COMPUTER APPROACHES TO TOTAL HIP REPLACEMENT EVALUATION JUST PRIOR TO OPERATION

by

John A. Cogan

A thesis
presented to Trinity College Dublin
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical Engineering

Dublin, Ireland, 1999

I declare that this thesis has not been submitted as an exercise for a degree at Trinity College Dublin or any other University. I declare that I am the sole author of this thesis.
I authorise Trinity College Dublin to lend this thesis to other institutions or individuals for the purpose of scholarly research.

[redacted]

John A. Cogan

I authorise Trinity College Dublin to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

[redacted]

John A. Cogan

Summary

> As my evidence will show, innovation in medical devices results by and large from engineering-based problem-solving ... that does not in general depend upon the recent generation of fundamental new knowledge [108].
> *Edward B. Roberts*

The surgical operation of total hip replacement (THR) is, after 30 years, regarded as a successful medical procedure. The average life of an artificial hip has, however, remained at about 15 years for more than a decade, at a time when life expectancy and the quality of life of the elderly have greatly improved. As a result the expectations of patients are being increasingly frustrated.

It is in this context that failure to carry out a thorough scientific and clinical evaluation of THR options just prior to the operation is no longer tenable. Decisive information on patient variables and pathology and on operation choices becomes available at this time. The universal availability of advanced computing power has removed the time and knowledge limitations on orthopaedic surgeons that formerly prevented them from accessing all relevant medical and non-medical data. This thesis defends the proposition that what is involved in selecting the correct prosthesis is essentially a data-mining exercise that falls into the category of *classification*: patient/prosthesis combinations must be classified as accept or reject.

THR evaluation techniques divide into pre-clinical and clinical methods. Pre-clinical evaluations such as finite element analysis and fit-and-fill analysis are quantitative engineering procedures; clinical techniques such as radiographic and hip scoring analyses are in the paramedical domain. These evaluation techniques are reviewed and shown to be capable of considerable development using computer-assisted imaging and analytical techniques. Their outputs can be synthesised and made available to the surgeon in the time frame just prior to the operation, using a graphical user interface (GUI).

Two avenues of investigation were explored in the thesis. The first used the outputs from a selection of THR evaluation techniques to drive the

decision-making process of a rule-based expert system. Some 30 explicit rules, proffered by surgeons, were used to simulate professional judgement and decision-making. As a variation on the use of explicit rules, the suitability of a fuzzy expert system was also investigated.

The second avenue was to use the versatility and power of neural networks to unlock the information embedded in medical databases. Because of the dearth of Irish data two neural networks were trained on the survival analysis data contained in Sweden's National Hip Register. The first network investigated the effect of patient-related variables on survivability, the second the effect of implant variables.

The major findings of the thesis are:

- Comprehensive THR evaluation just prior to the operation is feasible and justified.

- There is an urgent need to standardise and harmonise patient and prosthesis data at a time when there is a general move to computerise medical records.

- A rule-base expert system is the most realistic and viable short-term option for THR evaluation just prior to the operation. Lack of familiarity with the quantitative evaluation techniques, however, made it difficult to arrive at a set of explicit rules on which the surgeons consulted could agree.

- Neural networks have formidable potential to mine historical THR records but this potential will remain unrealised if databases do not accommodate a greatly extended range of prosthesis design variables.

- Neural networks do not replace the portfolio of other prosthesis evaluation methods (finite element analysis, fit-and-fill, hip score, radiographic analysis etc.) but rather provide a significant additional technique which enhances the capability to evaluate and select prostheses.

Acknowledgements

I wish to thank my supervisor Professor David Taylor for his continued support during the many incarnations of this thesis. I was particularly fortunate that he recognised and directed me towards the valuable resource that the Swedish Hip Register turned out to be.

I am also deeply indebted to all those friends who stood by me during what has been an arduous journey. Above all, I am very grateful to my parents for their financial support and unflinching encouragement.

Abstract

The unsatisfactory record of survival of the total hip replacement (THR) surgical procedure can be attributed in part to the fact that candidate patient/prosthesis combinations are not fully evaluated in the time period before the operation. Pre-clinical and clinical evaluation techniques were reviewed and, when harnessed to modern computing, were seen to have impressive analytical and imaging capabilities.

To make the output of these evaluations accessible to the surgeon in the time period prior to the operation a rule-based expert system was prototyped. This consisted of 30 explicit rules, elicited from surgeons, and driven by the output of three evaluation techniques: finite element analysis, fit-and-fill analysis and clinical hip score analysis. A variation of this approach in the form of a fuzzy logic expert system was also considered.

An alternative method of making information available to the surgeon in a timely manner is to mine the rich data contained in historical patient, prosthesis and other databases with a feed-forward neural network. Two neural networks were trained and tested on survival analysis data from the Swedish Hip Register.

The rule-based expert system is a realistic short-term option for THR evaluation prior to the operation, while neural networks have enormous potential in the longer term if medical databases can be improved to meet the challenge.

Contents

List of Figures

List of Tables

List of Programs

Chapter 1

Introduction

> Knowledge is the most fundamental resource of the modern economy and accordingly the most important process is learning [79].
> *Lundvall*

> Knowledge is not just another resource alongside the traditional factors of production—land, labour and capital—but the only meaningful resource today. [33]
> *Drucker*

This thesis is fundamentally about a new approach to learning, a new way of improving the success of a surgical procedure that has important consequences for the human condition. An enormous amount of current and historical data are available about the total hip replacement surgical procedure but critical and compelling time pressures made it impossible, before the advent of modern computing power, for individual surgeons to interrogate and mine these data effectively. The first requirement is to assemble, in the form of a computer database, comprehensive data about prostheses[1] and patients, and about patient/prosthesis combinations. Next a set of computer-based analytical and graphical techniques, which can aid the surgeon in evaluating possible patient/prosthesis selections, must be identified. But the essential challenge is to devise a computer application which can extract from the data, and the results of sometimes competing evaluations of these data, clear patterns to guide the surgeon in his decision to accept or reject a given implant for a particular patient.

This search for a computer application stimulated an intellectual quest which started with a decision-support (synthesis) model, which had subjective evaluation output weighting, and progressed to a rule-based expert system and a fuzzy expert system, and finished with the design and testing of two feed-forward neural networks or "connectionist" expert systems.

Knowledge is now recognised as the source of continuous economic growth and enhanced quality of life. The term *knowledge-based economies*, in common usage, gives recognition to the role of knowledge and technology in economically advanced societies. Knowledge-intensive services such as education, communications and information are the most rapidly growing economic sectors and the share of research-intensive[2] manufacturing has more than doubled in the last ten years [100]. Indeed, it is estimated that more than 50% of output in the major OECD economies is now knowledge-based. The advent of the new computer and telecommunications technologies has

[1] A prosthesis is a fabricated substitute for a diseased or missing part of the body, such as a limb, tooth, eye, or heart value. In this thesis it generally pertains to the head of the femur and the associated joint (see Figure 3.4).

[2] Research-intensive sectors are sectors where a significant proportion of sales revenue is spent on R&D (electronics, bio-technology, pharmaceuticals etc.).

been decisive in providing the capability for this transformation. These technologies are the instruments through which information (codified knowledge) is stored, transmitted and manipulated.

A recent European Commission Report pointed out that the application of the new information technologies in the health sector conspicuously lags that in other economic sectors, particularly industry, and that the opportunities for progress in that particular sector are correspondingly greater.

> Widespread implementation of the information and communications technologies (ICTs) in the health sector is expected in all Member States. This will result in improved availability and quality of health services and will provide support for rationalisation and improvement of cost-effectiveness of health service systems. A great potential for application is foreseen in all parts of the health sector, namely prevention, promotion, creative services and rehabilitation, as well as in social services [39].

1.1 KNOWLEDGE REQUIRED FOR HIP TECHNOLOGY

The part of the health sector identified for analysis and application of information and communications technologies in this thesis is the surgical operation of total hip replacement. Even though this is rated as a very successful surgical procedure the economic and social pressures to develop prostheses with significantly improved performance and to institute more cost-effective delivery systems are unrelenting. Waiting lists are expanding almost everywhere because longer life expectancy, combined with a lower birth rate, is changing the age distribution of the population. The number of first time hip replacements carried out world-wide in a single year is now in the order of ten million. In Great Britain alone the figure is 400,000 a year, and the number continues to rise [136]. This is compounded by the high proportion of patients who require revision surgery. The average life of an artificial hip is currently about 15 years but the variation between revision rates for different prosthesis designs and for different surgical procedures is so great as to render hip replacement more of an art than a science. Furthermore, the prostheses in current use in the UK for hip replacement vary in cost by a factor of nearly ten [97]. There is absolutely no evidence that these cost differences are reflected in prosthesis survival times or in the subsequent comfort and quality of life of treated patients [92].

