment of these non-linear models for use clinically. Larger data sets might include similar sized data sets with fewer missing values, or data sets with similar levels of missingness and more participants. As has been discussed previously, it can be hard to establish with certainty that missing data are missing at random, and biasses can be introduced in data sets with missing observations.

LCA and ANNs are more frequently used in very large data sets (Mitchell [1999]) where the use of standard statistical analyses becomes problematic: small effects can be found with very large sample sizes and it is important to recognize that a statistically significant finding may not be meaningful or useful. Variance diminishes as the sample size increases, but bias stays constant and when the bias is large, a small truly random sample or randomized study can be more valuable. One very large database of observational clinical data is the Clinical Practice Research Datalink (CPRD, formerly the General Practice Research Database, GPRD) which makes available anonymised data from the GP Practices which are registered with it. There is also an NHS Data linkage service at the Health and Social care Information Centre which is able to link CPRD data and Hospital Episode Statistics (HES) data in a suitably anonymised form, or to link study-collected data or disease register data with CPRD and HES data. HES is a data warehouse containing details of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England, collected during a patient's time at hospital and is submitted to allow hospitals to be paid for the care they deliver. The use of general medical or administrative data sources must be approached with caution, however, since its primary purpose is not research but administration and the needs for accuracy may differ in the details between these two purposes. For this reason, any study using these kinds of data needs to include a practitioner who can give appropriate insight into the likely accuracy of the variables, data quality control and reasons for missing data. The UK's national health service puts it in a unique position with respect to the possibility of unselected medical data, but it should be remembered that CPRD and HES data is observational, and the usual cautions with respect to bias in observational studies apply (Hammer et al. [2009]).

The increasing adoption of core outcome measures in effectiveness trials (COMET) is making it easier to combine the results of several trials in a specific disease group, since the agreed measures are reported alongside the researchers' other

outcomes. The repository above can take advantage of the fact that standard outcome measures have been proposed for low back pain research (Dayo et al. [1988]), specifically: bothersomeness or severity and frequency of LBP; RMQ or Oswestry Disability questionnaire (Roland and Fairbank [2000], Maughan and Lewis [2010]); SF-12 or EuroQoL; days of work absenteeism, cut down activities and bed rest; and a single question on overall satisfaction. However, the ability to undertake sub-group analysis on such data may not be straightforward, and the cautions issued for sub-group analysis in Cochrane review meta-analyses should be heeded (Higgins and Green [2011]). There may not be sufficient detail given in differing studies to be certain that the patients truly belong in the same sub-group, which may bias the analysis. It is important to consider whether the differences between sub-groups are clinically plausible, are supported by other evidence, or support different recommendations.

The Low Back Pain Repository project to collect anonymised data of therapist-delivered intervention for low back pain, currently has more than 9,000 cases in the repository and the aim is to reach 10,000. One challenge here is that, despite standard outcome measures having been proposed for low back pain research, even when the same or similar patient characteristics and outcomes were measured, the instruments for measuring them are not all the same, and very few measure troublesomeness. The repository divides the measures into domains, and some work to understand how these individual measures relate to each other will be needed. It will be interesting to discover how troublesomeness is related to the other measures. If it proves possible to model a proxy to troublesomeness, it will be possible validate these result on trials with troublesomeness recorded and also on these proxy outcomes. In addition not all interventions have been the same, so uncovering the similarity and differences of patient responses to disparate interventions may be challenging. Another challenge is the heterogeneity in the follow-up times and how these might be robustly rescaled to match. Since the data is anonymised, there is no way to know whether the same individual is represented multiple times in different trials, and to date the age distribution has not matched that in the population at large. Nevertheless, there is significant opportunity for validating a classification approach in such a large database.

The burden of data collection on both study participants and researchers can have an effect of the rates of missingness, so it may be better to collect those variables known to be informative and make effort to avoid missing data, than to try to

collect everything and risk much missingness. Exploration of the repository or other data sets present an opportunity to reveal what the minimal effective data collection would be to inform future research.

Research into a classification approach to tailoring interventions to patients with low back pain using LCA would be aided by a more extensive data set. Methodological advances which allow multiple imputation and robust pooling for non-linear models would like LCA and ANN would help to make best use of the relatively small data sets which are able to be collected during back pain intervention trials.

7.8.2 Basingstoke & Alton cardiac rehabilitation

Currently there is not a core outcome set in place for cardiac arrest clinical trials but there is a COMET initiative underway to rank the importance of potential outcome measures with the use of a Delphi process for all key stakeholder groups and to host a consensus meeting involving all key stakeholders to finalise a core outcome set (Collaboration [Accessed 25/05/2013]). Another source of data are disease registers. There is an adult cardiac surgery register (Bridgewater et al. [2010]), but not all patients who have a cardiac event or procedure have surgery. The purpose of the cardiac register is to measure the quality of care of adult cardiac surgery in the UK and provide information for quality improvement and research, so it may nor be ideal for investigation into long-term survival for a more general population of patients who have had a cardiac event or procedure.

For the ANN survival analysis, Bayesian methods for learning could be applied as they have been in other studies. Defining a suitable prior may reveal more detail, especially since a Bayesian approach might help overcome some of the limitations of a small data set (Eleuteri et al. [2007b], Taktak et al. [2008]). This ANN analysis has relied on identifying prognostic factors using the Cox model as in Joshi et al. [2005], but variable selection can be done using the ANN (Bourdès et al. [2007]), and in Biglarian et al. [2010] ANNs were compared to a Weibull model for determining prognostic factors in gastric cancer patients. The normalised importance was the measure of significance for the ANN and the p-value for the Weibull model. The markers were broadly the same, but order of significance was different. The performance as measured by correct prediction was superior in the ANN, but both models were good ($\geq$ 74% for death, survival and total). In the work of Eleuteri & Taktak, automatic relevance determination was used, a Bayesian method described in Bishop [1996] capable of determining the relative importance of input variables,

and allowing the inclusion of a large number of inputs without fear of overfitting (MacKay [1994]). It is also possible to assess the saliency of each weight by setting to zero and examining the change in the cost function. A number of feature-salience ranking techniques are discussed in Wang et al. [2000].

Development of good methods for interpreting ANNs could make them far more useful: a common problem is that users, and particularly clinicians, are quite reasonably reluctant to trust important decisions to systems they do not understand (Plate et al. [2000]). Illustrating with a data set on squamous-cell lung cancer, Plate *et al.* showed that plotting the effect on the output of a particular input variable for some selection of points in the input space can reveal which variables contribute to the output and even indicate interactions. There have been attempts to interpret the output of an ANN using rule extraction from the trained network, and there is a survey of these methods in Saad and WunschII [2007]. In addition, the ROC framework has been extended to evaluate performance of time-to-event data and a framework developed to represent the operation of ANNs as low-order Boolean rules that can be checked against domain expert knowledge (Lisboa et al. [2007])
It has recently been shown that another form of machine learning, support vector machines (SVM), can be adapted to perform survival analysis (Lama et al. [2011]). The basic SVM is a classification machine, but relevance vector machines (RVM) is an extension with a Bayesian learning algorithm providing posterior probabilistic outputs. Using a partial logistic framework, Lama *et al.* used SVMs to analyse data from a head and neck cancer trial.

Missing data is often a challenge in real-world applications. In an ANN-based decision support system, missing values in a key variable were substituted using the mean value, a random value, nearest neighbour or a neural network estimate and were compared to omitting the variable in training the decision support ANN (Pesonen et al. [1998]). The imputation methods gave broadly similar accuracy and omitting the variable from the input of the decision support ANN also gave acceptable results. Machine learning, including ANNs, can be used as techniques for providing imputed values before analysis (Jerez et al. [2010]).

Research into long-term survival following a cardiac event or procedure could be extended by using larger data sets and by extending the ANN approach into a Bayesian framework. Visualisation and interpretability of outputs is vital in order to make non-linear models acceptable for use in health care settings, especially if

used in decision support for treatment choices.

7.9 Summary

In the case of patients with non-specific low back pain, a cognitive behavioural approach was previously found to be effective on average and cost effective. The work in this thesis has shown that it is possible to use the patient characteristics that the intervention was designed to tackle to divide the patient cohort into classes. There is some evidence that there is a differential response to the intervention, although the evidence here is not robust. However, the ability of these methods to pick out the same effect modifiers as independent work on the same data set gives some evidence that the methods are finding true effects. One avenue of confirmatory work could be the use of the back pain repository to test out the hypothesis that classes of patients can be identified such that class membership is associated with differing outcomes following a cognitive behavioural approach. The same methodology could be used to evaluate the division of patients into classes and investigate association with outcome for other interventions.

Artificial neural networks have been applied to predicting outcomes based on the same characteristics that the cognitive behavioural intervention was designed to tackle. Compared to a logistic regression on these variables, and compared to a logistic regression using class membership based on these variables, predictions from the ANN were superior. The same methodology could be used to evaluate the likely response of patients to other interventions, and move towards a tailoring of interventions for patients that may increase the rate of recovery from first-treatments by distinguishing which of a range of treatments is most likely to be effective for a given individual.

In the case of patients recovering from a cardiac event or procedure, the standard survival analysis has shown that fitness is an invaluable prognostic indicator for both cardiovascular and all-cause mortality. Furthermore, when fitness is measured as VO_2max, then BMI is no longer statistically significant as a prognostic factor. The fact that BMI ceased to be a significant predictor of prognosis when fitness was measured in cardiac rehabilitation has important implications for other areas of medicine where BMI is a prognostic factor. It is important that proper distinction is made between correlated variables like BMI and fitness to ensure that patients are given correct advice. If prognostic value ascribed to BMI is actually the effect of fitness, but this was not measured, it may go some way to explaining the obesity paradoxes

being seen in (Ryan [2005], von Haehling et al. [2011]), as people attempting to lose weight typically increase their activity leading to an improvement in fitness, independent of any weight loss. In future work, whenever BMI and hip to waist ratio are measured, fitness should also be measured so that the correct factors can be identified and appropriate health advice can be given and optimal interventions may be designed.
An ANN has been successfully used to model survival in this cardiovascular rehabilitation cohort. The predicted survival by this method was a good fit to the Kaplan-Meier estimate, and more flexible than the Cox model fit. The hazards over time for individual covariates have been explored (as in Lisboa et al. [2009]) giving rise to testable hypotheses. These results suggest that the effect on the hazard of ACE inhibitors is modified by Statins and that the response of unfit patients over the age of 70 years is different from their fitter counterparts in the age range 60-69 years. Extensions to this work would ideally include a larger data set where each of the sets of variable combinations was adequately represented, particularly women and the age groups under 60 years about which the ANN approach can say little. In addition more detail may allow the testing of the hypotheses about interactions between variables. A Bayesian approach could be taken and variable selection could be performed within the ANN framework instead of adopting the variables selected in the optimal Cox model. Other machine learning techniques, such as radial basis function networks could be explored.

In the following chapter, conclusions are drawn on the value of using non-linear models to explore complex health care data sets.

Chapter 8

Conclusion

It is well-established that health care and health itself are complex systems and they are therefore expected to behave in a non-linear manner. Many of the conventional models used for health care data are linear. The aim of this research was to explore whether using non-linear models could better model health care and so capture the information contained in health care data sets. In order to achieve this goal, this research compared linear models with non-linear models for two health care data sets of complex interventions. This research has attempted to contribute to an increase in understanding of the complex interaction within people, within complex interventions and between people and the interventions, particularly in pursuit of decision support in tailoring interventions to patients and generation of hypotheses.

Latent class analysis was able to identify classes of patients with similar characteristics to each other and these classes were related to the patients response to a complex intervention for low back pain. It was not possible to identify with certainty the specific responses (positive or negative) of particular classes, which limited the use of the classification for tailoring interventions to patients. Tailoring decisions were best supported by the use of an artificial neural network, since its ability to predict likely outcome for an individual was superior to the other models investigated. The ability of these models to inform policy on the allocation of treatments for low back pain was limited by the modest accuracy and missing data. Nevertheless, it was demonstrated that this approach has potential, and that the non-linear models offered more accurate predictions than the linear model.

After a cardiac event or procedure cardiovascular mortality in the long-term is predicted by age, sex, fitness, secondary prevention medications and specific diagnosis of cardiac event or procedure. All-cause survival in this population is predicted

by these same variable with the addition of co-morbidity and systolic blood pressure. There was a different survival time for those who completed and did not complete the rehabilitation programme. For those who did complete the rehabilitation, fitness gained during the programme predicted increased survival times, and fitness was a more significant predictor of survival than BMI. using the same explanatory variables, the non-linear artificial neural network model was able to predict the survival probability as well as the linear model. The non-linear model of hazard rate gave rise to a number of hypotheses about interactions between covariates which may affect, which were nor obtainable from the linear model. These models are not yet sufficiently robust to inform policy, but if visualization of the model outputs was developed and the hypotheses generated were tested, then these models could be used. The non-linear model offered new hypotheses that are pertinent to tailoring interventions to patients using phenotype, rather than genotype.

The conclusion drawn is that non-linear models are useful in the analysis of complex health care data sets because of their ability to capture some of the complexity of the data that cannot be captured by linear models.

Bibliography

Ann Adams, Christopher D. Buckingham, Antje Lindenmeyer, John B. McKinlay, Carol Link, Lisa Marceau, and Sara Arber. The influence of patient and doctor gender on diagnosing coronary heart disease. *Sociology of Health & Illness*, 30 (1):1–18, 2008. ISSN 1467-9566. doi: 10.1111/j.1467-9566.2007.01025.x. URL http://dx.doi.org/10.1111/j.1467-9566.2007.01025.x.

Simon J Adamson, J Martin Bland, Elaine M Hay, Ruth E Johnson, Gareth T Jones, Henry Kitchener, Jennifer A Klaber Moffett, Gary J Macfarlane, Hugh MacPherson, Sionnadh McLean, Linsey Nelson, Chris Salisbury, Elaine Thomas, Helen E Tilbrook, and David J Torgerson. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. *BMJ*, 337, 2008. doi: http://dx.doi.org/10.1136/bmj.a1864.

H. Akaike. A new look at the statistical model identification. In *IEEE Trans Automat*, 1974.

L Akhmadeeva, B Veytsman, G Rayanova, and N Setchenkova. Dynamic tens in management of low back pain: cluster analysis for efficiency prediction (poster abstract). *European Journal of Neurology*, page 584, 2010.

Steven Allender, Peter Scarborough, Viv Peto, Mike Rayner, Jose Leal, Ramon Luengo-Fernandez, and Alastair Gray. European cardiovascular disease statistics. *European Heart Network*, 2008.

Federico Ambrogi, Nicola Lama, Patrizia Boracchi, and Elia Biganzoli. Selection of artificial neural network models for survival analysis with genetic algorithms. *Computational Statistics & Data Analysis*, 52(1): 30 – 42, 2007. ISSN 0167-9473. doi: 10.1016/j.csda.2007.05.001. URL http://www.sciencedirect.com/science/article/pii/S016794730700196X.

Ruth A. Anderson, L. Michele Issel, and Reuben R. Jr McDaniel. Nursing homes

as complex adaptive systems: Relationship between management practice and resident outcomes. *Nurs Res.*, 52:12–21, 2003.

Elliott Antman, Jean-Pierre Bassand, Werner Klein, Magnus Ohman, Jose Luis Lopez Sendon, Lars Rydén, Maarten Simoons, and Michal Tendera. Myocardial infarction redefined-a consensus document of the joint european society of cardiology/american college of cardiology committee for the redefinition of myocardial infarction: The joint european society of cardiology/ american college of cardiology committee. *J Am Coll Cardiol*, 36:959–969, 2000. doi: 10.1016/S0735-1097(00)00804-4.

Adri T. Apeldoorn, Raymond W. Ostelo, Hans van Helvoirt, Dirk L. Fritz, Julie M.and Knol, Maurits W. van Tulder, and Henrica CW de Vet. A randomized controlled trial on the effectiveness of a classification-based system for subacute and chronic low back pain. *Spine*, 37:1347–1356, 2012.

PO Astrand and K. Rodahl. *Text Book of Work Physiology: Physiological basis of exercise*. McGraw Hill, New York, USA, 1986.

Madhavi Bajekal, Shaun Scholes, Hande Love, Nathaniel Hawkins, Martin O'Flaherty, Rosalind Raine, and Simon Capewell. Analysing recent socioeconomic trends in coronary heart disease mortality in england, 2000 - 2007: A population modelling study. *PLoS Med*, 9(6), 2012. doi: 10.1371/journal.pmed.1001237.

Yaneer Bar-Yam. Improving the effectiveness of health care and public health: A multiscale complex systems analysis. *American Journal of Public Health*, 2006.

Martine J. Barons, Frances E. Griffiths, Nick Parsons, Anca Alba, Margaret Thorogood, Graham F. Medley, and Sallie Lamb. Matching patients to an intervention for back pain: classifying patients using a latent class approach. *Journal of Evaluation in Clinical Practice*, SUBMITTED, 2013a.

Martine J. Barons, Nick Parsons, Frances Griffiths, and Margaret Thorogood. A comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioural approach to treatment for non-specific low back pain. In *Computational Intelligence in Healthcare and e-health (CICARE), 2013 IEEE Symposium on*, pages 7–12, 2013b. doi: 10.1109/CICARE.2013.6583061.

Jurgen Barth, Martine Schumacher, and Christoph Herrmann-Lingen. Depression as a risk factor for mortality in patients with coronary heart disease: A meta-analysis. *Psychosom Med*, 66:802–813, 2004.

Darren John Beales, Anne Julia Smith, Peter Bruce O'Sullivan, and Leon Melville Straker. Low back pain and comorbidity clusters at 17 years of age: A cross-sectional examination of health-related quality of life and specific low back pain impacts. *Journal of Adolescent Health*, 50(5):509Ű516, 2012. ISSN 1054-139X. doi: 10.1016/j.jadohealth.2011.09.017. URL http://www.sciencedirect.com/science/article/pii/S1054139X11003405.

Katja Beesdo, Juergen Hoyer, Frank Jacobi, Nancy C. P. Low, Michael Hoefler, and Hans-Ulrich Wittchen. Association between generalized anxiety levels and pain in a community sample: Evidence for diagnostic specificity. *Journal of Anxiety Disorders*, (5), 2009.

LF Berkman, J Blumenthal, M Burg, RM Carney, D Catellier, MJ Cowan, Czajkowski SM, DeBusk R, J Hosking, A Jaffe, PG Kaufmann, P Mitchell, J Norman, LH Powell, JM Raczynski, and N Schneiderman. Effects of treating depression and low perceived social support on clinical events after myocardial infarction: The enhancing recovery in coronary heart disease patients (enrichd) randomized trial. *JAMA*, 289(23):3106–3116, 2003. doi: 10.1001/jama.289.23.3106. URL + http://dx.doi.org/10.1001/jama.289.23.3106.

H. J. N. Bethell, S. C. Turner, and M. A. Mullee. Cardiac rehabilitation in the community: 11 year follow-up after a randomized controlled trial. *Coronary Health Care*, 3:183–188, 1999.

Hugh J N Bethell and Mark A Mullee. A controlled trial of community based coronary rehabilitation. *British Heart Journal*, 64:370–375, 1990. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1224812/pdf/brheartj00060-0024.pdf.

Hugh J.N. Bethell, Allen Larvan, and Sally C. Turner. Coronary rehabilitation in the community. *Journal of the Royal College of General Practitioners*, 33:285 – 291, 1983. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1972763/pdf/jroyalcgprac00077-0031.pdf

Hugh J.N. Bethell, Robert Lewin, and Hasnain Dalal. Cardiac rehabilitation in the united kingdom. *Heart*, 95(2-3):271–275, 2009. doi: 10.1136/hrt.2007.134338. URL http://heart.bmj.com/content/95/4/271.full.pdf+html.

A Biglarian, E Hajizadeh, A Kazemnejad, and F Zayeri. Determining of prognostic factors in gastric cancer patients using artificial neural networks. *Asian Pac J Cancer Prev*, 11:533–536, 2010.

Akbar Biglarian, Enayatollah Bakhshi, Ahmad Reza Baghestani, Mahmood Reza Gohari, Mehdi Rahgozar, and Masoud Karimloo. Nonlinear survival regression using artificial neural network. *Journal of Probability and Statistics*, 2013, 2013. URL http://dx.doi.org/10.1155/2013/753930.

Tammo H.A. Bijmolt, Leo J. Paas, and Jeroen K. Vermunt. Country and consumer segmentation: multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, 21(4):323 – 340, 2004. ISSN 0167-8116. doi: 10.1016/j.ijresmar.2004.06.002. URL http://www.sciencedirect.com/science/article/pii/S0167811604000424. <ce:title>Special Issue on Global Marketing</ce:title>.

Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, USA, 1 edition, January 1996. ISBN 9780198538646. URL http://www.worldcat.org/isbn/0198538642.

Ingvar Bjelland, Tone Tangen Dahl, Alv A. amd Haug, and Dag Neckelmannd. The validity of the hospital anxiety and depression scale: An updated literature review. *Journal of Psychosomatic Research*, page 69Ű 77, 2002.

Felicity Boardman, Frances Griffiths, Renata Kokanovic, Maria Potiriadis, Christopher Dowrick, and Jane Gunn. Resilience as a response to the stigma of depression: A mixed methods analysis. *Journal of Affective Disorders*, 135 (1âĂŞ3):267 – 276, 2011. ISSN 0165-0327. doi: 10.1016/j.jad.2011.08.007. URL http://www.sciencedirect.com/science/article/pii/S0165032711004642.

C Bombardier, J Hayden, and D E Beaton. Minimal clinically important difference. low back pain: outcome measures. *The Journal of Rheumatology*, 28(2):431–438, 2001. URL http://www.jrheum.org/content/28/2/431.abstract.

Leonardo Bottaci, Philip J Drew, John E Hartley, Matthew B Hadfield, Ridzuan Farouk, Peter W R Lee, Iain M C Macintyre, Graeme S Duthie, and John R T Monson. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*, pages 469 – 472, 1997.

V.S. Bourdès, S. Bonnevay, P. J.G. Lisboa, M.S.H. Aung, S. Chabaud, T. Bachelot, D. Perol, and S. Negrier. Breast cancer predictions by neural networks analysis: a comparison with logistic regression. In *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, August 2007.

Leo Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16: 199Ű231, 2001.

Naomi Breslau, Beth A. Reboussin, James C. Anthony, and Carla L. Storr. The structure of posttraumatic stress disorder: Latent class analysis in 2 community samples. *Arch Gen Psychiatry*, 62 (12):1343–1351, 2005. doi: 10.1001/archpsyc.62.12.1343. URL http://archpsyc.ama-assn.org/cgi/content/abstract/62/12/1343.

N. E. Breslow. Analysis of survival data under the proportional hazards model. 43 (1):45–57, 1975.

Ben Bridgewater, On behalf of the Society for Cardiothoracic Surgery in GB, and Ireland. Cardiac registers: the adult cardiac surgery register. *Heart*, 96(18):1441–1443, 2010. doi: 10.1136/hrt.2010.194019. URL http://heart.bmj.com/content/96/18/1441.abstract.

James PR Brown, Alexander M Clark, Hayes Dalal, Karen Welch, and Rod S Taylor. Patient education in the management of coronary heart disease. *Cochrane Database of Systematic Reviews*, 2011. doi: 10.1002/14651858.CD008895.pub2. URL http://0-onlinelibrary.wiley.com.pugwash.lib.warwick.ac.uk/doi/10.1002/14651858.CD

RA Bruce. Principles in exercise testing. In Hellerstein HK, editor, *Exercise Testing and Exercise Training in Coronary Heart Disease*. Academic Press, New York, USA, 1973.

R.A. Bruce, F. Kusumi, and D. Hosmer. Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal*, 85(4):546 – 562, 1973. ISSN 0002-8703. doi: 10.1016/0002-8703(73)90502-4. URL http://www.sciencedirect.com/science/article/pii/0002870373905024.

B Buckley and A W Murphy. Do patients with angina alone have a more benign prognosis than patients with a history of acute myocardial infarction, revascularisation or both? findings from a community cohort study. *Heart*, 95(6):461–467, 2009. doi: 10.1136/hrt.2008.146944.

Christopher Burton. Heavy tailed distributions of effect sizes in systematic reviews of complex interventions. *PLoS ONE*, 7(3):e34222, 03 2012. doi: 10.1371/journal.pone.0034222.

Christopher Burton, Rachel A. Heath, David Weller, and Michael Sharpe. Evidence of reduced complexity in self-report data from patients with medically unexplained

symptoms. *Nonlinear dynamics psychology and life sciences*, 14(1):15–25, 2010. ISSN 1090-0578.

Michelle Campbell, Ray Fitzpatrick, Andrew Haines, Ann Louise Kinmonth, Peter Sandercock, David Spiegelhalter, and Peter Tyrer. Framework for design and evaluation of complex interventions to improve health. *BMJ*, 321(7262):694–696, 9 2000. doi: 10.1136/bmj.321.7262.694.

Neil C Campbell, Elizabeth Murray, Janet Darbyshire, Jon Emery, Andrew Farmer, Frances Griffiths, Bruce Guthrie, Helen Lester, Phil Wilson, and Ann Louise Kinmonth. Designing and evaluating complex interventions to improve health care. *BMJ*, pages 455–459, 2007.

Robert M. Carney, Kenneth E. Freedland, Brian Steinmeyer, James A. Blumenthal, Lisa F. Berkman, Lana L. Watkins, Susan M. Czajkowski, Matthew M. Burg, and Allan S. Jaffe. Depression and five year survival following acute myocardial infarction: A prospective study. *Journal of Affective Disorders*, 109 (1Ű2):133 – 138, 2008. ISSN 0165-0327. doi: 10.1016/j.jad.2007.12.005. URL `http://www.sciencedirect.com/science/article/pii/S0165032707004260`.

Elizabeth Casey, Joel Hughes, Donna Waechter, Richard Josephson, and James Rosneck. Depression predicts failure to complete phase-ii cardiac rehabilitation. *Journal of Behavioral Medicine*, 31:421–431, 2008. ISSN 0160-7715. URL `http://dx.doi.org/10.1007/s10865-008-9168-1`. 10.1007/s10865-008-9168-1.

Damon Centola. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702Ű34, 2007.

Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329:1194, 2010. doi: 10.1126/science.1185231.

Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science*, page 1269, 2011. doi: 10.1126/science.1207055.

Damon Centola, Robb Willer, and Michael Macy. The emperorŠs dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, 110(4): 1009Ű40, 2005.

Roberto Cervera, Faisal Bakaeen, Lorraine Cornwell, XingLi Wang, Joseph Coselli, Scott LeMaire, and Danny Chu. Impact of functional status on survival after coronary artery bypass graft in a veteran population. *Ann Thorac Surg*, 93:1950–1955, 2012.

S. L. Chapman and J. S. Pemberton. Prediction of treatment outcome from clinically derived mmpi clusters in rehabilitation for chronic low back pain. *The Clinical Journal of Pain*, pages 267–276, 1994.

ZhengMing Chen, Peter Sandercock, HongChao Pan, Carl Counsell, Rory Collins, LiSheng Liu, JingXian Xie, Charles Warlow, Richard Peto, on behalf of the CAST, and IST collaborative groups. Indications for early aspirin use in acute ischemic stroke: A combined analysis of 40 000 randomized patients from the chinese acute stroke trial and the international stroke trial. *Stroke*, 31(6):1240–1249, 2000. doi: 10.1161/01.STR.31.6.1240. URL `http://stroke.ahajournals.org/content/31/6/1240.abstract`.

Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *The New England journal of medicine*, 357:370–9, 2007.

Nicholas A. Christakis and James H. Fowler. Quitting in droves: Collective dynamics of smoking behavior in a large social network. *N Engl J Med.*, 358:2249Ű2258, 2008. doi: 10.1056/NEJMsa0706154.

D. Collett. *Modelling survival data in medical research*. 2nd edition, texts in statistical science. Chapman & Hall/CRC, Boca Raton, 2003.

Linda M Collins and Stephanie T Lanza. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, 2010.

COMET. Comet initiative, Accessed 25/05/2013. URL `http://www.comet-initiative.org/studies/details/284?result=true`.

Gregory F. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

D.R. Cox and E.J. Snell. A general definition of residuals (with discusion). *Journal of the Royal Statistical Society, A*, 30, 1968.

Peter Craig, Paul Dieppe, Sally Macintyre, Susan Michie, Irwin Nazareth, and Mark Petticrew. Developing and evaluating complex interventions: the new medical research council guidance. *BMJ*, 337, 9 2008. doi: 10.1136/bmj.a1655.

JA Critchley and S Capewell. Smoking cessation for the secondary prevention of coronary heart disease. *Cochrane Database of Systematic Reviews*, 2003. doi: 10.1002/14651858.CD003041.pub2.

Jean Dallongevillle, Dirk De Bacquer, Jan Heidrich, Guy De Backer, Christoph Prugger, Kornelia Kotseva, Michele Montaye, and Philippe on behalf of the EUROASPIRE Study Group Amouyel. Gender differences in the implementation of cardiovascular prevention measures after an acute coronary event. *Heart 2010;96:1744e1749. doi:*, 96:1744–1749, 2010. doi: 10.1136/hrt.2010.196170. URL `http://heart.bmj.com/content/96/21/1744.full.pdf+html`.

Bertil Damato, Antonio Eleuteri, Anthony C. Fisher, Sarah E. Coupland, and Azzam F.G. Taktak. Artificial neural networks estimating survival probability after treatment of choroidal melanoma. *Ophthalmology*, 115(9):1598 – 1607, 2008. ISSN 0161-6420. doi: 10.1016/j.ophtha.2008.01.032. URL `http://www.sciencedirect.com/science/article/pii/S0161642008000997`.

Leon Danon, Thomas A. House, Jonathan M. Read, and Matt J. Keeling. Social encounter networks: collective properties and disease transmission. *J. R. Soc. Interface*, page 2826Ű2833, 2012. doi: 10.1098/rsif.2012.0357.

RA Dayo, M Battie, AJ Beurskens, C Bombardier, P Croft, B Koes, A Malmivaara, M Roland, M Von Korff, and G Waddell. Outcome measures for low back pain research. a proposal for standardised use. *Spine*, 23(18):2003–2013, 1988.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):pp. 1–38, 1977. ISSN 00359246. URL `http://www.jstor.org/stable/2984875`.

William D'Hoore, André Bouckaert, and Charles Tilquin. Practical considerations on the use of the charlson comorbidity index with administrative data bases. *Journal of Clinical Epidemiology*, 49(12):1429 – 1433, 1996. ISSN 0895-4356. doi: 10.1016/S0895-4356(96)00271-5. URL `http://www.sciencedirect.com/science/article/pii/S0895435696002715`.

C. Dickens, L. McGowan, C. Percival, B. Tomenson, L. Cotter, A. Heagerty, and F. Creed. Depression is a risk factor for mortality after myocardial infarction fact or artifact? *Journal of the American College of Cardiology*, 49(18):1834–1840, 2007.

C Dionne. *Low Back Pain*, volume 31. IAS Press, Seattle, WA, 1999. ISBN 0-931092-25-6. doi: 10.1093/ije/31.2.506. URL http://ije.oxfordjournals.org/content/31/2/506.short.

Laura E. Dobson, Robert J. Lewin, Patrick Doherty, Phillip D. Batin, Simon Megarry, and Christopher P. Gale. Is cardiac rehabilitation still relevant in the new millennium? *J Cardiovasc Med*, 13:32Ű37, 2012.

S. Dolgobrodov, P. Moore, R. Marshall, R. Bittern, R. Steele, and A. Cuschieri. Artificial neural network: Predicted vs. observed survival in patients with colonic cancer. *Diseases of the Colon & Rectum*, 50(2):184–191, 2007. doi: doi:10.1007/s10350-006-0779-8. URL http://www.ingentaconnect.com/content/klu/10350/2007/00000050/00000002/00000779.

Wen Dong, Katherine Heller, , and Alex (Sandy) Pentland. Modeling infection with multi-agent dynamics. In *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP12)*, 2012.

Joan Dorn, John Naughton, Dai Imamura, and Maurizio Trevisan. Results of a multicenter randomized clinical trial of exercise and long-term survival in myocardial infarction patients the national exercise and heart disease project (nehdp). *Circulation*, 100:1764–1769, 1999. doi: 10.1161/01.CIR.100.17.1764.

Kerstin Dudas, Georg Lappas, and Annika Rosengren. Long-term prognosis after hospital admission for acute myocardial infarction from 1987 to 2006. *International Journal of Cardiology*, 155(3):400–405, 2012. URL http://linkinghub.elsevier.com/retrieve/pii/S0167527310008958?showall=true.

Kate M. Dunn, Kelvin Jordan, and Peter R. Croft. Characterizing the course of low back pain: A latent class analysis. *American Journal of Epidemiology*, 163(8):754–761, 2006. doi: 10.1093/aje/kwj100. URL http://aje.oxfordjournals.org/content/163/8/754.abstract.

Richard Dybowski and Vanya Gant. *Clinical applications of artificial neural networks*. Cambridge University Press, Cambridge, UK, 2001.

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

Elaine Eggleston, John Laub, and Robert Sampson. Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, 20:1–26, 2004. ISSN 0748-

4518. URL http://dx.doi.org/10.1023/B:JOQC.0000016696.02763.ce. 10.1023/B:JOQC.0000016696.02763.ce.

Michael Eid, Rolf Langeheine, and Ed Diener. Comparing typological structures across cultures by multigroup latent class analysis. *Journal of Cross-Cultural Psychology*, 34(2):195–210, 2003. doi: 10.1177/0022022102250427. URL http://jcc.sagepub.com/content/34/2/195.abstract.

Sandra Eldridge, Anne Spencer, Colin Cryer, Suzanne Parsons, Martin Underwood, and Gene Feder. Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. *Journal of Health Services Research & Policy*, 10(3):133–142, 2005.

A. Eleuteri, M.S.H. Aung, A.F.G. Taktak, B. Damato, and P.J.G Lisboa. Continuous and discrete time survival analysis: neural network approaches. In *Proceedings of the 29th annual international conference of the IEEE EMBS*, August 2007a.

A. Eleuteri, M.S.H. Aung, A.F.G. Taktak, B. Damato, and P.J.G. Lisboa. Bayesian neural networks for survival analysis: a comparative study. In *Proceedings of Computational Intelligence in Medicine*, 2007b.

Antonio Eleuteri, Roberto Tagliaferri, Leopoldo Milano, Sabino De Placido, and Michele De Laurentiis. A novel neural network-based survival analysis model. *Neural Networks*, 16(5Ũ6):855 – 864, 2003. ISSN 0893-6080. doi: 10.1016/S0893-6080(03)00098-4. URL http://www.sciencedirect.com/science/article/pii/S0893608003000984. <ce:title>Advances in Neural Networks Research: IJCNN '03</ce:title>.

K. Miriam Elfstrom, Dustin Hatefi, Patrick D. Kilgo, John D. Puskas, Vinod H. Thourani, Robert A. Guyton, and Michael E. Halkos. What happens after discharge? an analysis of long-term survival in cardiac surgical patients requiring prolonged intensive care. *Journal of Cardiac Surgery*, 27(1):13–19, 2012. ISSN 1540-8191. doi: 10.1111/j.1540-8191.2011.01341.x. URL http://dx.doi.org/10.1111/j.1540-8191.2011.01341.x.

Richard Emsley, Graham Dunn, and Ian R White. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research*, 19(3):237–270, 2010. doi: 10.1177/0962280209105014. URL http://smm.sagepub.com/content/19/3/237.abstract.

J. Eng. Predicting the presence of acute pulmonary embolism: A comparative analysis of the artificial neural network, logistic regression, and threshold models. *AJR*, pages 869–874, 2002.

Eric A. Engels, Michael D. Sinclair, Robert J. Biggar, Denise Whitby, Peter Ebbesen, James J. Goedert, and Joseph L. Gastwirth. Latent class analysis of human herpesvirus 8 assay performance and infection prevalence in sub-saharan africa and malta. *International Journal of Cancer*, 88(6):1003–1008, 2000. ISSN 1097-0215. doi: 10.1002/1097-0215(20001215)88:6<1003::AID-IJC26>3.0.CO;2-9. URL http://dx.doi.org/10.1002/1097-0215(20001215)88:6<1003::AID-IJC26>3.0.CO;2-9.

Jeremy Fairbank, Stephen E. Gwilym, John France, Scott D. Daffner, Joseph Dettori, Jeff Hermsmeyer, and Gunnar Andersson. The role of classification of chronic low back pain. *Spine*, page S19Ű S42, 2011.

Stephen W. Farrell, LeeAnn Braun, Carolyn E. Barlow, Yiling J. Cheng, and Steven N. Blair. The relation of body mass index, cardiorespiratory fitness, and all-cause mortality in women. *Obesity Research*, 10(6):417–423, 2002. ISSN 1550-8528. doi: 10.1038/oby.2002.58. URL http://dx.doi.org/10.1038/oby.2002.58.

Giovanni Filardo, Cody Hamilton, Paul A. Grayburn, Hui Xu, Jr Hebeler, Robert F., and Baron Hamman. Established preoperative risk factors do not predict long-term survival in isolated coronary artery bypass grafting patients. *Ann Thorac Surg*, 93(6):1943–1948, 2012. doi: 10.1016/j.athoracsur.2012.02.072. URL http://ats.ctsnetjournals.org/cgi/content/abstract/93/6/1943.

James H Fowler and Nicholas A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *BMJ*, page 2338, 2008. doi: 10.1136/bmj.a2338.

Kim Fox, Jeffrey S. Borer, A. John Camm, Nicolas Danchin, Roberto Ferrari, Jose L. Lopez Sendon, Philippe Gabriel Steg, Jean-Claude Tardif, Luigi Tavazzi, and Michal Tendera. Resting heart rate in cardiovascular disease. *Journal of the American College of Cardiology*, 50(9):823 – 830, 2007. ISSN 0735-1097. doi: 10.1016/j.jacc.2007.04.079. URL http://www.sciencedirect.com/science/article/pii/S0735109707018232.

Kim Fox, Ian Ford, P Gabriel Steg, Michal Tendera, Michele Robertson, and Roberto Ferrari. Heart rate as a prognostic risk factor in patients with coronary artery disease and left-ventricular systolic dysfunction (beautiful): a subgroup analysis of a randomised controlled trial. *The Lancet*, 372(9641):817

– 821, 2008. ISSN 0140-6736. doi: DOI: 10.1016/S0140-6736(08)61171-X. URL http://www.sciencedirect.com/science/article/B6T1B-4TB3N73-2/2/3984fe859177f63c09f

Sarah W Fraser and Trisha Greenhalgh. Coping with complexity: educating for capability. *BMJ*, 323:799Ű803, 2001.

Nancy Frasure-Smith, François Lespérance, and Mario Talajic. Depression and 18-month prognosis after myocardial infarction. *Circulation*, 91(4):999–1005, 1995. doi: 10.1161/01.CIR.91.4.999. URL http://circ.ahajournals.org/content/91/4/999.abstract.

A. Fukushima, K. Asshizawal, T. Yamaguchi, N. Matsuyama, H. Hayashi, I. Kida, Y. Imafuku, A. Egawa, S. Kimura, K. Nagaoki, S. Honda, S. Katsuragawa, K. Doi, and K. Hayashi. Application of an artificial neural network to high-resolution ct: Usefulness in differential diagnosis of diffuse lung disease. *AJR*, 183:297–305, 2004.

Yoshifumi Fukushima, Hirotoshi Ohmura, Hiroshi Mokuno, Kan Kajimoto, Takatoshi Kasai, Satoshi Hirayamaand Katsumi Miyauchi, Takashi Miida, Atsushi Amano, and Hiroyuki Daida. Non-high-density lipoprotein cholesterol is a practical predictor of long-term cardiac death after coronary artery bypass grafting. *Atherosclerosis*, 221:206–211, 2012. doi: 10.1016/j.atherosclerosis.2011.12.012.