The different kinds of knowledge which are encountered in the art/science of hip replacement reflect very powerfully the kinds of knowledge which are important in a knowledge-based economy more generally. The most fundamental dichotomy is between codified and non-codified knowledge. The latter is also referred to as *tacit* knowledge or know-how, and is often jealously guarded by specific trades and professional elites.

Codified knowledge is that part of knowledge which constitutes *information.* There is a constant and concerted effort to codify as much knowledge as possible and elevate it to the status of information. The inexorable increase in the codification of elements of knowledge, which were previously tacit, is the hallmark of the *information society* or the knowledge-based economy. Information can be stored, retrieved, transmitted and organised in a rich variety of ways. Sources of information include digitised databases, books, scientific journals, working papers, images, video clips, sound and voice recordings, graphical displays as well as electronic mail.

With specific reference to total hip replacement, codified knowledge can be classified into a "know-what" component and a "know-why" component. Know-what refers to knowledge about *facts*, such as facts about the patient, contained in the patient database: age, height, bone condition, level of activity, medical history, and facts about femoral components, contained in prosthesis database: size, length of stem, material properties etc. Know-why on the other hand refers to scientific knowledge such as that emanating from engineering principles, the properties of matter and the behaviour of human tissues. Know-why is central to a deeper understanding of the functioning and the continued technological development of prostheses. This knowledge is arrived at by using techniques such as finite element analysis, mechanical testing, radiostereometric analysis and other design evaluation procedures.

Non-codified or tacit knowledge, however, is also critical to progressing the art/science of hip replacement. This is rooted primarily in practical experience and is not normally transmitted through formal channels of communication. This component of knowledge is referred to as "know-how." In the case of hip replacement, the skill and experience of the orthopaedic surgeon, which are the foundation for his confidence in performing the operation, illustrate the centrality of this component of knowledge in complex medical procedures. Know-how tends to be retained by individuals and not transmitted to others until it is codified and transmuted into information. The networking of surgeons and the linkages and interactions between the repositories of the different elements of knowledge in hip replacement are instruments for the sharing and diffusing of know-how relating to hip prosthesis selection.

There is a long tradition of sharing/trading tacit knowledge between experts (in this case surgeons) in specialised fields but this is invariably confined to closely-knit groups comprising inner colleges or centres of excellence. Such a system operates to the detriment of the overwhelming majority of practitioners (and especially their patients in peripheral economies) who do not belong to a select inner circle. Dissemination of socially and economically significant medical knowledge is greatly restricted and, as a result, society's return on investment in health and human betterment is sub-optimal. There is unnecessary experimentation and rediscovery of facts, already known to researchers and some practitioners, by surgeons who do not have access to the full spectrum of available knowledge including elements of know-what, know-why and more especially know-how. The current situation in the field of hip replacement surgery is a metaphor for failure to exploit the potential of information technology for the capture and diffusion of existing knowledge.

1.2 NEW APPROACHES TO THE CAPTURE AND DIFFUSION OF TACIT MEDICAL KNOWLEDGE

During the eighties several medical disciplines developed "intelligent" decision support systems, based on artificial intelligence techniques, including expert systems. This was a major breakthrough in the accretion of knowledge because it enabled the codification of the critical tacit component of medical knowledge to the extent that it formalised the intelligent behaviour and the reasoning processes utilised by the world's leading medical experts in their decision-making.

The initial challenge for this thesis was to design an artificial intelligence instrument which combines the outputs of a portfolio of quantitative evaluation methods in such a way as to make them available and usable by the surgeon in the short time-period before an operation. The methods chosen for the prototype application were finite element analysis, fit-and-fill analysis and clinical hip score evaluation. (These methods and a number of others are elaborated on in Chapter 4.) The application greatly facilitated the generation of evaluation outputs but it provided a less than satisfactory answer to the relative weighting of these analytical evaluations. The decision-making process could not be adequately represented without incorporating a better means of weighting the results of these evaluations. This limitation was addressed by designing an expert system using rules which replicate the decision-making process of selected experts.

Rule-based expert systems, however, have only limited capacity to capture solutions to very complex problems such as those common in real life [24]. The current generation of expert systems has fallen short of the progress expected. Essentially, all the easy problems were quickly solved. When conventional expert systems ran up against highly non-linear and computationally complex problems these required prohibitive amounts of computer power—if they could be solved at all.

Fuzzy expert systems overcome some of the limitations of more traditional expert systems by being able to encapsulate real-world uncertainty and imprecision as an intrinsic part of the system. They do this by incorporating fuzzy logic and fuzzy set theory into the model. Fuzzy sets deal with subsets of the universe that do not have well-defined boundaries but fuzzy expert systems still depend on being able to formulate adequate rules with which to represent the system being modelled. To the extent that the rules extracted from experts capture the reality, the fuzzy expert system is a formidable instrument.

Difficulty in reaching a consensus on the rules to be used is a major limitation on the capacity to apply rule-based and fuzzy expert systems. Neural networks obviate this need to explicitly enumerate a set of rules but the resultant trade-off is that, whereas an expert system can explain its decision-making process neural network decision-making cannot be made transparent to the user. The user's confidence that the system works is largely based on the fact that neural networks have a proven track record.

Developments in neural networks since the mid-eighties have transformed the whole area of artificial intelligence. Neural network paradigms for a range of complex applications have emerged. More particularly, the so-called "feed-forward" neural network paradigm, which is the model used in this thesis, has proved extraordinarily flexible and amenable to algorithmic manipulation. As a result the recent surge in neural network activity appears to be gaining momentum.

The decision-making powers of a neural network are *not* embodied in a set of explicit rules but are implicitly built into the structure of the network itself through a system of values or weights given to the network's nodes. The task of designing a neural network for a specific application is something of a black art requiring experimentation and repeated trial and error. A neural network must be trained using a set of examples that is representative of the problem space under investigation. It was critical to the current research, therefore, to have adequate training-set data available.

The Swedish Hip Register has catalogued about a hundred and thirty thousand-replacement operations that took place in Sweden between 1978 and 1994. More particularly some eight and a half thousand of these operations were for revision surgery [83]. Thus this database alone provides a fairly substantial record of the survivability of hip implants, representing a variety of prosthesis characteristics and a range of different patient situations.

For the purposes of this thesis two neural networks were constructed: the first to predict the survivability of an implant based on patient variables; the second to predict the survivability of an implant based on implant variables. The main benefit of the first neural network is that, by exploiting rich historical data, it facilitates the prediction of the survivability of a prosthesis in a patient with a specific demographic profile (age, sex etc.) and pathology. The second neural network predicts the survivability of different implants and can go on to postulate the survivability of untested permutations of implant variables. The training and test data for both these neural networks were extracted from graphs in research papers reporting findings based on the Swedish Hip Register [1, 82, 83]. These graphs, however, did not permit a more comprehensive exploitation of the data which could be achieved by using a combination of patient and prosthesis variables to predict survivability.

1.3 OUTLINE OF THESIS STRUCTURE

This opening chapter described the context within which this work is being carried out. Knowledge is the main currency in all aspects of modern economies including the health sector and, in the unfolding of this thesis, increasingly sophisticated knowledge-based techniques which are made possible by currently available computing power, are applied to hip prosthesis design evaluation.

Chapter 2 outlines the research objectives of the thesis and the learning process through which progress was made towards a solution to the problem of hip prosthesis selection.

Chapter 3 describes the design issues and directions in total hip replacement. The relative merits of cemented as opposed to cement-less fixation are considered. Custom-made prostheses have in general not proven more reliable than standard off-the-shelf designs. Competing design criteria are described and discussed.

Chapter 4 reviews the pre-clinical and clinical methods currently used to evaluate hip prosthesis design and it prototypes a computer application which synthesises those evaluation methods that are amenable to quantification and graphical representation. This computer application gives the surgeon, in the time-frame before the operation, the capacity to make decisions on prosthesis selection with the aid of computer-assisted imaging and analytical techniques.

Chapter 5 augments and enhances the computer application, which was developed in Chapter 4, by adding a rule-based expert system which aims to build into the model, the judgement and decision-making powers of selected experts, when faced with the task of evaluating competing patient/prosthesis combinations.

Because of difficulty experienced in getting surgeons to agree on the explicit rules which guide their decision-making, Chapter 6 introduces a further modification to the original model: a fuzzy logic expert system replaces the more conventional expert system.

Up to this point the thesis has concentrated on using advanced engineering and medical analytical methods, together with information available in the surgeon's patient and prosthesis databases, to determine the best prosthesis for a given patient in the time-frame immediately prior to the operation. But since the underlying theme of the thesis is critically about effective knowledge management, cognisance must also be taken of the great wealth of codified and non-codified knowledge which resides throughout the world in national hip registers. This mine of information remains to be accessed and exploited with the power of modern computing.

Chapter 7 deals with the application of feed-forward neural networks to hip prosthesis design evaluation. Section 7.2 presents the source of empirical data for this thesis. The training sets for the two neural networks were taken from graphs in research papers that utilised data extracted from the Swedish Hip Register so the background to this register is described. Section 7.3 gives details of the design, construction and operation of the two neural networks. The trial and error procedure for deciding the number of nodes and layers for each network is explained and the complex process of training the networks is detailed. Section 7.3.3 evaluates the performance of the two experimental networks.

Chapter 8 reviews the research techniques used in the thesis and places them in the context of the rapidly expanding data mining category of *classification.* It then lists the conclusions which flow from the research and makes specific recommendations on steps to be taken to remove the data limitations that are currently inhibiting the realisation of the full potential of neural networks in particular, and knowledge based techniques in general.

Chapter 2

Objectives and Overview of Research

> Objectives must be reasonable. Neither surgeons nor engineers will ever make an artificial hip joint which will last for thirty years and at some time in this period enable the patient to play football [19].
> *Sir John Charnley*

The fundamental proposition underlying this thesis is that knowledge-based techniques, culminating in neural networks, have an important role to play in matching patient/prosthesis combinations and in predicting the expected life of implants in patients, who undergo hip replacement surgery, by codifying and exploiting the empirical evidence contained in clinical, patient and prosthesis databases.

2.1 RESEARCH OBJECTIVES AND EVALUATION METHODOLOGIES

The objectives are:

- to outline the context in which the new information and communications technologies can accelerate the growth of knowledge and understanding of total hip replacement and so extend the survivability of hip implants;
- to review design goals, design trade-offs and emerging design criteria in the field of total hip replacement;
- to critically review pre-clinical and clinical evaluation methodologies currently in use for hip prosthesis design and to determine which of these are amenable to quantification and/or graphical representation;
- to design a computer application to synthesise selected hip evaluation methods;
- to balance the quantification and graphical emphasis of the computerised evaluation model by adding a rule-based expert system to capture distilled professional expertise and non-codified knowledge;
- to replace the expert system's explicit rules with fuzzy rules, in order to overcome the former's limitations;
- to evaluate the role of neural networks in hip prosthesis design and selection, as a complement to expert systems, and to build and validate

a neural network using data from the Swedish Hip Register;

- to discuss the findings and to highlight limitations in the Swedish data for the purposes of hip design evaluation, and to indicate how these might be overcome.

2.2 EVOLUTION OF THESIS

The brief for this thesis was to apply artificial intelligence decision-making techniques to the field of hip replacement with the ultimate goal of helping surgeons in the complex and subjective process of selecting appropriate prostheses. At first sight, the highly unstructured approach to decision-making, which is a feature of this area of surgical practice, gave reason to expect that a combination of modern computing power and rule-based expert systems would achieve demonstrable success. But as the work proceeded the words of Piet Hein, the poet philosopher, proved to be salutary:

Problems worthy of attack,
Prove their worth by hitting back.

2.2.1 Synthesis of Current Evaluation Methods

The first approach was to attempt to integrate a number of critical analytical methods, used at the pre-clinical and clinical stages of hip implant surgery, to produce evaluation ratings for a selection of prostheses, with respect to individual patient parameters [21].

The three principal evaluation methods used in this thesis are: finite element analysis, which is used to highlight stress levels which are portent of bone resorption[1] and bone growth; fit-and fill-analysis, which indicates the stability of a patient/prosthesis combination; and clinical hip score evaluation which enables comparison with similar patient/prosthesis combinations in the past (see Chapter 4). These quantitative methods can deliver considerable analytical capability but tend to be specific to particular disciplines, to operate independently and, frequently, to be answerable only to their separate areas of expertise.

The purpose of the model was to co-ordinate and combine these methods to produce a composite evaluation of prospective patient/prosthesis combinations. This requires encapsulating the data from these methods in a decision-making mechanism where a predetermined weight is given to the separate evaluation outputs. A more considered approach is to incorporate an expert system to embody the decision-making process employed by human experts.

The data required to drive the evaluation methods and with them the decision-making capacity of the application, was stored and compiled in a group of databases which mirrored the structure of the problem. The effectiveness of the decision-making is dependent on the completeness of these databases.

---

[1]Bone resorption is the loss of bone due to disease (such as rheumatoid arthritis) or stress shielding. Bone grows in proportion to the stress placed on it (Wolff's law). Anything which distorts the normal flow of stress through the bone is termed stress shielding.

2.2.2 Rule-Based Expert Systems

The next stage was to add a rule-based expert system to the application. On the face of it, this was a straightforward task: surgeons can choose from a wide range of hip design variations to suit their diagnosis of an individual patient's circumstances and this has close parallels with many successful applications of conventional expert systems. The best-known expert system in medicine, developed in the 1970s, is *MYCIN*. The development of this expert system took place at Stanford University; E.H. Shortliffe, in particular, played an important role in its development. The MYCIN system was able to assist interns in the diagnosis and treatment of a number of infectious diseases, in particular meningitis and bacterial septicaemia [121].

The MYCIN system has significantly influenced the subsequent development of expert systems. Even now, this expert system and its derivatives are sources of ideas concerning the representation and manipulation of medical knowledge. The MYCIN system has also given an important impulse to the development of similar expert systems in fields other than medicine.

The *INTERNIST-I* system is another example of an expert system initially developed early in the 1970s. The system is still being developed by H.E. Pople and J.D. Myers at Pittsburgh University. Later in their research, Pople and Myers renamed the system *CADUCEUS*. One of the objectives of the INTERNIST/CADUCEUS project is to assist in the study of models for diagnosing diseases in internal medicine [87, 78].

In internal medicine several hundreds of different diseases may be discerned. An intern not only has to bear in mind all the clinical presentations of these diseases during the diagnostic process, but must also take into account possible combinations of symptoms i.e. symptoms that can be caused by the interaction of several diseases present in a patient at the same time. The number of diseases in internal medicine, and the possible combinations of clinical signs and symptoms, is so large that it is not possible to consider each in isolation. INTERNIST/CADUCEUS focuses on those diseases that are most likely, and the interaction between them, given the symptoms, clinical signs, and the results of laboratory tests obtained from the patient.

Returning to the problem of hip implant design selection, it was clear that an expert system could potentially be built around the rules by which orthopaedic surgeons make hip implant decisions. These rules would be complemented with rules elicited from the other scientific sources. This would help formalise the process of choosing a prosthesis, a process that has hitherto been arbitrary and unscientific for most surgeons. After interviewing a number of surgeons (who were at the time co-operating with the Bio-engineering Unit in TCD), it became evident that it would be very difficult to arrive at a consensus among the orthopaedic fraternity as to a canon or set of rules governing the choice of a prosthesis. The Irish experience in this regard was somewhat atypical in so far as the range of prostheses in use is, in practice, limited to proven prostheses and mainly to the Charnley. There is very little research or experimentation with new designs by comparison with larger economies such as the US and Sweden (see Chapter 3).

2.2.3 Fuzzy Expert Systems

The *fuzzy expert system* is an adaptation of the traditional expert system. Fuzzy expert systems are universal approximators and are well suited to modelling highly complex, invariably non-linear problems. They are able to approximate the behaviour of systems that display a variety of poorly understood and/or non-linear properties. Fuzzy expert systems execute at a faster rate than conventional rule-based systems and require fewer rules. They have the ability to explain their reasoning and as a result they provide an ideal way of addressing difficult problems. They can surmount many of the computational and complexity difficulties that were responsible for the failure of some earlier expert systems. Fuzzy logic, as conceived by its inventor Lotfi Zadeh, provides a method of reducing as well as explaining system complexity.

The rule-based expert system of Section 2.2.2 was replaced by an expert system using a set of fuzzy logic rules. The approach taken was to evaluate the risk of failure of a prosthesis in a way similar to that by which project risks are assessed [24]. The acceptance or rejection of a particular patient/prosthesis combination was based on a calculation of risk of failure.

2.2.4 Feed-Forward Neural Networks

The rule-based expert system, and the fuzzy expert system just described, are both based on the formulation of *explicit* rules. Explicit rules have been difficult to formulate, however, because the field of total hip replacement has not yet stabilised and there is some uncertainty as to how the technology will develop. This is not untypical of technologies in their early stages of development [132]. The absence of a dominant design has been very apparent in the case of total hip replacement and this may be due to the length of time and the number of subjects that must be looked at, in order to verify a new development. It is estimated that as many as 3,000 patients with a new device would have to be followed-up for 5 years, before that device could be proven significantly better—that is, enduring longer—than an alternative [55]. Despite the imperatives for the development of an improved prosthesis, progress has proceeded in an *ad hoc* fashion and this has not been conducive to productive research. Many mistakes could have been avoided if new hip implant technology had been implemented in an ordered fashion [81]. The case is made for moving away from artificial intelligent decision-making based on the formulation of *explicit* rules, and moving towards an artificial intelligence system which involves making the rules implicit. Hence the decision to pursue the neural network route.