C P Gale, S O M Manda, P D Batin, C F Weston, J S Birkhead, and Hall A S. Predictors of in-hospital mortality for patients admitted with st-elevation myocardial infarction: a real-world study using the myocardial infarction national audit project (minap) database. *Heart*, 94:1407Ű1412, 2008. doi: 10.1136/hrt.2007.127068. URL http://heart.bmj.com/content/94/11/1407.full.pdf+html.

Carol R. Geary and Karen L. Schumacher. Care transitions: Integrating transition theory and complexity science concepts. *Advances in Nursing Science*, 35(3): 236Ű248, 2012.

Pantaleo Giannuzzi, Pier Luigi Temporelli, Roberto Marchioli, Aldo Pietro Maggioni, Gianluigi Balestroni, Vincenzo Ceci, Carmine Chieffo, Marinella Gattone, Raffaele Griffo, Carlo Schweiger andLuigi Tavazzi, Stefano Urbinati, Franco Valagussa, Diego Vanuzzo, and for the GOSPEL Investigators. Global secondary prevention strategies to limit event recurrence after myocardial infarction: Results of the gospel study, a multicenter, randomized controlled trial from the italian cardiac rehabilitation network. *Archives of Internal Medicine*, 168:2194–2204, 2008. doi: 10.1001/archinte.168.20.2194. URL + http://dx.doi.org/10.1001/archinte.168.20.2194.

D Gill and S Hatcher. Antidepressants for depression in medical illness. *The Cochrane Database of Systematic Reviews (Complete Reviews)*, 2000. doi: 10.1002/14651858.CD001312.

H Gioftsos and DW Grieve. The use of artificial neural networks to identify patients wirth chronic low back pain conditions from patterns of sit-to-stand manoeveures. 11:275–280, 1996.

Janette Greenhalgh, Juliet Hockenhull, Naveen Rao, Yena Dundar, Rumona C Dickson, and Adrian Bagust. Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes. *Cochrane Database of Systematic Reviews*, 5, 2010. doi: 10.1002/14651858.CD004587.pub2. URL `http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD004587/frame.ht` http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD004587.pub2/full.

Frances Griffiths, Jeffrey Borkan, David Byrne, Benjamin F. Crabtree, Chris Dowrick, Jane Gunn, Renata Kokanovic, Sarah Lamb, Antje Lindenmeyer, Michael Parchman, Shmuel Reis, and Jackie Sturt. Developing evidence for how to tailor medical interventions for the individual patient. *Qualitative Health Research*, 20(12):1629–1641, 2010. doi: 10.1177/1049732310377453. URL `http://qhr.sagepub.com/content/20/12/1629.abstract`.

Frances Griffiths, Jonathan Cave, Felicity Boardman, Justin Ren, Teresa Pawlikowska, Robin Ball, Aileen Clarke, and Alan Cohen. Social networks - the future for health care delivery. *Social Science & Medicine*, 2012.

Morten Grundtvig. Patients with micro or other myocardial infarctions have equal long-term survival. *Scandinavian Cardiovascular Journal*, 46(2):76–80, 2012. doi: 10.3109/14017431.2011.653824. URL `http://informahealthcare.com/doi/abs/10.3109/14017431.2011.653824`.

Morten Grundtvig, Terje Hagen, Elin Amrud, and Aasmund Reikvam. Mortality after myocardial infarction: impact of gender and smoking status. *European Journal of Epidemiology*, 26:385–393, 2011. ISSN 0393-2990. URL `http://dx.doi.org/10.1007/s10654-011-9557-6`. 10.1007/s10654-011-9557-6.

D C Haas. Depression and disability in coronary patients: time to focus on quality of life as an end point. *Heart*, 92:8Ű10, 2006. doi: 10.1136/hrt.2005.066035.

Dirk Hallner and Monika Hasenbring. Classification of psychosocial risk factors (yellow flags) for the development of chronic low back and leg pain using artificial neural network. *Neuroscience Letters*, 361:151Ű154, 2004.

Gaël P Hammer, Jean-Baptist du Prel, and Maria Blettner. Avoiding bias in observational studies: A review article. *Deutsches Ärzteblatt International*, 106(41): 664Ű8, 2009. Translated from the original German by Rodney A. Yeates, M.A., Ph.D. Part 8 in a Series of Articles on Evaluation of Scientific Publications.

David J. Hand. Evaluating diagnostic tests: The area under the roc curve and the balance of errors. *Statistics in Medicine*, 29:1502–1510, 2010.

Edward L. Hannan. Predicting long-term survival for coronary artery bypass graft surgery and percutaneous coronary intervention. *Circulation*, 125 (12):1475–1476, 2012. doi: 10.1161/CIRCULATIONAHA.112.095943. URL `http://circ.ahajournals.org/content/125/12/1475.short`.

Henrik Haraldsson, Lars Edenbrandt, and Mattias Ohlsson. Detecting acute myocardial infarction in the 12-lead ecg using hermite expansions and neural networks. *Artificial Intelligence in Medicine*, 32(2):127 – 136, 2004. ISSN 0933-3657. doi: 10.1016/j.artmed.2004.01.003. URL `http://www.sciencedirect.com/science/article/pii/S0933365704000144`.

SL Hardoon, PH Whincup, LT Lennon, SG Wannamethee, S Capewell, and RW Morris. How much of the recent decline in the incidence of myocardial infarction in british men can be explained by changes in cardiovascular risk factors? evidence from a prospective population-based study. *CIRCULATION*, 117(5):598 – 604, 2008.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd edition, Springer series in statistics. Springer Science and Business Media, New York, USA, 2009.

Penelope Hawe, Alan Shiell, and Therese Riley. Complex interventions: how 'out of control' can a randomised controlled trial be? *British Medical Journal*, pages 1561–3, 2004.

The Health-Foundation. Complex adaptive systems. *Evidence Scan*, 2010.

Jeffrey J. Hebert, Shane L. Koppenhaver, and Bruce F. Subgrouping patients with low back pain: A treatment-based approach to classification. *Athletic Training*, 2011.

Balraj S. Heran, Jenny MH Chen, Shah Ebrahim, Tiffany Moxham, Neil Oldridge, Karen Rees, David R Thompson, and Rod S Taylor. Exercise-based cardiac rehabilitation for coronary heart

disease. 2011. doi: 10.1002/14651858.CD001800.pub2. URL http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD001800/frame.ht

Julian PT Higgins and Sally Green, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. The Cochrane Collaboration, 2011.

Jonathan C. Hill, Kate M. Dunn, Martyn Lewis, Ricky Mullis, Chris J. Main, Nadine E. Foster, and Elaine M. Hay. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Care & Research*, 59(5):632–641, 2008. ISSN 1529-0131. doi: 10.1002/art.23563. URL http://dx.doi.org/10.1002/art.23563.

Jonathan C Hill, David GT Whitehurst, Martyn Lewis, Stirling Bryan, Kate M Dunn, Nadine E Foster, Kika Konstantinou, Chris J Main, Elizabeth Mason, Simon Somerville, Gail Sowden, Kanchan Vohora, and Elaine M Hay. Comparison of stratified primary care management for low back pain with current best practice (start back): a randomised controlled trial. *The Lancet*, 378(9802): 1560 – 1571, 2011. ISSN 0140-6736. doi: 10.1016/S0140-6736(11)60937-9. URL http://www.sciencedirect.com/science/article/pii/S0140673611609379.

L Hooper, CD Summerbell, R Thompson, D Sills, FG Roberts, HJ Moore, and G Davey Smith. Reduced or modified dietary fat for preventing cardiovascular disease. *Cochrane Database of Systematic Reviews*, 2012. doi: 10.1002/14651858.CD002137.pub3.

Thomas House. Modelling behavioural contagion. *J. R. Soc. Interface*, 2011. doi: 10.1098/rsif.2011.0014. The author would like to thank Martine Barons and Matt Keeling for helpful comments relating to this work.

J Hughes, F Kee, K Bennett, M O'Flaherty, J Critchley, M Cupples, and S Capewell. Modelling coronary heart disease mortality in northern ireland between 1987 and 2007. *Journal of Epidemiology and Community Health*, 65(Suppl 2):A9–A10, 2011. doi: 10.1136/jech.2011.143586.21.

Jose M. Jerez, Ignacio Molina, Pedro J. Garcia-Laencina, Emilio Alba, Nuria Ribelles, Miguel Martin, and Leonardo Franco. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, 50:105–115, 2010. doi: 10.1016/j.artmed.2010.05.002.

Holly Johnson, Dr. Sergio Camorlinga, and Ryan DeForge. Healthcare organizations as complex adaptive systems: Lessons for nurse leaders. In *European Conference on Complex Systems 2010*. Complex Systems Society, 2010.

Judith Jolliffe, Karen Rees, Rod R.S. Taylor, David R. Thompson, Neil Oldridge, and Shah Ebrahim. Exercise-based cardiac rehabilitation for coronary heart disease (review). 2009. doi: 10.1002/14651858.CD001800.

Kate Jolly, Rod S Taylor, Gregory Y.H Lip, and Andrew Stevens. Home-based cardiac rehabilitation compared with centre-based rehabilitation and usual care: A systematic review and meta-analysis. *International Journal of Cardiology*, 111(3):343–351, 2006. URL `http://linkinghub.elsevier.com/retrieve/pii/S0167527305013306`.

Rashmi Joshi, Colin Reeves, and Charles Johnson. Beyond the cox model: Artificial neural networks for survival analysis. In *Proceedings of the 16th International Conference on Systems Engineering*, September 2003.

Rashmi Raj Joshi. *Modelling Survival Time Distributions of Cancer Data Using Artificial Neural Networks*. PhD thesis, Coventry University, 2004.

Rashmi Raj Joshi, Colin Reeves, and Charles Johnston. *Probabilistic Artificial Neural Networks for Malignant Melanoma Prognosis*. Springer Vienna, 2005. ISBN 978-3-211-27389-0.

T Kailath. *Linear Systems*. Prentice=Hall, Englewood Cliffs, NJ, 1980.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, 2002.

Kirsten I. Kaptein, Peter de Jonge, Rob H. S. van den Brink, and Jakob Korf. Course of depressive symptoms after myocardial infarction and cardiac prognosis: A latent class analysis. *Psychosomatic Medicine*, 68(5):662–668, 2006. doi: 10.1097/01.psy.0000233237.79085.57. URL `http://www.psychosomaticmedicine.org/content/68/5/662.abstract`.

David A. Katerndahl. Power laws in covariability of anxiety and depression among newly diagnosed patients with major depressive episode, panic disorder and controls. *Journal of Evaluation in Clinical Practice*, 15(3):565–570, 2009. ISSN 1365-2753. doi: 10.1111/j.1365-2753.2009.01166.x. URL `http://dx.doi.org/10.1111/j.1365-2753.2009.01166.x`.

Terence Kavanagh, Larry F. Mertens, Donald J.and Hamm, Joseph Beyene, Johanna Kennedy, Paul Paul Corey, and Roy J. Shephard. Prediction of long-term prognosis in 12 169 men referred for cardiac rehabilitation. *Circulation*, 2002. doi: 10.1161/01.CIR.0000024413.15949.ED.

Terence Kavanagh, Donald J Mertens, Larry F Hamm, Joseph Beyene, Johanna Kennedy, Paul Corey, and Roy J Shephard. Peak oxygen intake and cardiac mortality in women referred for cardiac rehabilitation. *Journal of the American College of Cardiology*, 42(12):2139–2143, 2003. doi: 10.1016/j.jacc.2003.07.028. URL + http://dx.doi.org/10.1016/j.jacc.2003.07.028.

Pamela K. Keel, Manfred Fichter, Norbert Quadflieg, Cynthia M. Bulik, Mark G. Baxter, Laura Thornton, Katherine A. Halmi, Allan S. Kaplan, Michael Strober, D. Blake Woodside, Scott J. Crow, James E. Mitchell, Alessandro Rotondo, Mauro Mauri, Giovanni Cassano, Janet Treasure, David Goldman, Wade H. Berrettini, and Walter H. Kaye. Application of a latent class analysis to empirically define eating disorder phenotypes. *Arch Gen Psychiatry*, 61(2):192–200, 2004. doi: 10.1001/archpsyc.61.2.192. URL http://archpsyc.ama-assn.org/cgi/content/abstract/61/2/192.

P Kent and H H Laurisden. Managing missing scores on the roland morris disability questionnaire. *Spine*, 36(22):1878–1884, 2011.

Peter Kent and Jennifer L Keating. Classification in nonspecific low back pain: What methods do primary care clinicians currently use? *Spine*, 30:1433–1440, 2005.

Peter Kent, Jennifer L Keating, and Charlotte Leboeuf-Yde. Research methods for subgrouping low back pain. *BMC Medical Research Methodology*, 10, 2010. doi: 10.1186/1471-2288-10-62.

M. Ketonen, P. Pajunen, H. Koukkunen, P. Immonen-Räihä, J. Mustonen, M. Mähönen, M. Niemelä, K. Kuulasmaa, P. Palomäki, M. Arstila, T. Vuorenmaa, A. Lehtonen, S. Lehto, H. Miettinen, J. Torppa, J. Tuomilehto, J. Airaksinen, K. Pyörälä, and V. Salomaa. Long-term prognosis after coronary artery bypass surgery. *International Journal of Cardiology*, 124(1): 72 – 79, 2008. ISSN 0167-5273. doi: 10.1016/j.ijcard.2006.12.026. URL http://www.sciencedirect.com/science/article/pii/S0167527307003439.

N.A. Khovanova, I.A. Khovanov, L. Sbano, F. Griffiths, and T.A. Holt. Characterisation of linear predictability and non-stationarity of subcutaneous glucose profiles. *Computer Methods and Programs in Biomedicine*, 2012.

Mika Kivimäki, Martin J Shipley, Jane E Ferrie, Archana Singh-Manoux, G David Batty, Tarani Chandola, Michael G Marmot, and George Davey Smith. Best-practice interventions to reduce socioeconomic inequalities of coronary heart disease mortality in uk: a prospec-

tive occupational cohort study. *The Lancet*, 372(9650):1648 – 1654, 2008. ISSN 0140-6736. doi: 10.1016/S0140-6736(08)61688-8. URL http://www.sciencedirect.com/science/article/pii/S0140673608616888.

E. David Klonsky and Thomas M. Olino. Identifying clinically distinct subgroups of self-injurers among young adults: A latent class analysis. *Journal of Consulting and Clinical Psychology, Special Section: Suicide and Nonsuicidal Self-Injury.*, 76(1):22–27, 2008. doi: 10.1037/0022-006X.76.1.22.

S Kodama, K Saito, S Tanaka, and et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA: The Journal of the American Medical Association*, 301(19):2024–2035, 2009. doi: 10.1001/jama.2009.681. URL + http://dx.doi.org/10.1001/jama.2009.681.

Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89 – 109, 2001. ISSN 0933-3657. doi: 10.1016/S0933-3657(01)00077-X. URL http://www.sciencedirect.com/science/article/pii/S093336570100077X.

Michael R. Von Korff and James C. Anthony. The {NIMH} diagnostic interview schedule modified to record current mental status. *Journal of Affective Disorders*, 4(4):365 – 371, 1982. ISSN 0165-0327. doi: 10.1016/0165-0327(82)90032-5.

S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. informatica 31:249Ű268, 2007.

Maria Teresa La Rovere, Chiara Bersano, Marco Gnemmi, Giuseppe Specchia, and Peter J Schwartz. Exercise-induced increase in baroreflex sensitivity predicts improved prognosis after myocardial infarction. *Circulation*, 2002.

Nicola Lama, Patrizia Boracchi, and Elia Biganzoli. Partial logistic relevance vector machines in survival analysis. *Journal of Applied Statistics*, 38(11):2445–2458, January 2011. URL http://ideas.repec.org/a/taf/japsta/v38y2011i11p2445-2458.html.

S Lamb, R Lall, Z Hansen, E Castelnuovo, E Withers, and et al. A multicentred randomised controlled trial of a primary-care based cognitive behavioural programme for low back pain. the back skills training (best) trial. *Health Technol Assess*, 14 (41), 2010a.

Sarah E Lamb, Zara Hansen, Ranjit Lall, Emanuela Castelnuovo, Emma J Withers, Vivien Nichols, Rachel Potter, and Martin R Underwood. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. *The Lancet*, 375(9718):916 – 923, 2010b. ISSN 0140-6736. doi: 10.1016/S0140-6736(09)62164-4.

GA Lancaster, MJ Campbell, S. Eldridge, A. Farrin, M. Marchant, S. Muller, R. Perera, TJ Peters, AT Prevost, and G. Rait. Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Statistical Methods in Medical Research*, 19(4):349–377, 2010. doi: 10.1177/0962280209359883. URL `http://smm.sagepub.com/content/19/4/349.abstract`.

Jennifer M Langworthy and Alan C Breen. Rationalizing back pain: the development of a classification system through cluster analysis. *Journal of manipulative and physiological therapeutics*, 20:303–310, 1997.

Henrik H Lauridsen, Jan Hartvigsen, Claus Manniche, Lars Korsholm, and Niels Grunnet-Nilsson. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskeletal Disorders*, 7, 2006. doi: 10.1186/1471-2474-7-82.

Paul F. Lazarsfeld and Neil W. Henry. *Latent Structure Analysis*. Houghton Mifflin, Boston, USA, 1968.

Chong Do Lee, Steven N Blair, and Andrew S Jackson. Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *The American Journal of Clinical Nutrition*, 69(3):373–380, 1999. URL `http://ajcn.nutrition.org/content/69/3/373.abstract`.

Chung-Yol Lee, Renato Kratter, Nicole Duvoisin, Aydin Taskin, and Julian Schilling. Cross-sectional view of factors associated with back pain. *International Archives of Occupational and Environmental Health*, 78:319–324, 2005. ISSN 0340-0131. doi: 10.1007/s00420-004-0589-z. URL `http://dx.doi.org/10.1007/s00420-004-0589-z`.

R J P Lewin, G Furze, J Robinson, K Griffith, S Wiseman, M Pye, and R Boyle. A randomised controlled trial of a self-management plan for patients with newly diagnosed angina. *British Journal of General Practice*, 52:194–201, 2002.

Tricia Y. Li, Jamal S. Rana, JoAnn E. Manson, Walter C. Willett, Meir J. Stampfer, Graham A. Colditz, Kathryn M. Rexrode, and Frank B. Hu. Obesity as compared

with physical activity in predicting risk of coronary heart disease in women. *Circulation*, 113(4):499–506, 2006. doi: 10.1161/CIRCULATIONAHA.105.574087. URL http://circ.ahajournals.org/content/113/4/499.abstract.

Oliver J Liakopoulos, Elmar W Kuhn, Ingo Slottosch, Gernot Wassmer, and Thorsten Wahlers. Preoperative statin therapy for patients undergoing cardiac surgery. 2012. doi: 10.1002/14651858.CD008493.pub2. URL http://0-onlinelibrary.wiley.com.pugwash.lib.warwick.ac.uk/doi/10.1002/14651858.CD

Steven James Linton. Risk factors for neck and back pain in a working population in sweden. *Work & Stress: An International Journal of Work, Health & Organisations*, 4:41–49, 1990.

Drew A. Linzer and Jeffrey B. Lewis. polca: An r package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1–29, 6 2011a. ISSN 1548-7660. URL http://www.jstatsoft.org/v42/i10.

Drew A. Linzer and Jeffrey B. Lewis. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1–29, 2011b. URL http://www.jstatsoft.org/v42/i10/.

Paulo J. G. Lisboa, Terence A. Etchells, Ian H. Jarman, Corneliu T. C. Arsene, M. S. Hane Aung, Antonio Eleuteri, Azzam F. G. Taktak, Federico Ambrogi, Patrizia Boracchi, and Elia Biganzoli. Partial logistic artificial neural network for competing risks regularized with automatic relevance determination. *IEEE Transactions on neural networks*, 20(9), 2009. doi: 10.1109/TNN.2009.2023654. URL http://ieeexplore.ieee.org.