Neural network techniques do not require an explicit statement of rules. Rather, they require a sufficiently large training set of examples that reflects the behaviour over an extended historical period of the system they are expected to model. Neural networks learn to reproduce these training examples to within an acceptable level of error. This is an iterative process that involves repeated adjustment to the weights pertaining to the nodes of the network. The principle is that once the network has been trained on a representative set of examples it will then be able to accurately predict and mimic the system when new input values are presented to it.

Two neural networks were constructed as part of this thesis. The first one to predict survivability of an implant based on patient variables; the second one to predict survivability of an implant based on implant variables. The ideal would be a single neural network, combining both implant-based and patient-based variables, but such a network could not be constructed, given the structure of the data available to the researcher.

Chapter 3

Design Issues and Directions in Total Hip Replacement

Traveller, there is not path; the path you must build as you walk.

Antonio Machado

The joint replacement procedure (also called arthroplasty) involves replacing a painful joint with an artificial joint, called a prosthesis and is appropriate when a patient experiences severe, incapacitating hip pain, due to conditions such as osteoarthritis, rheumatoid arthritis or injury.

In a healthy hip (see Figure 3.1), the smooth ball on the end of the thigh bone fits easily into the end of the hip socket to form the "ball and socket" joint. A layer of cartilage covers the ends of these bones, serving as a cushion which allows the ball to glide easily within the socket.

Severe pain and decreased movement can result if the cushion of cartilage degenerates as a result of osteoarthritis or other diseases (see Figure 3.2). The joint surfaces are allowed to rub against each other, becoming rough,

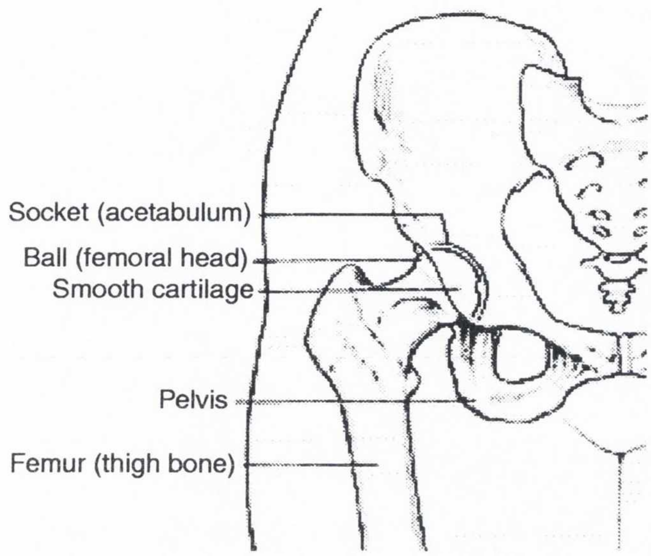


Figure 3.1: Healthy hip.

pitted and irritated.

The hip prosthesis (see Figure 3.3) consists of a specially designed ball and socket that replaces the worn hip joint. The ball and stem replace the worn ball of the thigh bone and a cup replaces the rough hip socket. The prosthesis has smooth surfaces that fit together and allow the ball to move easily and painlessly within the socket, much like a healthy hip.

3.1 BACKGROUND

Total hip replacement (THR) is an established surgical procedure for replacing a degenerated hip joint with a mechanical equivalent made from metal and plastic, to restore function to the joint. The procedure is typically elective and is indicated when the pain in the hip joint becomes severe, resulting in a reduction in the activity for the patient that does not resolve

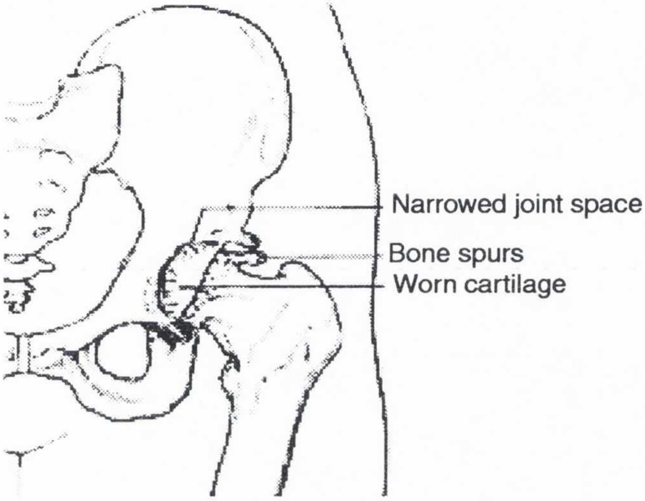


Figure 3.2: Problem hip.

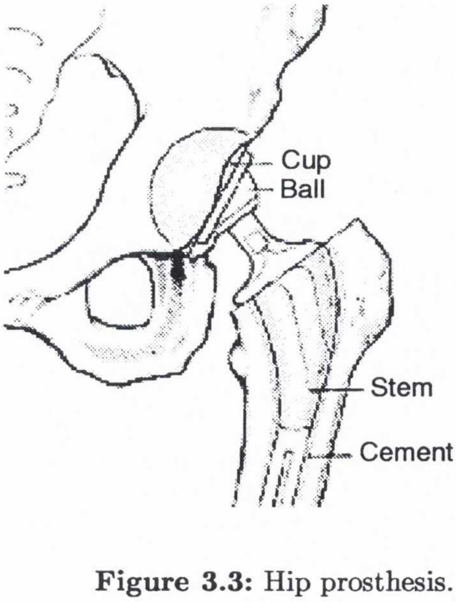


Figure 3.3: Hip prosthesis.

with conservative treatments e.g. use of a cane, anti-inflammatory drugs, or modification of activities [27]. Conditions that may warrant joint replacement include common disorders such as osteoarthritis, traumatic arthritis, osteonecrosis, congenital hip disease, and rheumatoid arthritis as well as less common conditions such as developmental and growth abnormalities, tumours, hip fractures, and other arthritic conditions [27]. In addition, the procedure may be used to treat traumatic injury of the hip.

Although improvements in preoperative planning, surgical technique and prosthetic design have significantly improved the clinical results achieved with THR, the elapsed time between the operation and the mechanical failure of joint components is still unacceptably short in very many cases. Follow-up studies indicate that the predominant mode of failure has been symptomatic aseptic loosening[1] of the prosthesis components in both cemented primary [17, 124, 93, 22, 69, 1, 95], cemented revision [71, 35, 123], cement-less primary [102, 13, 106, 3, 46, 120, 74], and cement-less revision [76, 49, 54, 36, 45, 63] surgery. The highest incidence of loosening occurs in revision and second revision cases, leading almost inevitably to further revision or re-constructive surgery for the patient. The high revision rates arising from aseptic loosening of existing implants, combined with the increasing number of patients who have complex bone stock loss which precipitates loosening, is expanding the demand for re-constructive surgery and extending the waiting lists for

[1] Aseptic loosening refers to loosening in the absence of living pathogenic organisms. The most likely course of events is that wear particles—polyethylene from the acetabular bearing surface, acrylic cement abraded from the cement-bone interface or even metal debris from elsewhere—collect in the joint and migrate into the cement-bone interface from the periphery. These small particles activate macrophages at the interface into inflammatory responses of local bone resorption—or lysis, thereby gradually debonding cement and bone.

THR. The desire to give hip implants to younger patients who have suffered accidents or who have congenital hip malfunctions is also putting pressure on demand. In these circumstances the performance of hip joints has become an issue of both medical and public concern and there is general acceptance that the status quo is not tolerable.

Many fundamental design issues remain to be resolved. The following sections attempt to summarise the literature where leading exponents take opposing positions on some very basic design considerations.

3.2 CEMENTED OR CEMENT-LESS?

At the most basic level of whether to use a cemented or a cement-less joint there remains a lack of certainty. A theoretical framework for explaining joint loosening has not been decisively established. Since the development of THR, the orthopaedic research community has pursued both cemented and cement-less designs for achieving good long-term fixation. For the elderly patient, cemented THR has a low incidence of loosening and is generally accepted as the standard method of treatment for reconstructing the hip. For the younger, more active, heavy, or bone stock deficient patient, the decision as to whether a cemented or a cement-less hip joint is preferable has not yet been clearly determined. Current research on the loosening of cemented components is extensive and embraces investigations into cementing materials, surgical techniques and component designs [14]. Simultaneously, improvements in the fixation of cement-less joints are being sought through the development of biological in-growth surfaces,[2] and press-fit designs.

It has been suggested that the choice of method of fixation should be decided on a case by case basis [14]. As against that there is an argument for using the cement-less approach for younger patients, where revisions may be inevitable. Sometimes in revision surgery the cement from the primary replacement can be very hard to remove and as a result the cemented components can be quite difficult to extract. In addition, the extra interface of a bone/cement/metal composite introduces additional complexity to the design. While bone in-growth occurs in the cement-less hip joint there is however a greater incidence of bone resorption. Recent applications of dual-energy X-ray absorptiometry have shown that the amounts of bone resorbed around cement-less stems may be quite substantial, and that the resorptive process continues even several years post-operatively [37, 75].

It has been demonstrated experimentally that some cement-less femoral stems simulate, more accurately than cemented femoral stems, the bone stress distribution which is characteristic of a healthy joint [56, 134]. Huiske's theoretical model [59] also supports this finding. The overriding criterion for successful cement-less stems is that they must achieve a quality of fit sufficiently high to limit micro-motion of the implant. The preliminary clinical evidence for some cement-less designs suggests that these designs are less prone to loosening, but may have a higher reported incidence of thigh pain [46, 15]. Long-term clinical follow-up is needed however, to fully establish the operating performance of cement-less designs, because empirical data currently available do not provide conclusive evidence.

[2] The attachment of porous materials to living tissue by tissue in-growth.

3.2.1 Loosening of the Prosthesis

Loosening of the joint components is first suspected when a patient reports pain in the joint. A radiographical study of the joint can then be used to detect evidence to support the loosening hypothesis. The radiographical criteria for indication of loose components can vary subtly between authors. In general, the indications are an increase in radiolucent lines (to 2mm or greater) between the cement and bone, cement fracture, and subsidence of the implant [118]. Quantification of the radiographical evidence can be carried out using the zonal analysis methods proposed by Gruen [44] for the femoral or stem component and the method of DeLee and Charnley [26] for the acetabular or cup component (see Figure 3.4). Although a strong correlation between radiographical evidence of loosening and symptomatic loosening exists, several authors report that false positives and false negatives can occur [22, 106, 63, 69, 77, 14]. Eventually, if failure of the joint occurs, the looseness of the component can be definitely confirmed at the time of surgery by applying sub-physiologic loads [5].

It has not been unambiguously established whether the acetabular component loosening precedes, accompanies, or is merely an indication of femoral component loosening. The frequency of reported loosening depends on whether the replacement is primary or revision, cemented or cement-less, or some permutation of these of methods. Morrey [93] constructed a mathematical model from the follow-up of several thousand cemented total hip replacements and showed that the incidence of stem loosening also depended on the age of the prosthesis. The loosening of the acetabular component was

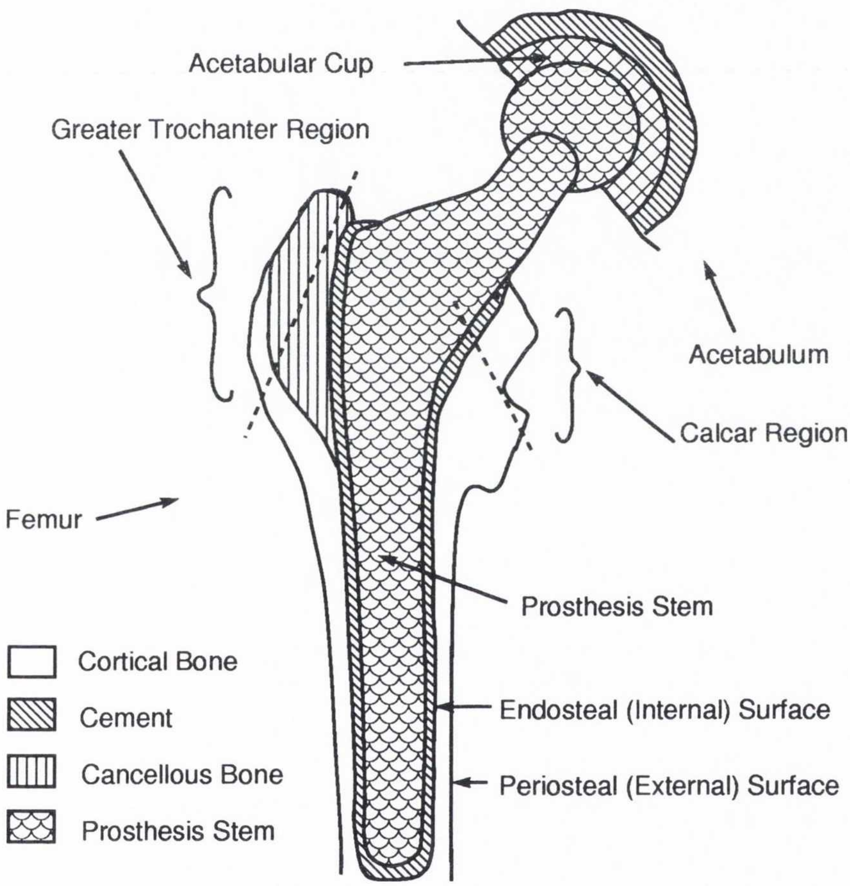


Figure 3.4: Artificial hip joint diagram.

relatively rare prior to the fifth postoperative year, but increased exponentially after the eighth year [93]. For 333 Charnley total hip replacements, performed at the Mayo clinic, it was found that the probability of loosening increased at a linear rate with the age of the prosthesis and that there was no indication of an exponential increase [69]. For cemented THR in very young patients, Chandler [17] reported that acetabular component loosening occurred twice as often as femoral component loosening. Standardisation of the different clinical analyses is necessary, however, if the test results are to be effectively compared [70].

Component loosening in cemented designs may be a function of many factors including surgical technique, biomechanics and biological mechanisms. Initially, it was presumed that insufficient loading of the host tissue was the primary factor, but emerging clinical evidence now suggests that joint failure can be induced by the presence of significant amounts of material debris in the joint. Schmalzried et al. [117] suggested that the femoral component loosening was mechanical in nature and resulted from the de-bonding and fracturing of the cement. The presence of small particles of high density polyethylene migrating along the cement/bone interface for acetabular components has led others to conclude that the mechanism for the cup failure may be biological in nature. Brien discovered high levels of metal in the synovial fluid of cemented titanium-alloy implants [9] in cases where loosening had occurred, but this leaves the question unresolved as to which came first the loosening or the debris. The degree to which degeneration is biological as opposed to mechanical has not been established because the specific factors, which determine cemented joint loosening, remain to be clarified.

In cement-less designs, the specifics of the loosening process are also uncertain. It has been proposed that inadequate preparation of the femoral canal [104], under-sizing of the femoral component [45], incomplete apposition of the stem to the bone [74], insufficient stability [53], the presence of significant amounts of bio-material debris [135] and retained cement, in the case of revision surgery [63], can all contribute to the mechanical instability of joints.

In short, the nature and causes of loosening are very poorly understood and this seriously impairs the clinical development of hip prosthesis replacement.

3.3 STANDARD VERSUS CUSTOM PROSTHESIS DESIGN

The question often arises as to whether an established clinical rationale exists for the use of custom implant components as opposed to standard off-the shelf systems. Given the considerable cost and timeliness advantages of off-the-shelf systems what evidence is there of the clinical superiority of customised components? This is another area of femoral design where convincing scientific evidence in favour of one or other system is not available. It is conceded that custom manufactured femoral components may be justified in the case of fracture, abnormal anatomy, complicated bone stock, or revision surgery, but the advantages of such prostheses in more routine circumstances has been questioned [3, 16]. Amstutz presents clinical results with standard, off-the-shelf, press-fit prostheses of superior design, for an average follow-up period of 28 months. He argues that, given the good results obtained for relatively complex problems, the use of custom femoral components is not warranted. On the other hand, Bargar is an advocate of custom designs and argues that the anatomy of each patient's femur is highly individual [5]. An interesting issue that may influence the future trend in this matter is the fact that custom designed components increase the legal responsibility of the surgeon and the engineer.

In the case of standard off-the-shelf designs, the implant size is chosen by placing two-dimensional templates over a roentgenograph of the joint at the time of surgery. The accuracy of this method has been questioned [15, 26], but it is popular, inexpensive, and generally produces good results for simple cases. Plastic trial components can also be used during the operation to evaluate the available sizes and to select the best implant. Careful preoperative planning can reduce the time spent under surgery and the associated amount of time the patient has to be anaesthetised. However, it is difficult to completely predict the complexity of the reconstruction and some decisions about choice from available range will still need to be made at the time of the operation.

Several different approaches have been tried to improve the fit of off-the-shelf femoral components. Many of the standard implant systems have an expanded range of implant sizes and THR instrumentation to provide a broad spectrum of prostheses for the surgeon to choose between. These may be based on data extracted from extensive studies of cadaver femurs [106, 110] and are known as anatomical shapes. More recently, modular components are being considered to provide further variability. Ultimately the complexity of the case may warrant a unique or custom design based on the geometry of the patient's femoral anatomy and in that event modern computer-aided design and manufacturing techniques have made it possible to produce unique femoral stem geometries.

3.4 ISSUES IN THE DESIGN OF FEMORAL STEMS

This section concentrates on the femoral component used in total hip replacement. However, many of the implications drawn apply equally to the acetabular component.

3.4.1 General Design Goals

The primary bio-mechanical design goals for the femoral component are a stable implant with no primary support point failure, well fixed in-growth surfaces, and physiological load transfer. Criteria for bone preservation, prosthesis extractability, and optimal joint geometry must also be considered.