Paulo J.G. Lisboa, Elia M. Biganzoli, Azzam F. Taktak, Terence A. Etchells, Ian H. Jarman, M.S.Hane Aung, and Federico Ambrogi. Assessing flexible models and rule extraction from censored survival data. In *Proceedings of IEEE International Joint Conference on Neural Networks*, August 2007. doi: 10.1109/IJCNN.2007.4371207.

P.J.G. Lisboa, H. Wong, P. Harris, and R. Swindell. A bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artificial Intelligence in Medicine*, 28(1): 1 – 25, 2003. ISSN 0933-3657. doi: 10.1016/S0933-3657(03)00033-2. URL http://www.sciencedirect.com/science/article/pii/S0933365703000332.

R. J. A. Little and D. A. Rubin. *Statistical analysis with missing data*. 2nd edition, series in probability and statistics. Wiley-Interscience, New Jersey, USA, 2002.

D.J.C. MacKay. Bayesian non-linear modelling for the prediction competition. In *ASHRAE Transactions, V.100, Pt.2*, pages 1053–1062, Atlanta Georgia, 1994. ASHRAE.

Todd A. MacKenzie, Jeremiah R. Brown, Donald S. Likosky, YingXing Wu, and Gary L. Grunkemeier. Review of case-mix corrected survival curves. *The Annals of Thoracic Surgery*, 93(5):1416 – 1425, 2012. ISSN 0003-4975. doi: 10.1016/j.athoracsur.2011.12.094. URL `http://www.sciencedirect.com/science/article/pii/S0003497512002779`.

John Maindonald and W. John Braun. *DAAG: Data Analysis And Graphics data and functions*, 2012. URL `http://CRAN.R-project.org/package=DAAG`. R package version 1.15.

B Martin, R. Arena, T. Hauer, S. Aggarwal, and Stone J. Cardiovascular fitness and mortality following contemporary cardiac rehabilitation. *Canadian Journal of Cardiology*, 27, 2011. doi: 10.1016/j.cjca.2011.07.092.

K.and Watanabel H. Matsuki, Y.and Nakamura, T. Aokil, H. Nakatal, S. Katsuragawa, and K. Doi. Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution ct evaluation with receiver operating characteristic analysis. *AJR*, 178:657–663, 2002.

Elaine F. Maughan and Jeremy S. Lewis. Outcome measures in chronic low back pain. *European Spine Journal*, page 1484Ű1494, 2010. doi: 10.1007/s00586-010-1353-6.

C McCarthy, F Arnall, N Strimpakos, A Freemont, and J Oldham. The biopsychosocial classification of nonspecific low back pain: A systematic review. *Physical Therapy*, 9:17–30, 2004. doi: 10.1179/108331904225003955.

Christopher James McCarthy, Christopher Roberts, Matthew Gittins, and Jacqueline Ann Oldham. A process of subgroup identification in non-specific low back pain using a standard clinical examination and cluster analysis. *Physiother. Res. Int.*, 17:92Ű100, 2011.

Charles McCreary. Empirically derived mmpi profile clusters and characteristics of low back pain patients. *Journal of Consulting and Clinical Psychology*, 53:558–560, 1985.

M. McGonigle-Chalmers and I. Kusel. The emergence of linear sequencing children: A continuity account and a formal model. In D. McFarland, K. Stenning,

and M. McGonigle-Chalmers, editors, *The Complex Mind: An Inter-disciplinary Approach*. Palgrave-MacMillan, 2012.

Margaret McGonigle-Chalmers, Kimberly Bodner, Alicia Fox-Pitt, and Laura Nicholson. Size sequencing as a window on executive control in children with autism and aspergerâĂŹs syndrome. *Journal of Autism and Developmental Disorders*, 38(7):1382–1390, 2008. ISSN 0162-3257. doi: 10.1007/s10803-007-0396-y. URL `http://dx.doi.org/10.1007/s10803-007-0396-y`.

Ghada W Mikhail. Coronary heart disease in women. *BMJ*, 331(7515):467–468, 9 2005. doi: 10.1136/bmj.331.7515.467.

Todd D. Miller. The exercise treadmill test: Estimating cardiovascular prognosis. *Cleveland Clinic Journal of Medicine*, 75(6):424–430, 2008.

Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997. ISBN 0070428077, 9780070428072.

Tom M. Mitchell. Machine learning and data mining. *Commun. ACM*, 42(11): 30–36, November 1999. ISSN 0001-0782. doi: 10.1145/319382.319388. URL `http://doi.acm.org/10.1145/319382.319388`.

Lars Morso, Peter M. Kent, and Hanne B. Albert. Are self-reported pain characteristics, classified using the paindetect questionnaire, predictive of outcome in people with low back pain and associated leg pain? *Clin J Pain*, 27(6), 2011.

Ian T. Nabney. *NETLAB. Algorithms for Pattern Recognition*. Advances in Pattern Recognition. Springer, 2002. ISBN 1852334401.

Office for National Statistics. URL `http://www.ons.gov.uk/`.

NECSI Team. About complex systems, 2000. URL `http://www.necsi.edu/guide/study.html`. accessed online 5 June 2012.

Rosalind J. Neuman, Andrew Heath, Wendy Reich, Kathleen K. Bucholz, Pamela A. F. Madden, Lingwei Sun, Richard D. Todd, and James J. Hudziak. Latent class analysis of adhd and comorbid symptoms in a population sample of adolescent female twins. *Journal of Child Psychology and Psychiatry*, 42 (7):933–942, 2001. ISSN 1469-7610. doi: 10.1111/1469-7610.00789. URL `http://dx.doi.org/10.1111/1469-7610.00789`.

NICE. *Low back pain. Early management of persistent non-specific low back pain.* CG88. National Institute for Health and Clinical Excellence, London, UK, 2009.

Michael K. Nicholas. The pain self-efficacy questionnaire: Taking pain into account. *European Journal of Pain*, 11(2):153–163, 2007. ISSN 1532-2149. doi: 10.1016/j.ejpain.2005.12.008. URL http://dx.doi.org/10.1016/j.ejpain.2005.12.008.

Alain J. Nordmann, Peter Hengstler, Thomas Harr, James Young, and Heiner C. Bucher. Clinical outcomes of primary stenting versus balloon angioplasty in patients with myocardial infarction: A meta-analysis of randomized controlled trials. *Cochrane Database Syst Rev.*, 2, 2005. doi: 10.1002/14651858.CD005313. URL http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD005313/frame.ht http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005313/abstract.

Karen L. Nylund, Tihomir Asparouhov, and Bengt O. MuthÃľn. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4):535–569, 2007. doi: 10.1080/10705510701575396.

Ann Oakley, Vicki Strange, Chris Bonell, Elizabeth Allen, and Judith Stephenson. Process evaluation in randomised controlled trials of complex interventions. *BMJ*, 332:413Ű416, 2006. doi: 10.1136/bmj.332.7538.413.

Francesca O'Boyle, Neeraj Mediratta, Brian Fabri, Mark Pullan, John Chalmers, James McShane, Mathew Shaw, and Michael Poullis. Long-term survival after coronary artery bypass surgery stratified by euroscore. *European Journal of Cardio-Thoracic Surgery*, 42(1):101–107, 2012. doi: 10.1093/ejcts/ezr253. URL http://ejcts.oxfordjournals.org/content/42/1/101.abstract.

C M O'Connor, D J Whellan, K L Lee, and et al. Efficacy and safety of exercise training in patients with chronic heart failure: Hf-action randomized controlled trial. *JAMA*, 301(14):1439–1450, 2009. doi: 10.1001/jama.2009.454. URL + http://dx.doi.org/10.1001/jama.2009.454.

B Oerkild, M Frederiksen, JF Hansen, and E Prescott. Home-based cardiac rehabilitation is an attractive alternative to no cardiac rehabilitation for elderly patients with coronary heart disease: results from a randomised clinical trial. *BMJ Open*, 2012.

Nick Parsons, Michelle Odumenya, Antoinette Edwards, Fiona Lecky, and Giles Pattison. Modelling the effects of the weather on admissions to uk trauma units: a cross-sectional study. *Emergency Medicine Journal*, 28, 2011. doi: 10.1136/emj.2010.091058.

Michael J. Pencina, Ralph B. Sr D'Agostino, Martin G. Larson, Joseph M. Massaro, and Ramachandran S. MD Vasan. Predicting the 30-year risk of cardiovascular disease: The framingham heart study. *Circulation*, 119:3078–3084, 2009.

Marco I Perez, Vijaya M Musini, and James M Wright. Effect of early treatment with anti-hypertensive drugs on short and long-term mortality in patients with an acute cardiovascular event. *Cochrane Database of Systematic Reviews*, 2009. doi: 10.1002/14651858.CD006743.pub2. URL http://0-onlinelibrary.wiley.com.pugwash.lib.warwick.ac.uk/doi/10.1002/14651858.CD

Erkki Pesonen, Matti Eskelinen, and Martti Juhola. Treatment of missing data values in a neural network based decision support system for acute abdominal pain. *Artificial Intelligence in Medicine*, 13(3):139 – 146, 1998. ISSN 0933-3657. doi: 10.1016/S0933-3657(98)00027-X. URL http://www.sciencedirect.com/science/article/pii/S093336579800027X.

Tony A. Plate, Joel Bert, John Grace, and Pierre Band. Visualising the function computed by a feedforward neural network. *Neural Computation*, pages 1337–1353, 2000.

Plexus. Applying complexity science to health and healthcare. In *Applying Complexity Science to Health and Healthcare*. Plexus Institute www.PlexusInstitute.org, 2003.

Paul E Plsek and Trisha Greenhalgh. The challenge of complexity in health care. *BMJ*, 323:625–8, 2001.

Paul E Plsek and Tim Wilson. Complexity, leadership, and management in healthcare organisations. *BMJ*, 323:746Ű9, 2001.

Catrine Edström Plüss, M Rydell Karlsson, NH Wallen, E. Billing, and C. Held. Effects of an expanded cardiac rehabilitation programme in patients treated for an acute myocardial infarction or a coronary artery by-pass graft operation. *Clinical Rehabilitation*, 22(4):306–318, 2008. doi: 10.1177/0269215507085379. URL http://cre.sagepub.com/content/22/4/306.abstract.

Catrine Edström Plüss, Ewa Billing, Claes Held, Peter Henriksson, Anna Kiessling, Monica Rydell Karlsson, and Håkan N Wallen. Long-term effects of an expanded cardiac rehabilitation programme after myocardial infarction or coronary artery bypass surgery: a five-year follow-up of a randomized controlled study. *Clinical Rehabilitation*, 25(1):79–87, 2011. URL http://www.ncbi.nlm.nih.gov/pubmed/20702510.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL http://www.R-project.org. ISBN 3-900051-07-0.

Parwis B. Rahmanian, David H. Adams, Javier G. Castillo, Joanna Chikwe, Carol A. Bodian, and Farzan Filsoufi. Impact of body mass index on early outcome and late survival in patients undergoing coronary artery bypass grafting or valve surgery or both. *The American Journal of Cardiology*, 100(11): 1702 – 1708, 2007. ISSN 0002-9149. doi: 10.1016/j.amjcard.2007.07.017. URL http://www.sciencedirect.com/science/article/pii/S0002914907016244.

Aline Ramond, Céline Bouton, Isabelle Richard, Yves Roquelaure, Christophe Baufreton, Erick Legrand, and Jean-François Huez. Psychosocial risk factors for chronic low back pain in primary careŮa systematic review. *Family Practice*, 28 (1):12–21, 2011. doi: 10.1093/fampra/cmq072.

Peter Ravdin and Gary Clark. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*, 22:285–293, 1992. ISSN 0167-6806. URL http://dx.doi.org/10.1007/BF01840841. 10.1007/BF01840841.

Beth A. Reboussin, Eun-Young Song, Anshu Shrestha, Kurt K. Lohman, and Mark Wolfson. A latent class analysis of underage problem drinking: Evidence from a community sample of 16Ű20 year olds. *Drug and Alcohol Dependence*, 83(3): 199 – 209, 2006. ISSN 0376-8716. doi: 10.1016/j.drugalcdep.2005.11.013. URL http://www.sciencedirect.com/science/article/pii/S0376871605003509.

Colin Reeves and Charles Johnston. Neural networks for estimating probability distributions for survival analysis. In Keith Burnham and Olivier Haas, editors, *Intelligent and Adaptive Systems in Medicine*, Series in Medical Physics and Biomedical Engineering, pages 259–280. Taylor & Francis, U.K., 2008. doi: 10.1201/9781420012088.ch8. URL http://dx.doi.org/10.1201/9781420012088.ch8.

Olivier Regnier-Coudert, John McCall, Robert Lothian, Thomas Lam, Sam McClinton, and James NŠDow. Machine learning for improved pathological staging of prostate cancer: A performance comparison on a range of classifiers. *Artificial Intelligence in Medicine*, 55(1):25 – 35, 2012. ISSN 0933-3657. doi: 10.1016/j.artmed.2011.11.003. URL http://www.sciencedirect.com/science/article/pii/S0933365711001461.

Giuseppe Rengo, Gennaro Galasso, Dino F. Vitale, Giuseppe Furgi, Carmela Zincarelli, Luca Golino, Grazia D. Femminella, Federico Piscione, Franco Rengo, and Dario Leosco. An active lifestyle prior to coronary surgery is associated with improved survival in elderly patients. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 65A(7):758–763, 2010. doi: 10.1093/gerona/glp216. URL `http://biomedgerontology.oxfordjournals.org/content/65A/7/758.abstract`.

Brian D. Ripley and Ruth M. Ripley. Neural networks as statistical methods in survival analysis. In *Clinical Applications of Artificial Neural Networks*, pages 298–326. University Press, 2001.

Ruth Ripley. *Neural Network Models for Breast Cancer Prognosis*. PhD thesis, St Cross College, Univeristy of Oxford, England, Trinity Term 1998.

Ruth M. Ripley, Adrian L. Harris, and Lionel Tarassenko. Non-linear survival analysis using neural networks. *Statistics in Medicine*, 23(5):825–842, 2004. ISSN 1097-0258. doi: 10.1002/sim.1655. URL `http://dx.doi.org/10.1002/sim.1655`.

Jeanine E Roeters van Lennep, H.Tineke Westerveld, D.Willem Erkelens, and Ernst E van der Wall. Risk factors for coronary heart disease: implications of gender. *Cardiovascular Research*, 53 (3):538–549, 2002. doi: 10.1016/S0008-6363(01)00388-1. URL `http://cardiovascres.oxfordjournals.org/content/53/3/538.abstract`.

LUIS A. Rohde, GENÁRIO BARBOSA, GUILHERME POLANCZYK, MARIANA EIZIRIK, ERIK R. RASMUSSEN, ROSALIND J. NEUMAN, and RICHARD D. TODD. Factor and latent class analysis of dsm-iv adhd symptoms in a school sample of brazilian adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(6):711 – 718, 2001. ISSN 0890-8567. doi: 10.1097/00004583-200106000-00017.

Martin Roland and Jeremy Fairbank. The roland-morris disability questionnaire and the oswestry disability questionnaire. *Spine*, 2000.

MO Roland and RW Morris. A study of the natural history of back pain. part 1: Development of a reliable and sensitive measure of disability in low back pain. *Spine*, 1983.

Abel Romero-Corral, Victor M Montori, Virend K Somers, Josef Korinek, Randal J Thomas, Thomas G Allison, Farouk Mookadam, and Francisco Lopez-

Jimenez. Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *The Lancet*, 368(9536):666 – 678, 2006. ISSN 0167-5273. URL http://linkinghub.elsevier.com/retrieve/pii/S0140673606692519.

Donna Ryan. Risks and benefits of weight loss: challenges to obesity research. *European Heart Journal Supplements*, 7(suppl L):L27–L31, 2005. doi: 10.1093/eurheartj/sui083.

Emad W. Saad and Donald C. WunschII. Neural network explanation using inversion. *Neural Networks*, 20(1):78 – 93, 2007. ISSN 0893-6080. doi: 10.1016/j.neunet.2006.07.005. URL http://www.sciencedirect.com/science/article/pii/S0893608006001730.

Neal W. Sanders and N. Horace Mann. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. *Computers in Biology and Medicine*, 30:287–298, 2000.

Patrick D. Savage, MaryEllen Antkowiak, and Philip A. Ades. Failure to improve cardiopulmonary fitness in cardiac rehabilitation. *Journal of Cardiopulmonary Rehabilitation and Prevention*, 29, 2009.

Peter Scarborough, Prachi Bhatnagar, Kremlin Wickramasinghe, Kate Smolina, Colin Mitchell, and Mike Rayner. *Coronary Heart Disease Statistics, 2010 edition.* Department of Public Health, University of Oxford, Oxford, UK, 2010.

Friederike Scharhag-Rosenberger, Tim Meyer, Nina GäSSler, Oliver Faude, and Wilfried Kindermann. Exercise at given percentages of vo2max: Heterogeneous metabolic responses between individuals. *Journal of science and medicine in sport / Sports Medicine Australia*, 13:74–79, 2010. doi: 10.1016/j.jsams.2008.12.626.

Norbert Schmitz, Johannes Kruse, and Joachim Kugler. Disabilities, quality of life, and mental disorders associated with smoking and nicotine dependence. *Am J Psychiatry*, 160:1670Ű1676, 2003.

D. Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982. doi: 10.1093/biomet/69.1.239. URL http://biomet.oxfordjournals.org/content/69/1/239.abstract.

Gideon Schwarz. *Annals of Statistics*, 6, 1978.

David M. Shahian, Sean M. O'Brien, Shubin Sheng, Frederick L. Grover, John E. Mayer, Jeffrey P. Jacobs, Jocelyn M. Weiss, Elizabeth R. DeLong, Eric D. Peterson, William S. Weintraub, Maria V. Grau-Sepulveda, Lloyd W. Klein, Richard E. Shaw, Kirk N. Garratt, Issam D. Moussa, Cynthia M. Shewan, George D. Dangas, and Fred H. Edwards. Predictors of long-term survival after coronary artery bypass grafting surgery / clinical perspective. *Circulation*, 125(12):1491–1500, 2012. doi: 10.1161/CIRCULATIONAHA.111.066902. URL `http://circ.ahajournals.org/content/125/12/1491.abstract`.

Alan Shiell, Penelope Hawe, and Lisa Gold. Complex interventions or complex systems? implications for health economic evaluation. *BMJ*, 336(7656):1281–1283, 6 2008. doi: 10.1136/bmj.39569.510521.AD.

Michael S Shutty Jr. and Douglas E DeGood. Cluster analyses of responses of low back pain patients to the scl-90: Comparison of empirical versus rationally derived subscales. *Rehabilitation Psychology*, 32, 1987.

B.N. Singh. Increased heart rate as a risk factor for cardiovascular disease. *European Heart Journal Supplements*, Supplement G(5):G3ŮG9, 2003.

Jane S Skinner and Angela Cooper. Secondary prevention of ischaemic cardiac events. *BMJ: Clinical Evidence*, 08:206, 2011.

Thorkild I.A Sorensen, Aila Rissanen, Maarit Korkeila, and Jaakko Kaprio. Intention to lose weight, weight changes, and 18-y mortality in overweight individuals without co-morbidities. *PLoS Med*, 2(6):e171, 06 2005. doi: 10.1371/journal.pmed.0020171.

B. D. Spycher, M. Silverman, A. M. Brooke, C. E. Minder, and C. E. Kuehni. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *European Respiratory Journal*, 31(5):974–981, 2008. doi: 10.1183/09031936.00153507. URL `http://erj.ersjournals.com/content/31/5/974.abstract`.

Back screening tool STarT, accessed 18/05/13. URL `http://www.keele.ac.uk/sbst/clinicalpartners/`.