Analytical methods for quantifying each of these criteria have not yet been fully developed and verified. Numerical experiments which model the load state of the implant-bone system, through the use of three-dimensional Finite Element Analysis (FEA), can provide considerable insight into the ability of a prosthesis to achieve the desired loading configuration [61, 72], but the optimal load state of a femur with a stem inserted has not been identified.

The natural load state of the patient's healthy femur can be taken as the baseline and, if only one hip joint is degenerated, numerical simulations on the normal femur provide useful data. Stauffer [124], however, points out that intra-medullary[3] fixation is not a natural load state for the femur, and that the biological response to this is unpredictable. Structural analysis can also be performed with fully three-dimensional bone-implant FEA but this analysis requires very intensive computations, even with simplified models.

The quality of the bone tissue at the interface will critically affect load transfer characteristics and ultimately the stability of the joint. The local loading of the tissue and the relative motion between the bone and the implant are the main mechanical parameters which affect the interface stability [72]. In addition, it is generally recognised that good compressive loads across the in-growth surfaces are needed to facilitate the initial biological fixation.

Extractability of the femoral prosthesis stem is an additional requirement. For an anatomic stem,[4] extractability is generally guaranteed because undercuts are not present in the geometry and the surgical protocol specifies the removal of the greater trochanter for implanting the stem. For more complex designs, extractability of the joint can be calculated using adaptations of robotic path verification algorithms [29].

3.4.2 Design Trade-Offs

The design of the femoral component is marked by a series of trade-offs between conflicting design criteria. One of the best observed is the trade-off

[3]Within the bone marrow.

[4]An anatomic stem is a stem whose shape matches the shape of the femoral cavity.

in cement-less THR between the strength of the stem and the level of stress induced in the bone. This can be tracked by monitoring the resultant bone deposition or absorption associated with changes in the load state of the bone tissue. Increasing the size of the femoral component may increase the resistance to fracture of the stem, but can lead to an increase in the stress in the bone and this in turn can lead to resorption or loosening [125]. Bobyn et al. [8] confirmed that increasing the rigidity of the stem increases the amount of stress shielding[5] of the bone.

A second trade-off, and one more difficult to analyse, exists between achieving good proximal loads and preventing relative motion of the implant. Consider a prosthesis that is designed to completely fill the endosteal canal of the femur. Motion at the interface would be effectively limited purely by the geometrical constraint (this function is normally handled by the bone cement). If the stem is reduced, and the degree of contact between stem and bone surface falls below one hundred percent, equilibrium may not be obtained and motion of the implant may begin. The stability of the femoral stem will depend on the direction of the implant loads and on the material constraints at the load transfer sites. Three point contact in several planes can be used as an approximated criterion to determine if equilibrium is sustainable, but only an estimate of relevant planes is possible. In addition, designing for the static bone configuration may not be very accurate when considering the contact tolerances necessary for biological fixation. The shape of the femoral cavity may not remain static over the life of the design

[5]Bone grows in proportion to the stress placed on it. Anything which distorts the normal flow of stress through the bone is termed stress shielding.

and this can start a cycle of poor loading, bone remodelling, and relative motion.

Another well-recognised trade-off is between the amount of bone material removed at the time of surgery (to shape the femur to accept the implant) and maintaining the structural integrity of the femur. Bargar [5] suggests that the removal of the soft cancellous tissue is not so critical, but that removal of the harder more dense tissue should be avoided. The disadvantages of reaming endosteal bone listed by Bargar include: (1) weakening of the supporting bone, (2) creation of stress concentrations, (3) non-uniformity of stress transfer, and (4) the potential for undesirable long-term remodelling. On the other hand a marginal amount of additional reaming might improve the quality of the contact significantly in the case where a small bone spur may exist in the canal, out of the range of vision of the surgeon. Contact with the hard bone may have indicated the need to stop reaming when in fact it would be better to remove the small spur.

A more quantifiable understanding of these trade-offs is needed to improve the THR design process.

3.4.3 Defining the Fit of the Prosthesis

The question as to what exactly is meant by fit in the context of hip implants and how the fit of the femoral stem affects the mechanical strength of a hip joint are still largely unresearched areas. The variability of existing design criteria is perhaps best reflected in the fact that over sixty-two different off-the-shelf stems are available on the commercial market in the UK alone [97].

The variations include: shapes ranging from anatomical to modular; surface treatments ranging from beaded to sintered meshing; location and extent of stem in-growth surfaces ranging from partial to fully coated; and stem shapes that vary from straight to highly curved. In addition, geometric features such as flutes, notches and collars vary from design to design. The material used can be metal alloys, ceramics, or composites. Finally, the surgical techniques for implanting the femoral stem can vary significantly.

From a non-technical viewpoint, the fit of an object can be defined as its ability to perform its intended function. Within the context of mechanical engineering, fit commonly refers to how closely two surfaces match and the concept is primarily treated as a geometric question. In the bio-engineering literature most authors define femoral component fit anatomically i.e. as the match between an implant surface and a geometric representation of the endosteal surface of the femur [111, 116, 3, 127, 67]. For THR, however, the fit of the femoral stem will ultimately be measured by how well the design performs in clinical practice but there is very little empirical data on how the geometric fit relates to complications such as thigh pain [74, 15], implant micro-motion [53], long term fixation and positive bone adaptation [134, 59]. Knowledge of how the joint biomechanics relate to the total hip pathogenesis needs to be improved. Quantitative relationships between quality of joint fit and subsequent failure have not been established.

3.4.4 Current Limitations in Measuring Prosthesis Fit

The ability to acquire more knowledge about component fit is critically limited by difficulties in collecting sufficient radiological information to describe changes in the bone-state after the implant has been inserted. The length of time needed to complete a clinical trial is another major complicating factor.

Although advances in computed tomography (CT) imaging facilitate an anatomically precise description of the patient's hip prior to surgery, the ability to image the bone-joint configuration after surgery is considerably more limited. The image is degraded by artefacts[6] which appear as starburst streaking[7] radiating from the metallic orthopaedic hardware. In addition there are other technical complications referred to as beam hardening, partial volume effects, scatter and aliasing [111]. For titanium implants, the artefact issue is not too problematic, but for cobalt-chrome stems, the artefact problem is a significant one. In the latter case a blurring of the bone-prosthesis interface makes it difficult to take measurements in critical regions.

Robertson has investigated the use of techniques to reduce the difficulty with artefacts [110]. Two-dimensional imaging modalities like DEXA scanning can provide a high quality image of the bone/implant system, but the associated tissue remodelling is inherently a three dimensional problem that needs to be tackled with three dimensional techniques. Research in the area of extended CT may improve the geometric contours generated but a com-

[6]Fragments of the implant.

[7]Starburst streaking is the descriptive term given to the impression artefacts cause on a CT.

pletely quantitative method for validating design choices is some way into the future.

With regard to the length of time needed to complete a clinical trial it is generally accepted that at least a 5 year follow-up is needed for femoral prostheses and an 8–10 year follow-up for acetabular components [93]. This is a long time to wait and it can be difficult to track patients over such an extended period. Examples of the difficulties which such a long time frame cause for THR research can be found in the literature [103]. Following up patient records can be very frustrating because very little information about bone changes since the time of the implant is typically recorded. Quantitative information is not collected until the time of failure. Substantive progress in identifying the mechanisms responsible for joint failure may have to await the availability of good post-operative, intermediate and long-term three-dimensional information from clinical trials. The most far-reaching follow-up to date is that carried out in Sweden where since 1979 all re-operations[8] after THR have been recorded.

3.5 EMERGING DESIGN CRITERIA

Efforts to improve the current state of knowledge continue. Hospitals which see several hundred custom joint cases each year may construct a custom joint design centre, like the Dana Centre for Orthopaedic Implants at the Hospital for Special Surgery in New York City. If internal facilities are not available it is possible for the surgeon to develop custom designs in association

[8] Any new hip operation on a patient who has previously undergone a total hip replacement on the same side.

with one of the industrial sites such as Bio-met. When the design engineer works directly with the surgeon in discussing the joint designs the chance of a satisfactory outcome is enhanced.

Notwithstanding the large number of unresolved issues that still obtain a number of generic rationales for designing femoral prostheses are now emerging. These can be summarised [11] as: designs which give priority to bone (fit-and-fill); designs which give priority to the prosthesis; and designs which identify priority regions on the implant where loading is required. It is proposed to evaluate each of these rationales and to describe their theoretical and empirical bases. The rationales are considered for cement-less femoral prostheses. The addition of cement would further complicate the discussion.

3.5.1 Fit-and-Fill Method

The fit-and-fill rationale proposes that the femoral component should be designed to replicate the geometry of the endosteal canal of the bone, while allowing for eventual extractability. This approach gives priority to the bone. The fit of a femoral stem is generally defined as the amount of contact between the surface of the implant and the surface of the canal. Fill is defined as the volume of space occupied by the implant as a percentage of the corresponding volume of canal. Subtle variations in how fit-and-fill is quantified exist in the literature and these can affect interpretation of the data.

When using two dimensional X-ray films, the fit is approximated at several levels on the projection data [12, 3, 44]. For volumetric studies, most authors [114, 126, 51] assess each quantity at several points along the longitudinal axis of the femur, using cross-sectional data as shown in Figure 3.5.

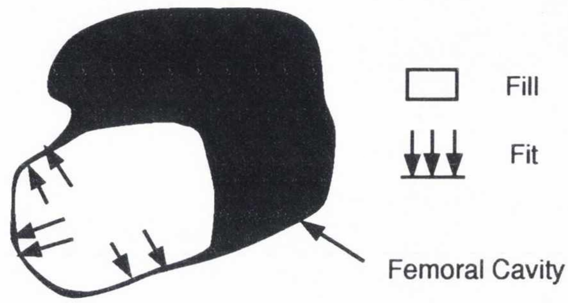


Figure 3.5: Quantification of fit-and-fill.[8]

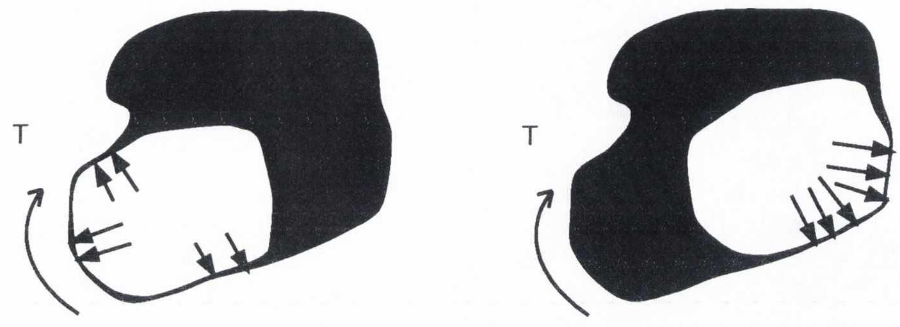
There are many reasons why the fit-and-fill approach can fail to generate adequate designs. One of the primary reasons is the potential geometric error in constructing contours to build up the endosteal canal representation. The model is typically constructed from trans-axial CT scans and the precise location of the endosteal canal is determined by subjective interpretation of "threshold" colour to delineate the boundary. A geometric and material phantom can be used to calibrate the data but subjective interpretation of where the boundary exists is still necessary. In addition, the finite discretization of the bone geometry and the polygonalization scheme used to represent the contour can also introduce error. Mean accuracy of measurements for both proximal and distal zones has been reported as 0.8mm ($\pm$0.7mm) by Rubin et al. [115]. When designing for intimate contact, errors of this magnitude can lead to poor contact between the in-growth surface and the bone.

[8] A trans-axial cross section of the proximal femur with a femoral stem is shown. Several approximations of the fit-and-fill have been suggested. In two-dimensional analysis, fill is generally defined as the percent of the endosteal area occupied by the implant cross-sectional area. The fit is defined as the implant-bone line contact expressed as a percentage of the total implant perimeter length.

A more practical disadvantage of contour generation is the demands it places on the design engineer's time, who has to work within the limitation of the current partially automated methods for extracting contours. Hauser [52] found that contour extraction for relatively normal femurs can take 1–2 hours and, for more complex cases, in the order of 2–3 hours. Similar time demands have been reported elsewhere [38]. In most cases, the design engineer must adapt the data and guide the process personally, because a completely automated algorithm does not exist. Operator intervention time is of the order of 60%, so faster processors can only have a moderate impact on the time required.

From the mechanical perspective, the fit-and-fill method does not necessarily produce a design with a state of equilibrium adequate for the stability of the implant. For example, a femoral component is shown in Figure 3.6 depicting two different femoral stem positions. The fit (area of contact between the bone and the prosthesis) is the same for both cases, but only the position in Figure 3.6a will satisfy equilibrium. For this example, it is assumed that the contact distribution identified in a single cross-section is typical of that which obtains throughout the proximal femur. The distribution of the contact points is just as important as the degree of contact: the nearer the numbers of contact points approaches complete contact around the perimeter of the implant the greater the stability of the design. However, the contact achieved at the time of surgery, with current implantation

[9] Two positions of the same implant are shown for the proximal femur. The endosteal surface contour shown has been extracted from a proximal CT trans-axial scan. The criteria for fit-and-fill would be the same for each position, but criteria for equilibrium is better satisfied by Figure 3.6a.



(a) 30% contact. (b) 30% contact.

Figure 3.6: Equilibrium of the femoral stem.[9]

methods, does not always result in what was planned for.

In revision THR, the generation of a contour can be severely impeded by the corruption of the reported CT value, due to the metal particles introduced by the orthopaedic hardware. In addition, cement debris and abnormal geometry can make detection of the contour a very approximate process. This situation can place a heavy burden on the experience and know-how of the engineer. In addition, for complex revision cases, a closed contour in the proximal femur may not be attainable due to a shift of the bone bed, so the definition of fit-and-fill has to be interpreted flexibly.

The fit-and-fill approach also suffers from another important limitation. The load carrying capacity of the femur is described not only by the characteristics of the canal itself, but also by the material density and distribution of the bone volume that defines the canal [34, 72]. A drawback of the anatomical design, for femoral prostheses, is the weak correlation that exists between pairs of measurements taken from the endosteum[11] and the periosteum[12] [98]. Despite the importance of matching the shape of the proximal femur, few published studies have examined endosteal geometry in sufficient detail for use in the design of femoral components. It has been suggested instead that implant designs may be based upon relationships derived from periosteal femoral geometry extrapolated to describe the shape of the endosteal cavity. The shape of the internal canal does not necessarily reflect the shape of the femur. It is also possible to have the same canal shape and size for two different femurs with different bone material distribution and density. These different bone configurations may have significantly different behaviours in the loaded state. Surgical technique may also vary this relationship in terms of how much reaming and broaching of the cavity is carried out. Thus the fit-and-fill approach is insensitive to the overall construction of the femur and only considers the shape of the femoral cavity.

The fit-and-fill approach is now under scrutiny by the orthopaedic community: Horne [57] is a particularly vocal critic. More scientific research needs to be conducted before the claims of a fit-and-fill approach can be justified.

3.5.2 Prosthesis Priority Method

The second generic rationale for designing femoral prostheses is the category of design that gives priority to the implant. When operating according to

[11]The endosteum refers to the medullary membrane; a thin membrane lining the inner surface of bone in the central medullary cavity.

[12]The periosteum is the thick fibrous membrane covering the entire outer surface of a bone except its articular cartilage.

this philosophy, the surgeon starts with the prosthesis shape and alters the bone bed to fit the implant. In primary replacement, this approach generally satisfies the design goals (see Section 3.4.1) and minimal modification of the bone suffices. For more complex reconstructions, such as revision THR, the shaping of the bone may compromise the integrity of the bone bed: in that event the prosthesis priority approach may not be appropriate. It is not always clear where the boundary line between bone priority and prosthesis priority lies. However, the removal of bone can only undermine the viability of further reconstruction operations.

3.5.3 Load Transfer and Bone Contact Quality Method

A third generic rationale for the design of prostheses is the load transfer and bone contact quality approach. In this case, priority regions on the implant, where loading is desired, are identified. Investigators at the Hospital for Special Surgery in New York City have developed a cement-less hip prosthesis that is designed to maximise metaphyseal[13] contact to give optimum load sharing [109]. The New York City study identified the three areas shown in Figure 3.7 as priority contact regions for the surface of the anatomic stem. The first region is located on the medial wall, wrapping around the anterior face and extending slightly to the posterior. This helps the implant to fit well within the "C" shaped region and thereby resist torsional loads. Several authors have emphasised the need for the stem to have adequate rotational

[13]Metaphysis refers to the region between the shaft and the head of the bone.

fixation to prevent rotational loosening [106, 99, 107]. The second region identified is located on the lateral side and is intended to resist forces transmitted along the line of peak joint load. The third region is located laterally on the distal stem to prevent rocking due to the bending moments. The degree of regional contact thus obtained should satisfy the equilibrium conditions needed for stability, in addition to providing good contact for the in-growth surfaces.

This approach does not differ greatly from the fit-and-fill approach if complete fit-and-fill is achieved and the priority regions are well contacted. Typically, 100% contact is not achieved with the load transfer method at the time of surgery and some compromises have to be made with fixed shape femoral components.

What remains to be understood is how the fit of the femoral component will change under various loading conditions and where does the bone/implant contact take place when the joint is fully loaded. It should be noted that most of the preoperative planning that is currently being done incorporates a static bone model. Within the Cornell University-Hospital for Special Surgery group, theoretical models are now being developed to answer questions on the dynamic nature of joint loading and how the shape of the component affects load transfer. This is expected to lead to a better understanding of successful implant design and particularly of dynamic loading conditions.