P.F. Sullivan, W. Smith, and D. Buchwald. Latent class analysis of symptoms associated with chronic fatigue syndrome and fibromyalgia. *Psychological Medicine*, 32(05):881–888, 2002.

. Surgeon General. *The Health Consequences of Smoking: A Report of the Surgeon General*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA, 2004.

A. Taktak, L. Antolini, M. Aung, P. Boracchi, I. Campbell, B. Damato andE. Ifeachor, N. Lama, P. Lisboa, C. Setzkorn, V. Stalbovskaya, and E. Biganzoli. Double-blind evaluation and benchmarking of survival models in a multi-centre study. *Comp. Biol. Med.*, 2006. doi: 10.1016/j.compbiomed.2006.10.001.

Azzam Taktak, Antonio Eleuteri, Min Aung, Paulo Lisboa, Laurence Desjardins, and Bertil Damato. External validation of a bayesian neural network model in survival analysis. In *Proceedings of Seventh IEEE International Conference on Machine Learning and Applications*, 2008.

Navdeep Tangri, David Ansell, and David Naimark. Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. *Nephrol Dial Transplant*, page 2972Ű2981, 2008. doi: 10.1093/ndt/gfn187.

Rod S Taylor, Allan Brown, Shah Ebrahim, Judith Jolliffe, Hussein Noorani, Karen Rees, Becky Skidmore, James A Stone, David R Thompson, and Neil Oldridge. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. *The American journal of medicine*, 2004. doi: 10.1016/j.amjmed.2004.01.009.

NECSI The New England Complex Systems Institute. About complex systems: Overview: what is the study of complex systems?, Accesed May 2013. URL `http://www.necsi.edu/guide/study.html`. Accessed 09/05/2013.

T. M. Therneau, P.M Grambsch, and T.R. Fleming. Martingale-based residuals for survivla models. *Biometrika*, 77(1), 1990.

T.M. Therneau and P.M. Grambsch. *Modelling survival data: extending the Cox model*. Springer, New York, 2000.

BD Thombs, P de Jonge, JC Coyne, and et al. Depression screening and patient outcomes in cardiovascular care: A systematic review. *JAMA*, 300(18):2161–2171, 2008. doi: 10.1001/jama.2008.667.

Margaret Thorogood, Melvyn Hillsdon, and Carolyn Summerbell. Changing behaviour. *Clinical Evidence*, 08, 2006.

Kristian Thygesen, Joseph S. Alpert, Allan S. Jaffe, Maarten L. Simoons, Bernard R. Chaitman, and Harvey D. White. Third universal definition of myocardial infarction. *Circulation*, 126(16):2020–2035, 2012. doi: 10.1161/CIR.0b013e31826e1058. URL http://circ.ahajournals.org/content/126/16/2020.short.

Stefan Topolski. Understanding health from a complex systems perspective. *Journal of Evaluation in Clinical Practice*, 15, 2009.

Sally C. Turner. *Relationship of Physical Fitness, Depression and Mortality in a Cardiac Rehabilitation Cohort*. PhD thesis, University of Warwick, U.K., 2007.

Sally C. Turner, Julia A. Evans, Hugh J.N. Bethell, Jonathan Goddard, and Mark A. Mullee. Patient characteristics and outcomes of cardiac rehabilitation. *Journal of Cardiopulmonary Rehabilitation*, 22(4):253–260, 2002. ISSN 0883-9212.

Sally C. Turner, Julia A. Evans, Hugh J.N. Bethell, and Jonathan Goddard. Psychological assessments for cardiac rehabilitation patients. *International Journal of Cardiology*, 92(2-3):201 – 207, 2003. ISSN 0167-5273. doi: 10.1016/S0167-5273(03)00086-X. URL http://www.sciencedirect.com/science/article/pii/S016752730300086X.

Liz Twigg, Graham Moon, and Sarah Walker. *The smoking epidemic in England*. NHS Health Development Agency, Holborn Gate, 330 High Holborn, London WC1V 7BA, 2004.

Belgin Unal, Julia Alison Critchley, and Simon Capewell. Explaining the decline in coronary heart disease mortality in england and wales between 1981 and 2000. *Circulation*, 109(9):1101–1107, 2004. doi: 10.1161/01.CIR.0000118498.35499.B2. URL http://circ.ahajournals.org/content/109/9/1101.abstract.

Martin Underwood, Dipesh Mistry, Ranjit Lall, and Sallie Lamb. Predicting response to a cognitiveâĂŞbehavioral approach to treating low back pain: Secondary analysis of the best data set. *Arthritis Care & Research*, 63 (9):1271–1279, 2011. ISSN 2151-4658. doi: 10.1002/acr.20518. URL http://dx.doi.org/10.1002/acr.20518.

Noah Vale, Alain J Nordmann, Gregory G Schwartz, James de Lemos, Frank Colivicchi, Furio den Hartog, Petr Ostadal, Stella M Macin, Anho H Liem, Edward Mills, Neera Bhatnagar, Heiner C Bucher, and Matthias Briel. Statins for acute coronary syndrome. *Cochrane Database of Systematic Reviews*, 2011a. doi: 10.1002/14651858.CD006870.pub2. URL http://0-onlinelibrary.wiley.com.pugwash.lib.warwick.ac.uk/doi/10.1002/14651858.CD

Noah Vale, Alain J Nordmann, Gregory G Schwartz, James de Lemos, Furio Colivicchi, Frank den Hartog, Petr Ostadal, Stella M Macin, Anho H Liem, Edward Mills, Neera Bhatnagar, Heiner C Bucher, and Matthias Briel1. Statins for acute coronary syndrome. *Cochrane Database of Systematic Reviews*, 6, 2011b. ISSN 0167-5273. doi: 10.1002/14651858.CD006870.pub2. URL http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD006870/frame.ht http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006870.pub2/abstract.

Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011. URL http://www.jstatsoft.org/v45/i03/.

JOOST P. van Melle, PETER DE JONGE, TITIA A. SPIJKERMAN, JAN G. P. TIJSSEN, JOHAN ORMEL, DIRK J. VAN VELDHUISEN, ROB H. S. VAN DEN BRINK, and MAARTEN P. VAN DEN BERG. Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: A meta-analysis. *Psychosom Med.*, 66:814–22, 2004.

Albert H.M. van Straten, Sander Bramer, Mohamed A. Soliman Hamad, André A.J. van Zundert, Elisabeth J. Martens, Jacques P.A.M. Schönberger, and Andre M. de Wolf. Effect of body mass index on early and late mortality after coronary artery bypass grafting. *The Annals of Thoracic Surgery*, 89(1): 30 – 37, 2010. ISSN 0003-4975. doi: 10.1016/j.athoracsur.2009.09.050. URL http://www.sciencedirect.com/science/article/pii/S0003497509019572.

Luc Vanhees, Robert Fagard, Lutgarde Thijs, and Antoon Amery. Prognostic value of training-induced change in peak exercise capacity in patients with myocardial infarcts and patients with coronary bypass surgery. *The American journal of cardiology*, 76(14):1014–1019, 1995.

W. N. Venables and Brian D. Ripley. *Modern applied statistics with S.* Springer, New York, 4th edition, August 2002. ISBN 0387954570. URL http://www.stats.ox.ac.uk/pub/MASS4. ISBN 0-387-95457-0.

Konstantinos Veropoulos. *Machine learning approaches to medical decision making.* PhD thesis, University of Bristol, UK, May 2001.

Johan W.S. Vlaeyen and Steven J. Linton. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3):317 – 332, 2000. ISSN 0304-3959. doi: 10.1016/S0304-3959(99)00242-0. URL http://www.sciencedirect.com/science/article/pii/S0304395999002420.

Stephan von Haehling, Wolfram Doehner, and Stefan D. Anker. Revisiting the obesity paradox in heart failure: new insights? *European Journal of Heart Failure*, 13(2):130–132, 2011. doi: 10.1093/eurjhf/hfq237. URL http://eurjhf.oxfordjournals.org/content/13/2/130.short.

Martha E. Wadsworth, James J. Hudzaik, Andrew C. Heath, and Thomas M. Achenbach. Latent class analysis of child behavior checklist anxiety/depression in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(1):106 – 114, 2001. ISSN 0890-8567. doi: 10.1097/00004583-200101000-00023. URL http://www.sciencedirect.com/science/article/pii/S0890856709608224.

W Wang, D Partridge, and P Jones. A ciomparative study of feature-salience ranking techniques. *Neural Computing & Applications*, pages 101–112, 2000.

John E. Jr. Ware, Mark Kosinski, and Susan D. Keller. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3):220–233, 1996.

CH Weiss, F Moazed, CA McEvoy, BD Singer, I Szleifer, LA Amaral, M Kwasny, CM Watts, SD Persell, DW Baker, JI Sznajder, and RG Wunderink. Prompting physicians to address a daily checklist and process of care and clinical outcomes: a single-site study. *Am J Respir Crit Care Med*, 184:680 – 686, 2011.

Robert R West, Dee A Jones, and Andrew H Henderson. Rehabilitation after myocardial infarction trial (ramit): multi-centre randomised controlled trial of comprehensive cardiac rehabilitation in patients following acute myocardial infarction. *Heart*, 2011. doi: 10.1136/heartjnl-2011-300302.

Ben Whalley, Karen Rees, Philippa Davies, Paul Bennett, Shah Ebrahim, Zulian Liu, Robert West, Tiffany Moxham, David R Thompson, and Rod S Taylor. Psychological interventions for coronary heart disease. *Cochrane Database of Systematic Reviews*, 2011. doi: 10.1002/14651858.CD002902.pub3. URL http://0-onlinelibrary.wiley.com.pugwash.lib.warwick.ac.uk/doi/10.1002/14651858.CD

David G T Whitehurst, Stirling Bryan, Martyn Lewis andJonathan Hill, and Elaine M Hay. Exploring the costŰutility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. *Ann Rheum Dis*, 71:1796Ű1802, 2012. doi: 10.1136/annrheumdis-2011-200731.

WC Willett, JE Manson, MJ Stampfer, and et al. Weight, weight change, and coronary heart disease in women: Risk within the normal weight range. *JAMA*, 273(6):461–465, 1995. doi: 10.1001/jama.1995.03520300035033. URL + http://dx.doi.org/10.1001/jama.1995.03520300035033.

Paula Williamson, Douglas Altman, Jane Blazeby, Michael Clarke, and Elizabeth Gargon. Driving up the quality and relevance of research through the use of agreed core outcomes. *Journal of Health Services Research & Policy*, 17(1):1–2, 2012. doi: 10.1258/jhsrp.2011.011131. URL http://hsr.sagepub.com/content/17/1/1.short.

D. Randall Wilson and Tony R. Martinez. The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16(10):1429–1451, 2003.

Tim Wilson and Tim Holt. Complexity and clinical care. *BMJ*, 323:685Ű8, 2001.

Y. Wu, K.and Vyborny C.J. Giger, M.L.and Doi, R.A. Schmidt, and C.E. Metz. 1993.

Chih-Chien Yang. Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50(4):1090 – 1104, 2006. ISSN 0167-9473. doi: 10.1016/j.csda.2004.11.004.

Javier Zamora, Francisco Kovacs, VÃŋctor Abraira, Carmen FernÃąndez, and Pablo LÃązaro. The social tariff of eq-5d is not adequate to assess quality of life in patients with low back pain. *Quality of Life Research*, 16: 523–531, 2007. ISSN 0962-9343. doi: 10.1007/s11136-006-9140-x. URL http://dx.doi.org/10.1007/s11136-006-9140-x.

Chapter 9

Appendix

9.1 Publications

9.1.1 Published

Martine J. Barons, Nick Parsons, Frances Griffiths, and Margaret Thorogood. A comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioural approach to treatment for non-specific low back pain. In *IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE) 2013*, pages 7-11, 2013.

9.1.2 Submitted

Martine J. Barons, Frances E. Griffiths, Nick Parsons, Anca Alba, Margaret Thorogood, Graham F. Medley, and Sallie Lamb. Matching patients to an intervention for back pain: classifying patients using a latent class approach. *Journal of Evaluation in Clinical Practice*, SUBMITTED, 2013a.

9.2 Diagnostic plots

9.2.1 Survival models with imputed data

Comparison of optimised all-cause survival models built on imputed data sets 1 to 5						
Variable		**Hazard Ratios**				
Inputed Data set		1	2	3	4	5
Age category	under 50	1	1	1	1	1
	50-59	2.02 ***	2.02 ***	2.03 ***	2.02 ***	2.02 ***
	60-69	3.08 ***	3.07 ***	3.09 ***	3.07 ***	3.08 ***
	70+	6.06 ***	6.06 ***	6.07 ***	6.04 ***	6.07 ***
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	1.76 ***	1.76 ***	1.76 ***	1.76 ***	1.76 ***
	Low baseline	2.75 ***	2.76 ***	2.76 ***	2.76 ***	2.76 ***
D'Hoore Co-morbidity score	None	1	1	1	1	1
	1 (least)	1.17	1.17	1.17	1.17	1.17
	2	1.48 ***	1.49 ***	1.48 ***	1.48 ***	1.49 ***
	3	1.87 ***	1.87 ***	1.87 ***	1.87 ***	1.87 ***
	4 (most)	2.67 ***	2.68 ***	2.67 ***	2.67 ***	2.68 ***
Statins	Yes	0.74 ***	0.74 ***	0.74 ***	0.74 ***	0.74 ***
	No	1	1	1	1	1
Aspirin	Yes	0.62 ***	0.63 ***	0.63 ***	0.63 ***	0.62 ***
	No	1	1	1	1	1
Diagnostic Category	MI	1	1	1	1	1
	CABG	0.69 ***	0.69 ***	0.69 ***	0.69 ***	0.69 ***
	PCI	0.68 *	0.68 *	0.68 *	0.68 *	0.68 *
	MI + PCI	0.79	0.78	0.79	0.79	0.79
	Angina	0.83	0.83	0.82	0.83	0.82
	Other cardiac	0.94	0.94	0.94	0.94	0.94
Sex	Male	1	1	1	1	1
	Female	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
Systolic blood pressure before		0.998	0.997	0.998	0.998	0.998
Ace inhibitor	Yes	1.18 *	1.19 *	1.19 *	1.19 *	1.19 *
	No	1	1	1	1	1
Resting heart rate		1.006 **	1.006 **	1.006 **	1.006 **	1.006 **

Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Table 9.1: Data sets with imputed values, 2714 cases. All-cause survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.16.

Comparison of optimised all-cause survival models built on imputed data sets 6 to 10						
Variable		**Hazard Ratios**				
Inputed Data set		6	7	8	9	10
Age category	under 50	1	1	1	1	1
	50-59	2.02 ***	2.02 ***	2.02 ***	2.03 ***	2.03 ***
	60-69	3.07 ***	3.08 ***	3.08 ***	3.08 ***	3.09 ***
	70+	6.05 ***	6.07 ***	6.08 ***	6.04 ***	6.08 ***
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	1.76 ***	1.76 ***	1.76 ***	1.76 ***	1.76 ***
	Low baseline	2.75 ***	2.76 ***	2.76 ***	2.76 ***	2.76 ***
D'Hoore Co-morbidity score	None	1	1	1	1	1
	1 (least)	1.17	1.17	1.17	1.17	1.17
	2	1.48 ***	1.48 ***	1.48 ***	1.48 ***	1.49 ***
	3	1.87 ***	1.87 ***	1.87 ***	1.87 ***	1.87 ***
	4 (most)	2.67 ***	2.68 ***	2.67 ***	2.67 ***	2.68 ***
Statins	Yes	0.74 ***	0.74 ***	0.74 ***	0.74 ***	0.74 ***
	No	1	1	1	1	1
Aspirin	Yes	0.63 ***	0.63 ***	0.63 ***	0.63 ***	0.63 ***
	No	1	1	1	1	1
Diagnostic Category	MI	1	1	1	1	1
	CABG	0.69 ***	0.69 ***	0.69 ***	0.69 ***	0.69 ***
	PCI	0.68 *	0.68 *	0.68 *	0.68 *	0.68 *
	MI + PCI	0.79	0.78	0.79	0.79	0.79
	Angina	0.83	0.83	0.83	0.83	0.83
	Other cardiac	0.94	0.94	0.94	0.94	0.94
Sex	Male	1	1	1	1	1
	Female	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
Systolic blood pressure before		0.998	0.997	0.998	0.998	0.998
Ace inhibitor	Yes	1.19 *	1.19 *	1.19 *	1.19 *	1.19 *
	No	1	1	1	1	1
Resting heart rate		1.006 **	1.006 **	1.006 **	1.006 **	1.006 **

Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Table 9.2: Data sets with imputed values, 2714 cases. All-cause survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.16.

Comparison of optimised all-cause survival models built on imputed data sets 11 to 15						
Variable		**Hazard Ratios**				
Inputed Data set		11	12	13	14	15
Age category	under 50	1	1	1	1	1
	50-59	2.03 ***	2.02 ***	2.02 ***	2.02 ***	2.02 ***
	60-69	3.08 ***	3.08 ***	3.08 ***	3.07 ***	3.08 ***
	70+	6.07 ***	6.07 ***	6.06 ***	6.05 ***	6.06 ***
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	1.76 ***	1.76 ***	1.76 ***	1.76 ***	1.76 ***
	Low baseline	2.76 ***	2.76 ***	2.76 ***	2.75 ***	2.75 ***
D'Hoore Co-morbidity score	None	1	1	1	1	1
	1 (least)	1.17	1.17	1.17	1.17	1.17
	2	1.48 ***	1.49 ***	1.48 ***	1.48 ***	1.49 ***
	3	1.87 ***	1.87 ***	1.87 ***	1.87 ***	1.87 ***
	4 (most)	2.67 ***	2.67 ***	2.67 ***	2.67 ***	2.67 ***
Statins	Yes	0.74 ***	0.74 ***	0.74 ***	0.74 ***	0.74 ***
	No	1	1	1	1	1
Aspirin	Yes	0.63 ***	0.63 ***	0.63 ***	0.63 ***	0.63 ***
	No	1	1	1	1	1
Diagnostic Category	MI	1	1	1	1	1
	CABG	0.69 ***	0.69 ***	0.69 ***	0.69 ***	0.69 ***
	PCI	0.68 *	0.68 *	0.68 *	0.68 *	0.68 *
	MI + PCI	0.79	0.79	0.79	0.79	0.79
	Angina	0.83	0.83	0.83	0.83	0.82
	Other cardiac	0.94	0.94	0.94	0.94	0.94
Sex	Male	1	1	1	1	1
	Female	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
Systolic blood pressure before		0.998	0.997	0.998	0.998	0.998
Ace inhibitor	Yes	1.18 *	1.19 *	1.19 *	1.19 *	1.19 *
	No	1	1	1	1	1
Resting heart rate		1.006 **	1.006 **	1.006 **	1.006 **	1.006 **

Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Table 9.3: Data sets with imputed values, 2714 cases. All-cause survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.16.

Comparison of optimised all-cause survival models built on imputed data sets 16 to 20						
Variable		**Hazard Ratios**				
Inputed Data set		16	17	18	19	20
Age category	under 50	1	1	1	1	1
	50-59	2.03 ***	2.02 ***	2.02 ***	2.03 ***	2.03 ***
	60-69	3.08 ***	3.08 ***	3.08 ***	3.09 ***	3.08 ***
	70+	6.08 ***	6.05 ***	6.07 ***	6.07 ***	6.07 ***
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	1.76 ***	1.76 ***	1.76 ***	1.76 ***	1.76 ***
	Low baseline	2.76 ***	2.76 ***	2.76 ***	2.76 ***	2.76 ***
D'Hoore Co-morbidity score	None	1	1	1	1	1
	1 (least)	1.17	1.17	1.17	1.17	1.17
	2	1.48 ***	1.48 ***	1.49 ***	1.48 ***	1.48 ***
	3	1.87 ***	1.87 ***	1.87 ***	1.87 ***	1.87 ***
	4 (most)	2.68 ***	2.67 ***	2.68 ***	2.67 ***	2.67 ***
Statins	Yes	0.74 ***	0.74 ***	0.74 ***	0.74 ***	0.74 ***
	No	1	1	1	1	1
Aspirin	Yes	0.63 ***	0.63 ***	0.63 ***	0.63 ***	0.62 ***
	No	1	1	1	1	1
Diagnostic Category	MI	1	1	1	1	1
	CABG	0.69 ***	0.69 ***	0.69 ***	0.69 ***	0.69 ***
	PCI	0.69 *	0.68 *	0.68 *	0.68 *	0.68 *
	MI + PCI	0.79	0.79	0.79	0.79	0.79
	Angina	0.83	0.83	0.82	0.83	0.82
	Other cardiac	0.94	0.94	0.94	0.94	0.94
Sex	Male	1	1	1	1	1
	Female	0.62 ***	0.62 ***	0.62 ***	0.63 ***	0.62 ***
Systolic blood pressure before		0.998	0.997	0.998	0.998	0.998
Ace inhibitor	Yes	1.19 *	1.19 *	1.19 *	1.19 *	1.19 *
	No	1	1	1	1	1
Resting heart rate		1.006 **	1.006 **	1.006 **	1.006 **	1.006 **

Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Table 9.4: Data sets with imputed values, 2714 cases. All-cause survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.16.

Comparison of optimised cardiovascular survival models built on imputed data sets 1 to 5						
Variable		**Hazard Ratios**				
Inputed Data set		1	2	3	4	5
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	2.17 ***	2.17 ***	2.17 ***	2.17 ***	2.17 ***
	Low baseline	4.13 ***	4.13 ***	4.13 ***	4.13 ***	4.13 ***
Statins	Yes	0.52 ***	0.53 ***	0.52 ***	0.52 ***	0.52 ***
	No	1	1	1	1	1
Age category	under 50	1	1	1	1	1
	50-59	1.53 .	1.53 .	1.53 .	1.53 .	1.53 .
	60-69	2.42 ***	2.42 ***	2.42 ***	2.42 ***	2.42 ***
	70+	3.83 ***	3.83 ***	3.83 ***	3.83 ***	3.83 ***
Diagnostic						
Category	MI	1	1	1	1	1
	CABG	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
	PCI	0.49 **	0.49 **	0.49 **	0.49 **	0.49 **
	MI + PCI	0.73	0.73	0.73	0.73	0.73
	Angina	0.71 .	0.71 .	0.71 .	0.71 .	0.71 .
	Other cardiac	1.004	1.004	1.004	1.004	1.004
Aspirin	Yes	0.57 ***	0.57 ***	0.57 ***	0.57 ***	0.57 ***
	No	1	1	1	1	1
Sex	Male	1	1	1	1	1
	Female	0.53 ***	0.53 ***	0.53 ***	0.53 ***	0.53 ***
Ace inhibitor	Yes	1.37 **	1.37 **	1.37 **	1.37 **	1.37 **
	No	1	1	1	1	1
Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**						

Table 9.5: Data sets with imputed values, 2714 cases. Cardiovascular survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.17.

Comparison of optimised cardiovascular survival models built on imputed data sets 6 to 10						
Variable		**Hazard Ratios**				
Inputed Data set		6	7	8	9	10
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	2.17 ***	2.17 ***	2.17 ***	2.17 ***	2.17 ***
	Low baseline	4.13 ***	4.13 ***	4.12 ***	4.12 ***	4.13 ***
Statins	Yes	0.52 ***	0.52 ***	0.52 ***	0.52 ***	0.52 ***
	No	1	1	1	1	1
Age category	under 50	1	1	1	1	1
	50-59	1.53 .	1.53 .	1.53 .	1.53 .	1.53 .
	60-69	2.42 ***	2.42 ***	2.42 ***	2.42 ***	2.42 ***
	70+	3.83 ***	3.83 ***	3.81 ***	3.81 ***	3.83 ***
Diagnostic						
Category	MI	1	1	1	1	1
	CABG	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
	PCI	0.49 **	0.49 **	0.49 **	0.49 **	0.49 **
	MI + PCI	0.73	0.73	0.73	0.73	0.73
	Angina	0.71 .	0.71 .	0.71 .	0.71 .	0.71 .
	Other cardiac	1.004	1.004	1.009	1.009	1.004
Aspirin	Yes	0.57 ***	0.57 ***	0.58 ***	0.57 ***	0.57 ***
	No	1	1	1	1	1
Sex	Male	1	1	1	1	1
	Female	0.53 ***	0.53 ***	0.53 ***	0.53 ***	0.53 ***
Ace inhibitor	Yes	1.37 **	1.37 **	1.38 **	1.38 **	1.37 **
	No	1	1	1	1	1
Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**						

Table 9.6: Data sets with imputed values, 2714 cases. Cardiovascular survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.17.

Comparison of optimised cardiovascular survival models built on imputed data sets 11 to 15.						
Variable		**Hazard Ratios**				
Inputed Data set		11	12	13	14	15
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	2.17 ***	2.17 ***	2.17 ***	2.17 ***	2.17 ***
	Low baseline	4.13 ***	4.13 ***	4.13 ***	4.12 ***	4.13 ***
Statins	Yes	0.52 ***	0.53 ***	0.52 ***	0.52 ***	0.52 ***
	No	1	1	1	1	1
Age category	under 50	1	1	1	1	1
	50-59	1.53 .	1.53 .	1.53 .	1.53 .	1.53 .
	60-69	2.42 ***	2.42 ***	2.42 ***	2.42 ***	2.42 ***
	70+	3.83 ***	3.83 ***	3.83 ***	3.81 ***	3.83 ***
Diagnostic						
Category	MI	1	1	1	1	1
	CABG	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
	PCI	0.49 **	0.49 **	0.49 **	0.49 **	0.49 **
	MI + PCI	0.73	0.73	0.73	0.73	0.73
	Angina	0.71 .	0.71 .	0.71 .	0.71 .	0.71 .
	Other cardiac	1.004	1.004	1.004	1.009	1.004
Aspirin	Yes	0.57 ***	0.57 ***	0.57 ***	0.58 ***	0.57 ***
	No	1	1	1	1	1
Sex	Male	1	1	1	1	1
	Female	0.53 ***	0.53 ***	0.53 ***	0.53 ***	0.53 ***
Ace inhibitor	Yes	1.37 **	1.37 **	1.37 **	1.38 **	1.37 **
	No	1	1	1	1	1
Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**						

Table 9.7: Data sets with imputed values, 2714 cases. Cardiovascular survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.17.

Comparison of optimised cardiovascular survival models built on imputed data sets 16 to 20.						
Variable		**Hazard Ratios**				
Inputed Data set		16	17	18	19	20
Fitness:						
	High baseline	1	1	1	1	1
	Mid baseline	2.17 ***	2.17 ***	2.17 ***	2.17 ***	2.17 ***
	Low baseline	4.12 ***	4.12 ***	4.13 ***	4.13 ***	4.13 ***
Statins	Yes	0.52 ***	0.52 ***	0.52 ***	0.52 ***	0.52 ***
	No	1	1	1	1	1
Age category	under 50	1	1	1	1	1
	50-59	1.53 .	1.53 .	1.53 .	1.53 .	1.53 .
	60-69	2.42 ***	2.42 ***	2.42 ***	2.42 ***	2.42 ***
	70+	3.81 ***	3.81 ***	3.83 ***	3.83 ***	3.83 ***
Diagnostic						
Category	MI	1	1	1	1	1
	CABG	0.62 ***	0.62 ***	0.62 ***	0.62 ***	0.62 ***
	PCI	0.49 **	0.49 **	0.49 **	0.49 **	0.49 **
	MI + PCI	0.73	0.73	0.73	0.73	0.73
	Angina	0.71 .	0.71 .	0.71 .	0.71 .	0.71 .
	Other cardiac	1.009	1.009	1.004	1.004	1.004
Aspirin	Yes	0.58 ***	0.58 ***	0.57 ***	0.57 ***	0.57 ***
	No	1	1	1	1	1
Sex	Male	1	1	1	1	1
	Female	0.53 ***	0.53 ***	0.53 ***	0.53 ***	0.53 ***
Ace inhibitor	Yes	1.38 **	1.38 **	1.37 **	1.37 **	1.37 **
	No	1	1	1	1	1
Pr(>Chi) Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**						

Table 9.8: Data sets with imputed values, 2714 cases. Cardiovascular survival model, showing values of model coefficients to compare with complete-cases and pooled estimates in Table 6.17.

9.2.2 Plots of the multiply imputed variable distributions

For key variables with substantial proportion of missing data, the distributions of the observed and imputed values are shown in figures 9.1 to 9.12.

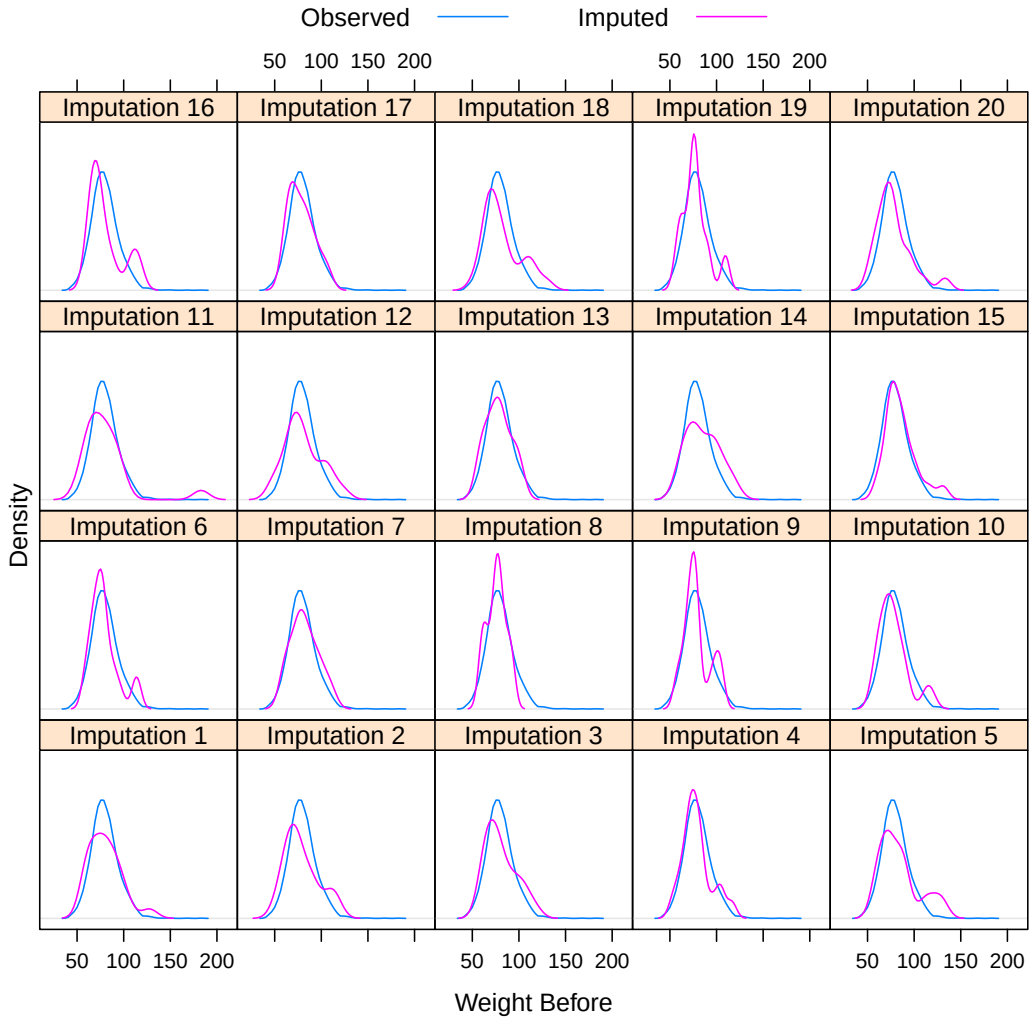


Figure 9.1: Distribution of observed and imputed values for weight at baseline for the cardiovascular rehabilitation cohort.

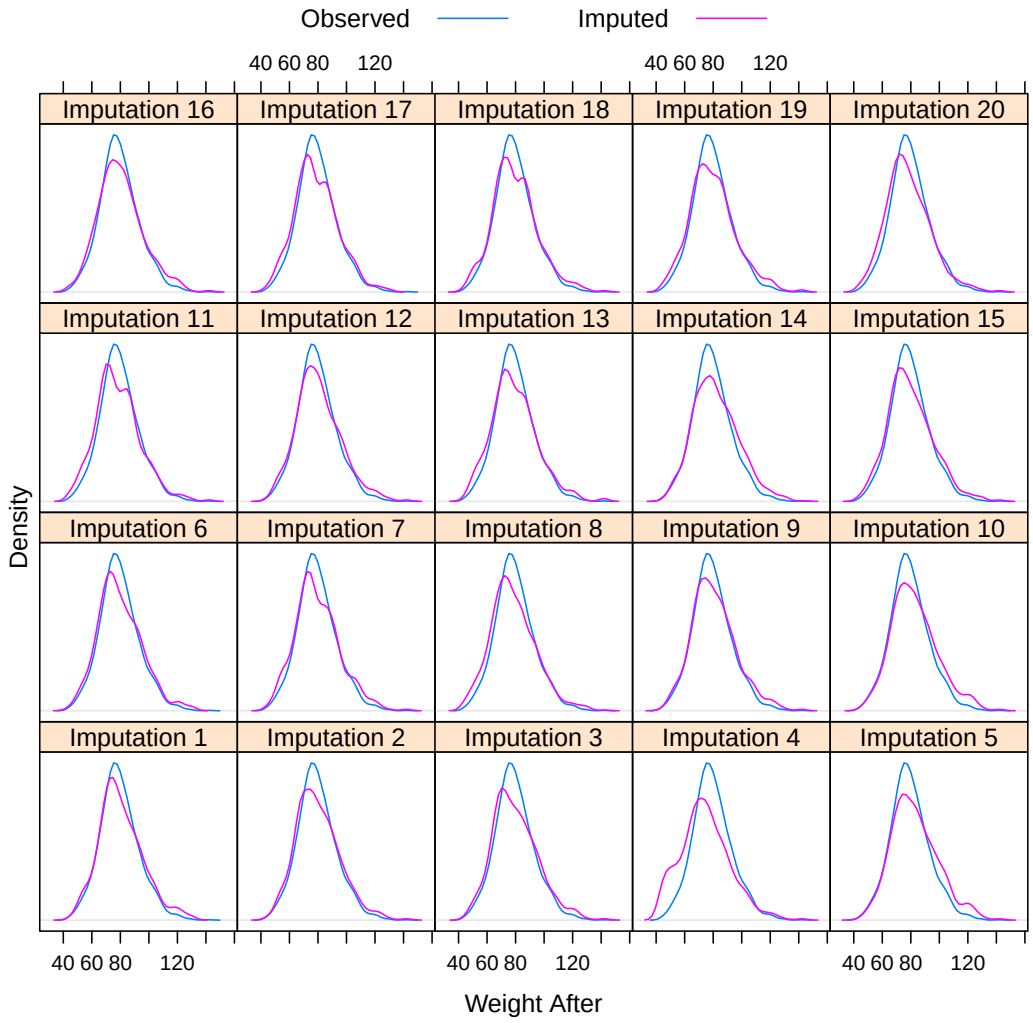


Figure 9.2: Distribution of observed and imputed values for weight at graduation for the cardiovascular rehabilitation cohort.

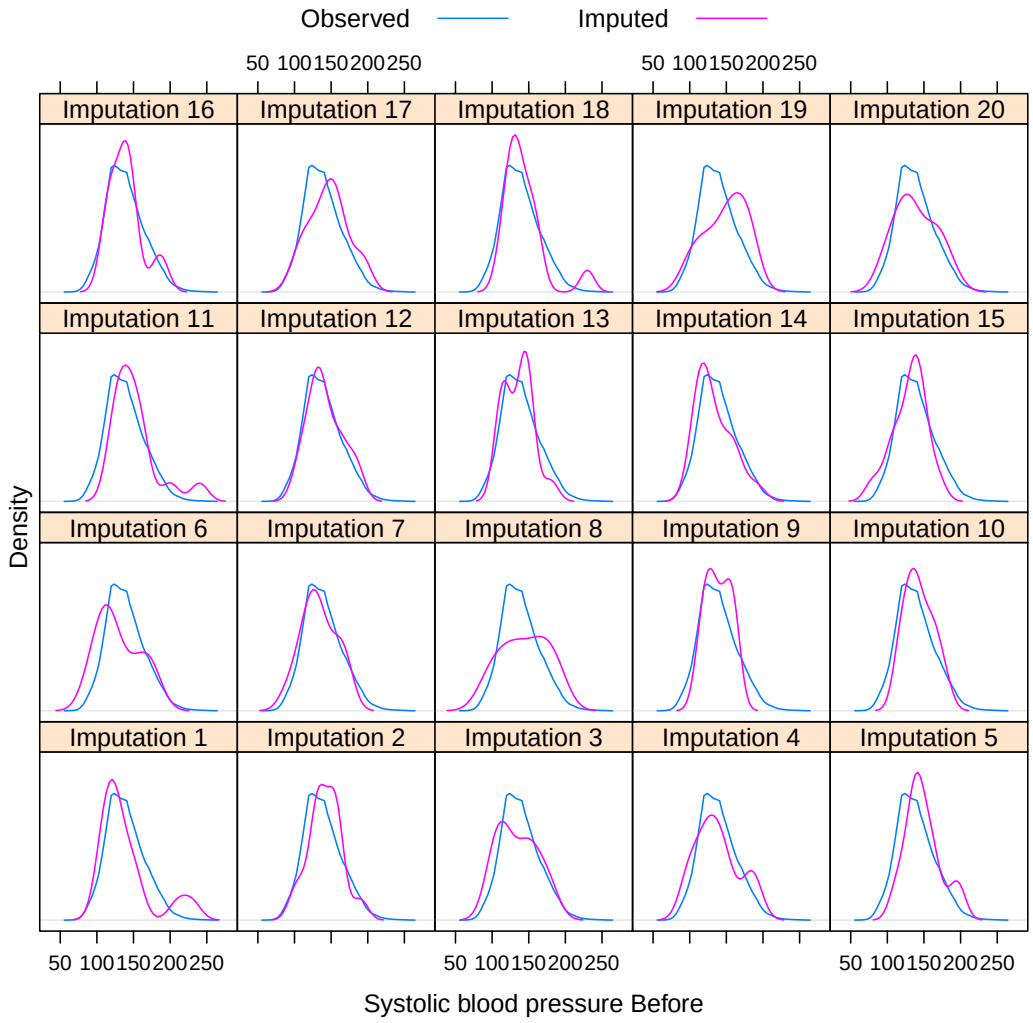


Figure 9.3: Distribution of observed and imputed values for systolic blood pressure at baseline for the cardiovascular rehabilitation cohort.

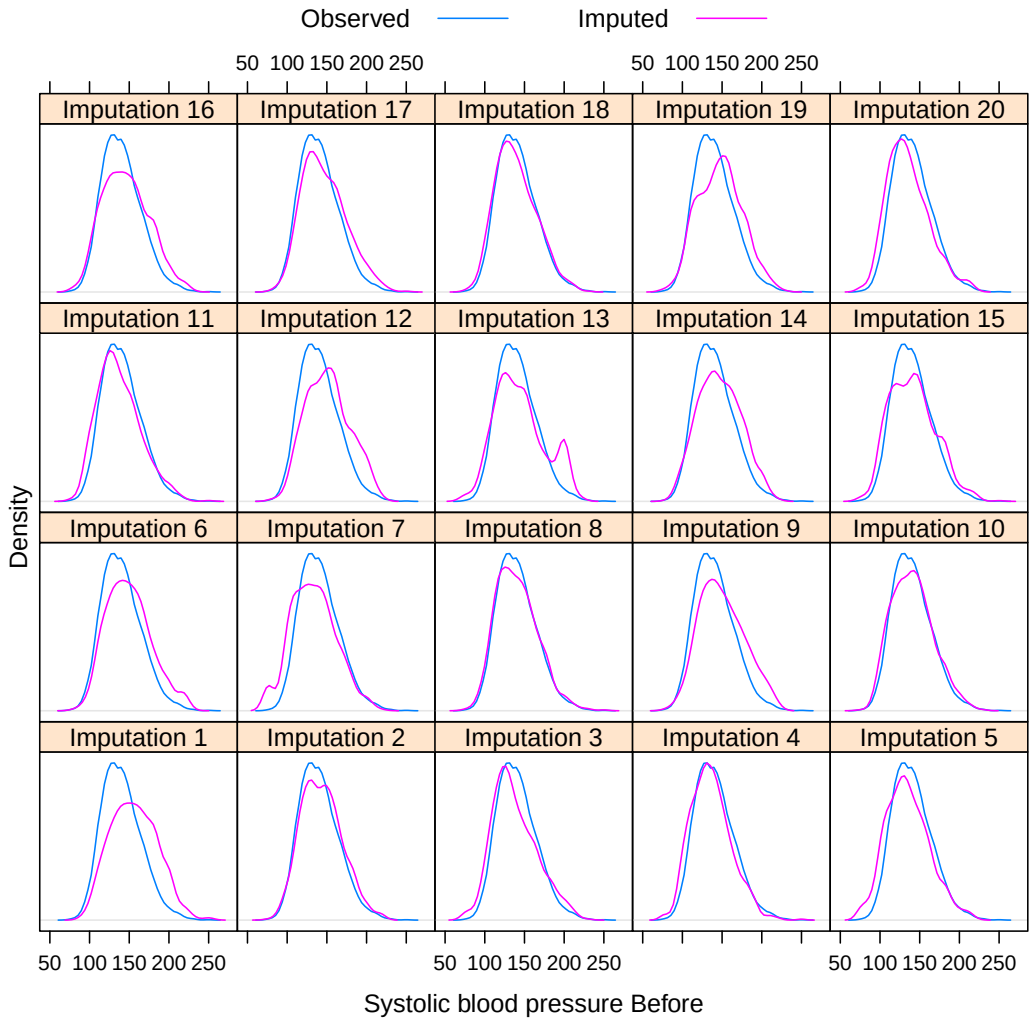


Figure 9.4: Distribution of observed and imputed values for systolic blood pressure at graduation for the cardiovascular rehabilitation cohort.

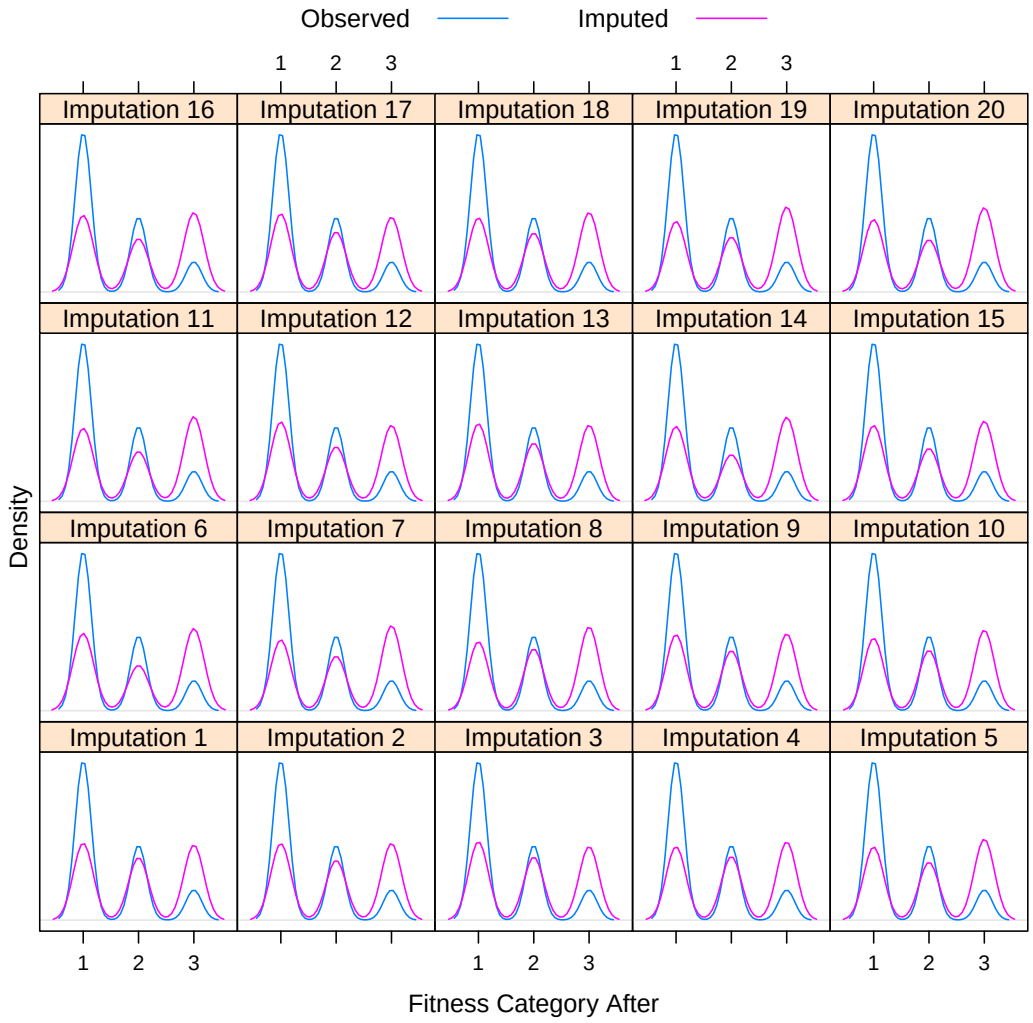


Figure 9.5: Distribution of observed and imputed values for fitness category at graduation for the cardiovascular rehabilitation cohort.

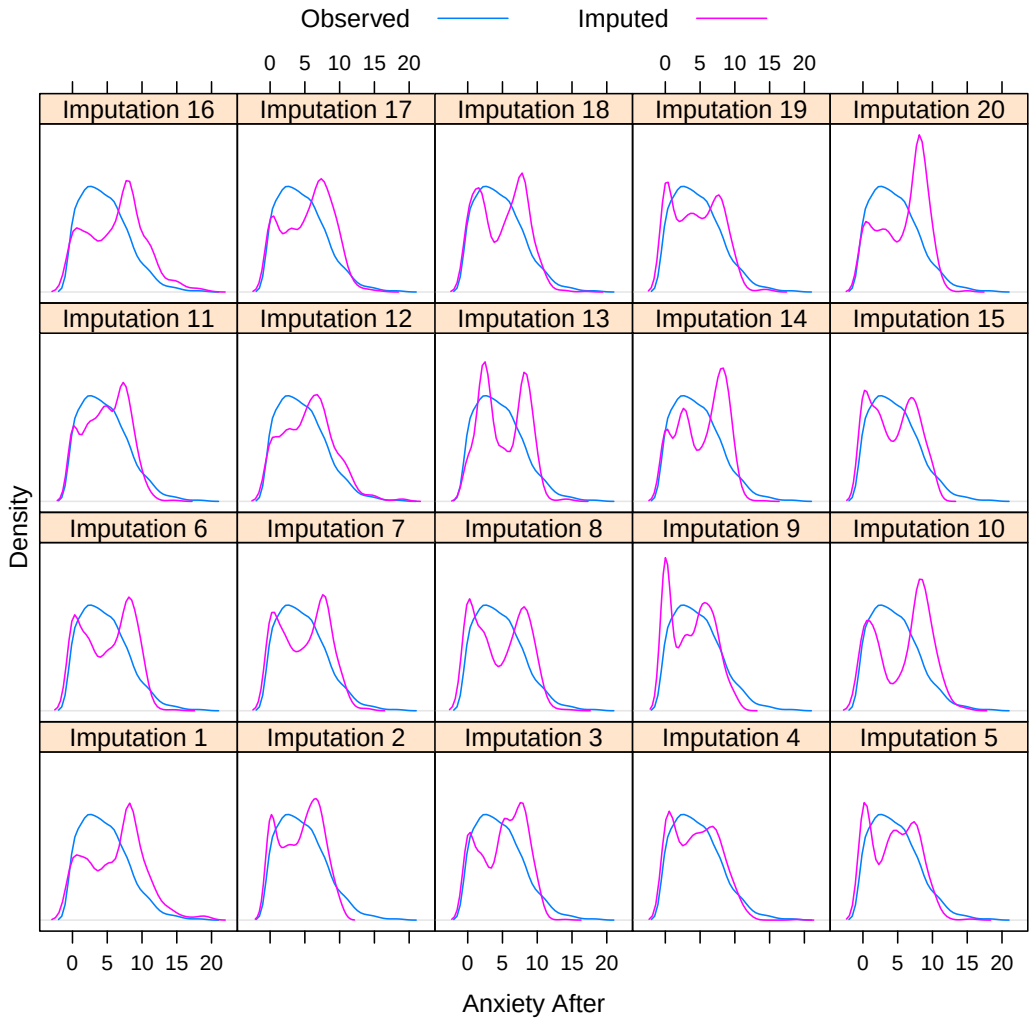


Figure 9.6: Distribution of observed and imputed values for anxiety at graduation for the cardiovascular rehabilitation cohort.

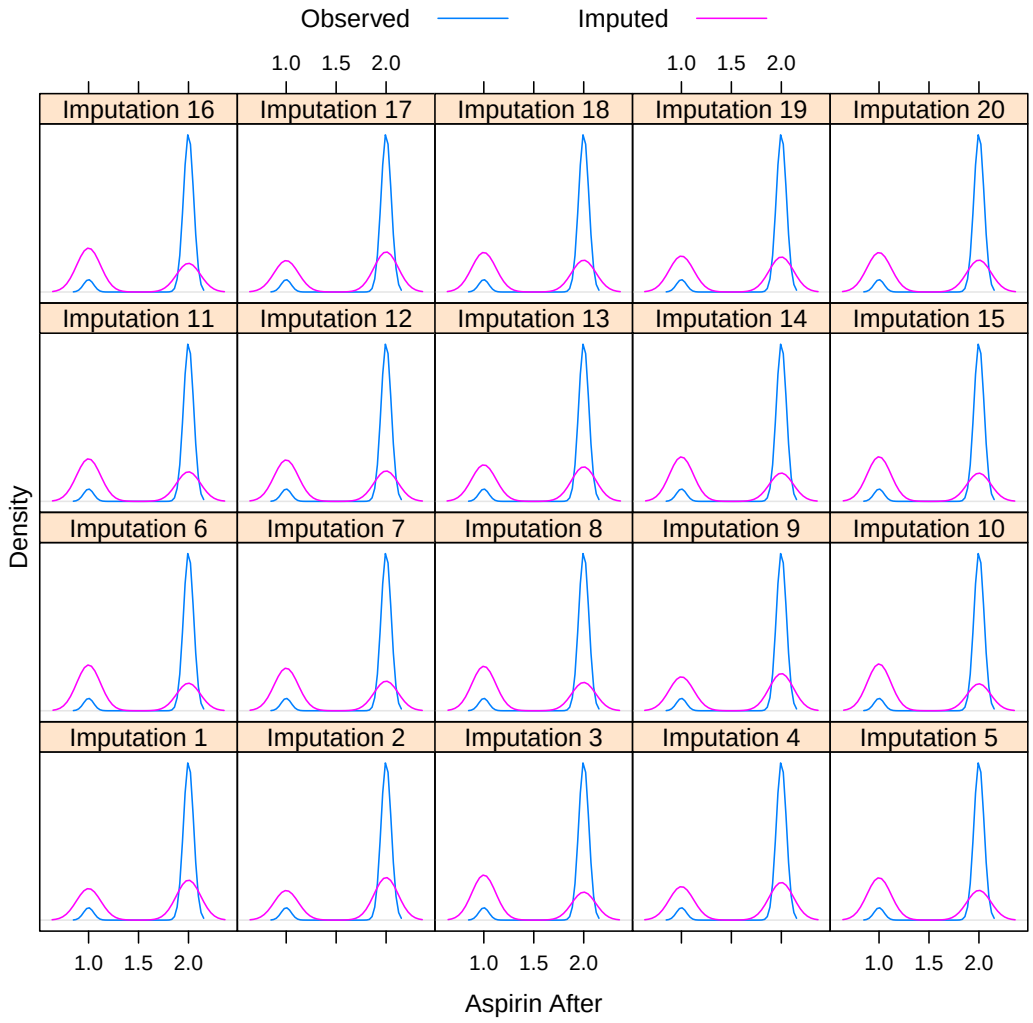


Figure 9.7: Distribution of observed and imputed values for aspirin prescription at graduation for the cardiovascular rehabilitation cohort.

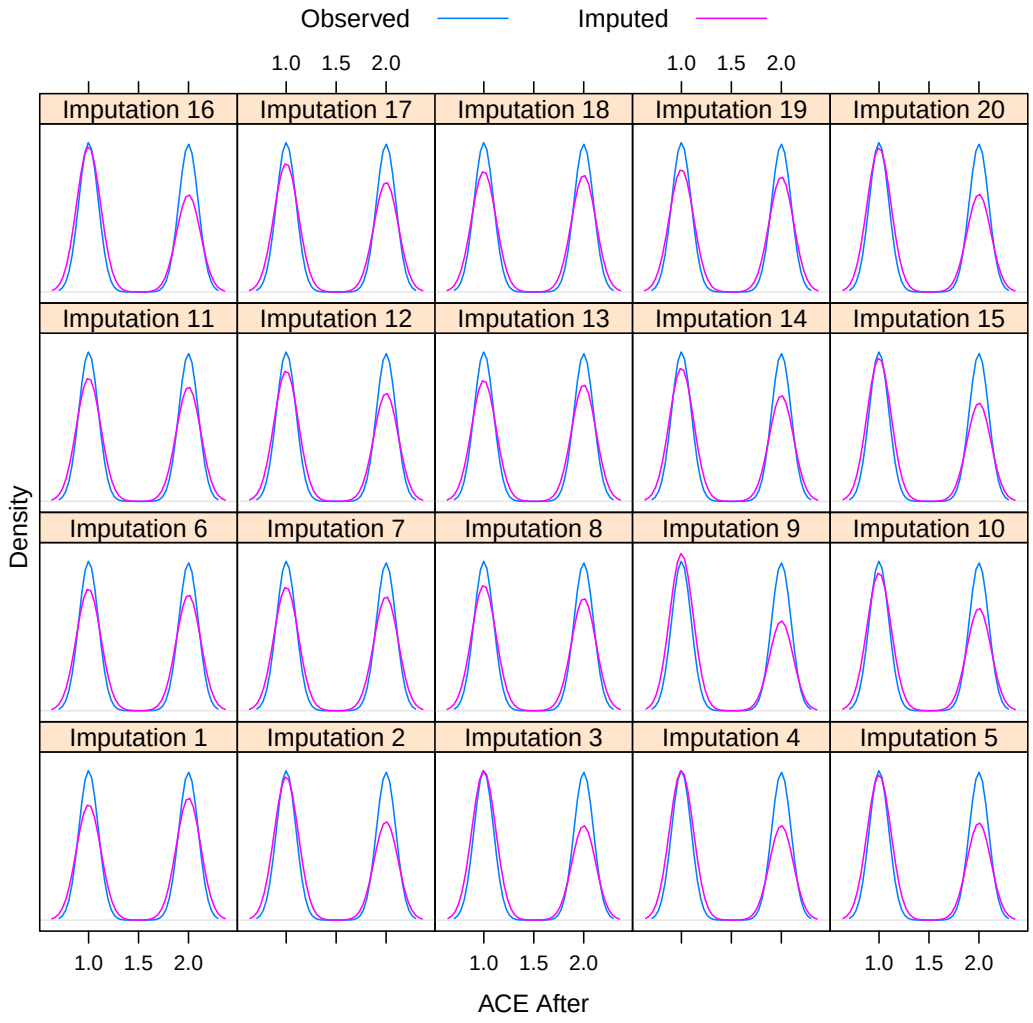


Figure 9.8: Distribution of observed and imputed values for ACE inhibitor prescription at graduation for the cardiovascular rehabilitation cohort.

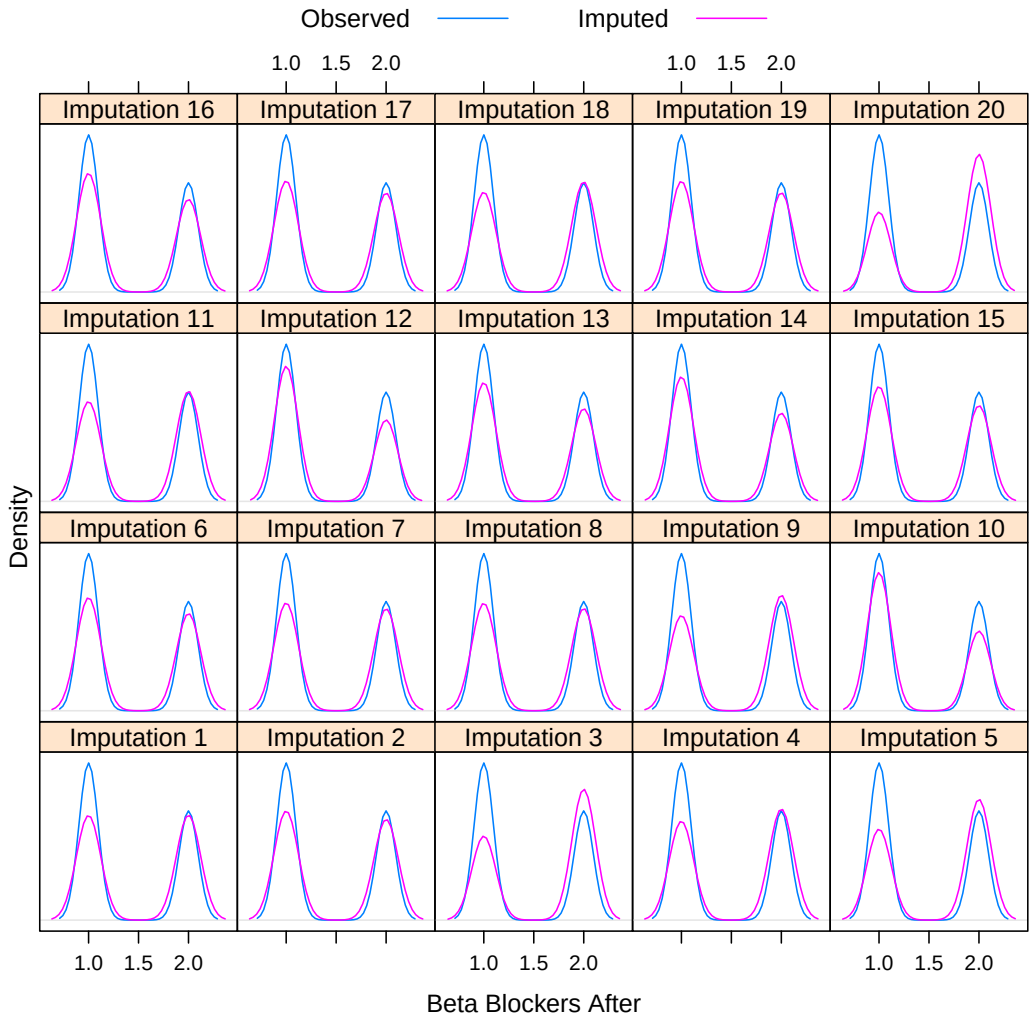


Figure 9.9: Distribution of observed and imputed values for beta blocker prescription at graduation for the cardiovascular rehabilitation cohort.

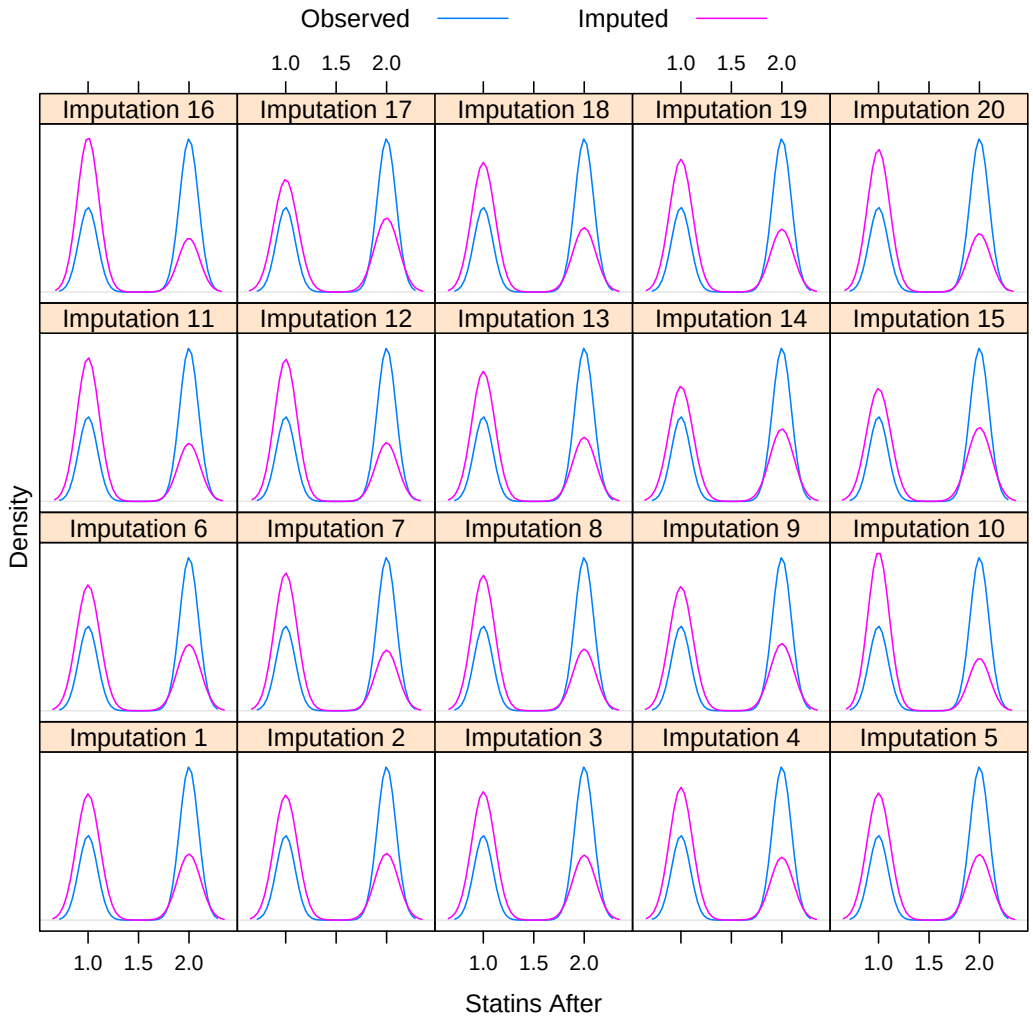


Figure 9.10: Distribution of observed and imputed values for statin prescription for the cardiovascular rehabilitation cohort.

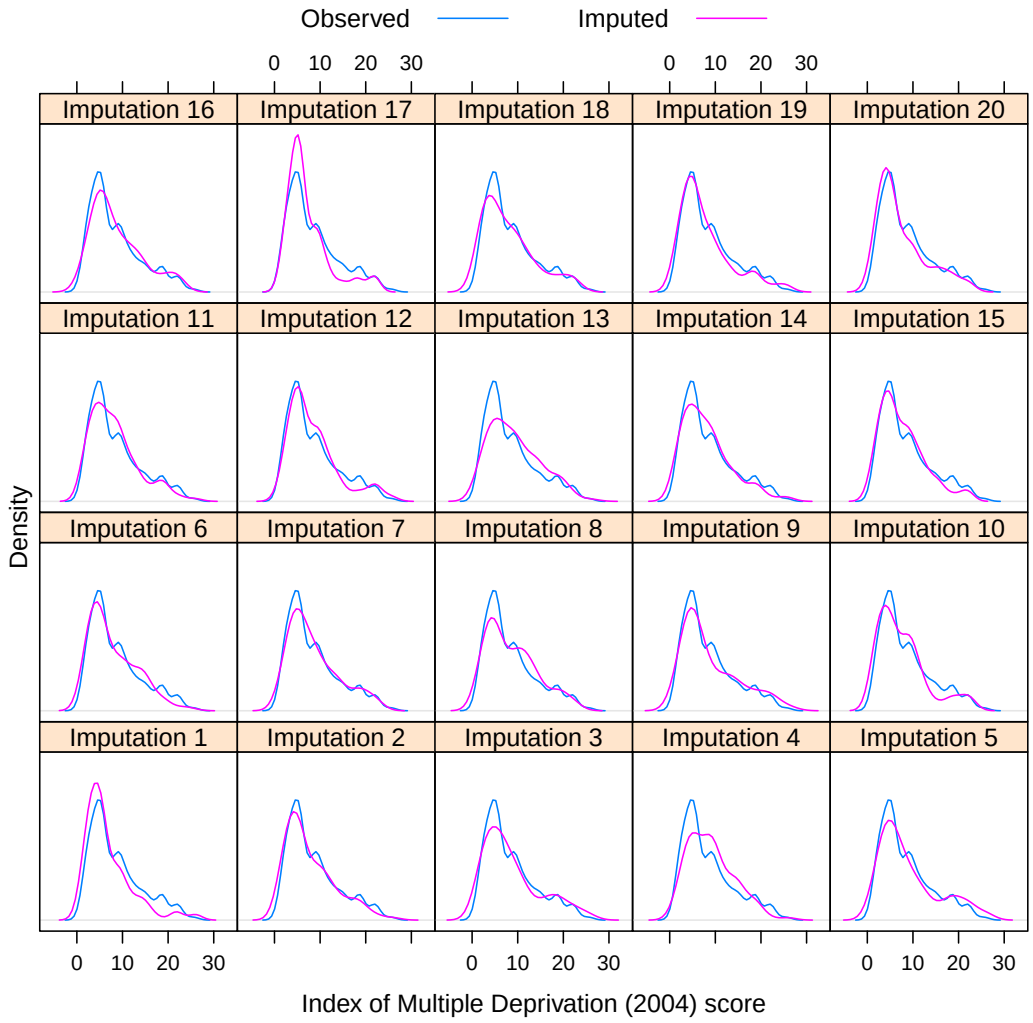


Figure 9.11: Distribution of observed and imputed values for deprivation at baseline for the cardiovascular rehabilitation cohort.

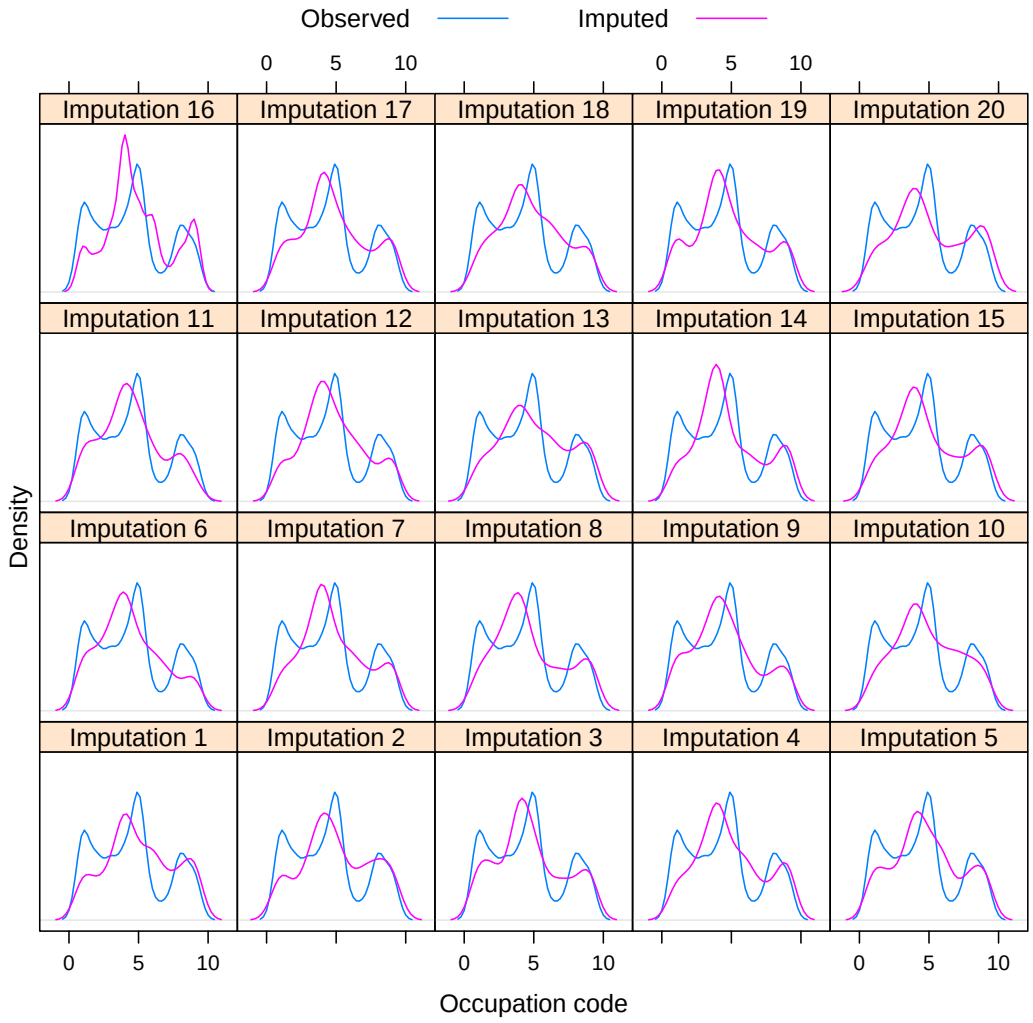


Figure 9.12: Distribution of observed and imputed values for occupation at baseline for the cardiovascular rehabilitation cohort.

9.2.3 Diagnostic plots for the proportional hazards assumption

These are plots of the Schoenfeld residuals versus log(time) for each of the predictors in the model including a lowess smoothing curve. See page 80 for an introduction.

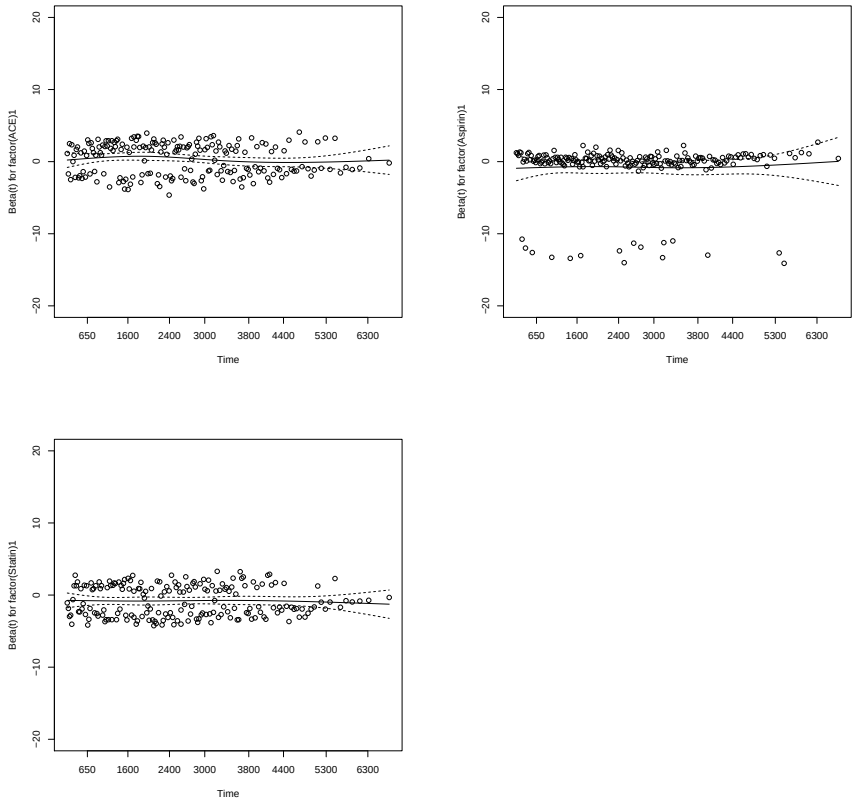


Figure 9.13: Schoenfeld residuals versus log(time)for ACE inhibitor, Aspirin and Statin prescriptions.

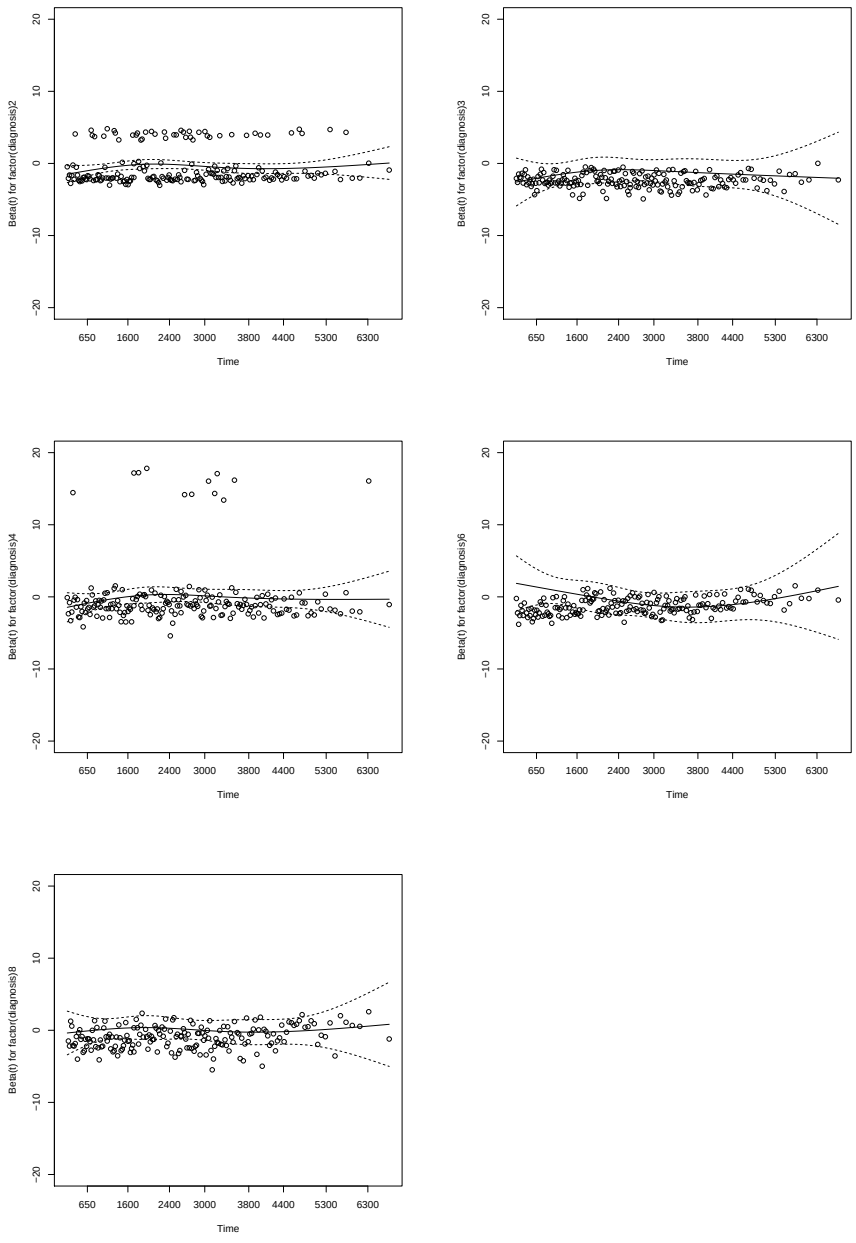


Figure 9.14: Schoenfeld residuals versus log(time). Diagnosis 1 is myocardial infarction, 2 is coronary artery bypass graft, 3 is percutaneous coronary intervention, 4 is angina, 8 is myocardial infarction and percutaneous coronary intervention and 6 is other cardiac diagnoses.

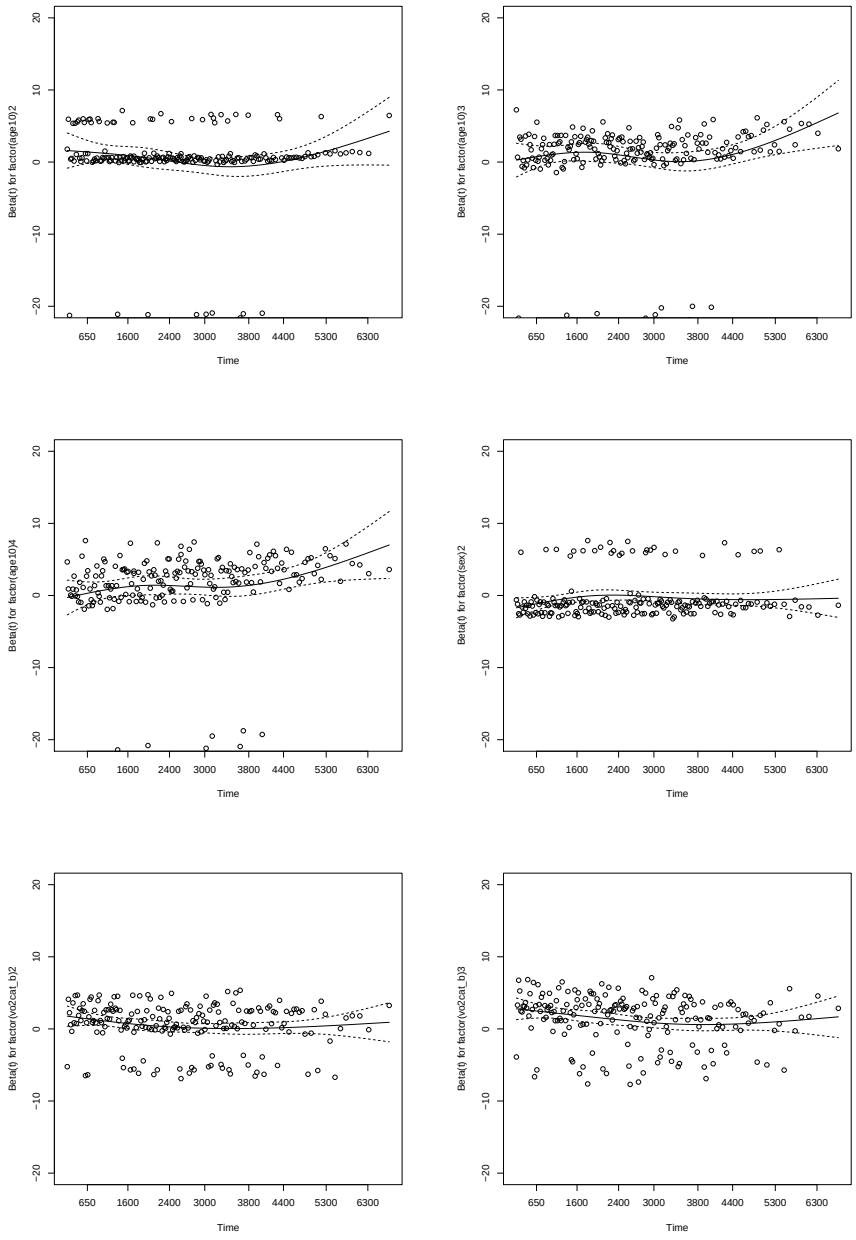


Figure 9.15: Schoenfeld residuals versus log(time). Age 1 is under 50 years, 2 is 50-59, 3 is 60-69 and 4 is 70 years and over. Sex 2 is female and fitness at baseline 1 is high, 2 is mid and 3 is low.