[13] Anterior view of the femoral stem shown in a graphic representation of the proximal femur. The priority regions for the surface of the anatomic femoral stem are identified. Three zones are chosen to establish three-point contact.

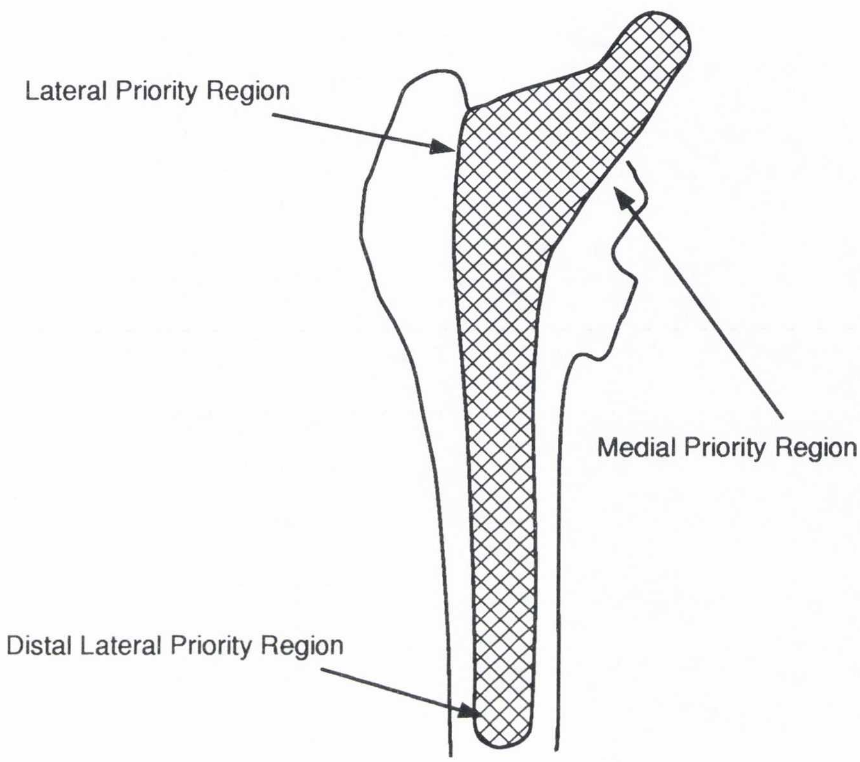


Figure 3.7: Priority regions for loading.[13]

3.6 DISCUSSION AND CONCLUSIONS

This chapter summarised the empirical background to the loosening of hip implants and then explored the complex issues and current difficulties encountered in prosthesis design. The chapter concluded with an overview of some of the more progressive prosthesis design philosophies that are found in the literature but concentrated on the fit-and-fill technique.

To match the evolution of established prosthesis design philosophies, a portfolio of evaluation techniques has emerged. These techniques include a range of mechanical and geometric testing and computer models designed to predict stresses and strains in the loaded prosthesis. This testing is now mainly confined to the design and prototyping stage of new and re-engineered prostheses.

In the next chapter it is proposed to develop and test a computer application that can combine and automate a range of evaluation techniques to test patient/prosthesis combinations at the time the patient presents for surgery.

It is confidently expected that this additional procedure will improve the quality of hip prostheses for most patients over and above that achieved when testing is carried out on a generic rather than on an individual basis.

Chapter 4

Synthesis of Current THR Evaluation Methods

> You should think of problem-solving paradigms as possible ingredients, not as complete solutions. In creating particular problem-solving systems, you may never use any paradigm by itself. Instead, you will mix them together, developing your own blends tailored to the problem domains you face [137].
>
> *Winston*

The rich variety of pre-clinical and clinical evaluation methods, used mainly at the prosthesis design stage, is not adequately reflected in the progress to-date in the matching of the correct prosthesis to the individual patient. The proposition explored in this chapter is that the overall selection process can be significantly improved by developing and deploying a computer application to synthesise a number of these evaluation methods.

Until now evaluations have been concentrated *ex-ante* on prospective new designs of prosthesis (pre-clinical) and *ex-post* on evaluating these designs after many examples have been inserted (clinical). In the approach being pursued in this chapter, an attempt is made to move the emphasis to an evaluation of specific patient/prosthesis combinations as an inherent part of THR surgical preparation and procedure. The justification for doing this was the conviction that a critical factor in limiting the success of THR is the inability of the surgeon to fully exploit and co-ordinate all the knowledge available to him about both patient and prosthesis.

The challenge was to harness the widespread availability of computing power and emerging knowledge-based techniques to make available to the surgeon, in the period just prior to the operation, the capacity to do a thorough evaluation of all the relevant data. The proposed system was intended to be labour-intensive rather than capital-intensive, comprising principally a PC, a scanner, and the associated software, and to be within the reach of all orthopaedic centres. Thus the evaluation methods would be accessible all centres and not just to larger teaching hospitals and to areas of orthopaedic research concentration.

The proposed computer application consists of prosthesis and patient consultation databases and initially three evaluation models expandable to five.

With respect to the data, they fall into two categories: data which are specific to the individual patient/prosthesis combination, and data about similar patient/prosthesis combinations which have taken place historically and from which inferences about the likely success or failure of the present combination can be drawn.

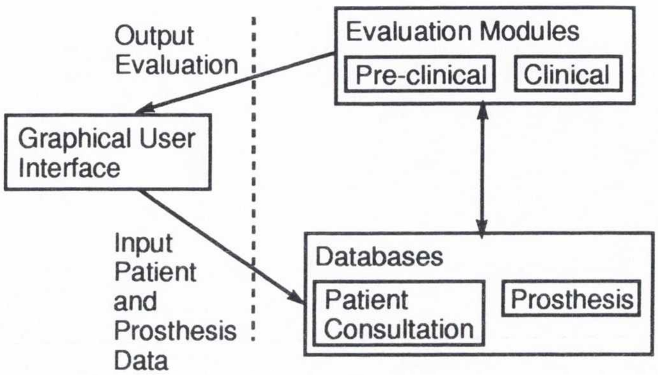


Figure 4.1: Simple representation of application structure.

With respect to the evaluation modules, they can be divided into pre-clinical and clinical. Pre-clinical evaluations are concerned with the mechanical and biological properties of implants, cements and other materials used in the hip operation. These evaluations have been traditionally carried out with generic femurs, independent of specific patient variables. Clinical evaluations are inferences and conclusions made *ex-post* on the basis of physical and radiographic examinations of patients post-operatively. A period of at least 5 years is needed for satisfactory clinical trials for femoral prostheses.

The approach taken in this thesis is to bring the analytical power, and the benefits of both pre-clinical and clinical evaluation, to the point where they can operate on all the relevant data, namely just prior to the hip operation; and to bring the lessons to be learned from these evaluation modules closer to where they can have maximum impact.

A simple representation of the application structure is shown in Figure 4.1.

The next section will describe in some detail the evaluation methods chosen and the reasons why some other evaluation methods were not considered at this stage.

4.1 OVERVIEW OF POTENTIAL EVALUATION METHODS

Evaluation methods fall into two categories, pre-clinical and clinical.

4.1.1 Pre-clinical Methods

A range of pre-clinical testing is now widely used at the research and design stage to identify potential problems with the chemical and material properties of proposed new implants and bone cements. Animal models, notwithstanding their relatively high cost, are used in these tests whenever feasible [81]. When it comes to evaluating individual patient/prosthesis combinations, however, minimal use is made of these powerful evaluation methods. It must be conceded at the outset that all pre-clinical methods cannot be adapted for use in the short period before the operation but it can be shown that two pre-clinical methods are particularly suited for performing "hands-on" evaluation.

4.1.1.1 Fit-and-Fill (Geometric) Analysis

Fit-and-fill analysis is capable of being adapted for use in evaluating patient/prosthesis combinations prior to an operation. The shape and dimensions of the prosthetic components largely determine the load-transfer mechanism and stress patterns in THR [60]. In a cemented THR, shape determines the thickness profile of the cement mantle.[1] In non-cemented THR, the shape of the prosthesis determines the fit. Shape and dimensions are complex entities in designing THR components because of natural variations in the bone geometry individual of patients [98]. Optimal fit was highlighted earlier as a design objective for prosthetic components, and many manufacturers advertise advantages for their products in this respect (such as anatomic shape, optimal fit, proximal fit, canal filling and the like). If the fit-and-fill analysis is carried out prior to the operation the chances of obtaining a satisfactory fit is substantially increased as opposed to a fit-and-fill analysis carried out at the design stage of the prosthesis, when calculations must be based on a standard femur and the wide variation in femoral dimensions (see Table 4.1) is not taken into account.

Fit-and-fill analysis is also a tool for pre-clinically testing the likelihood of bone in-growth failure occurring, due to gaps and relative motion at the implant-bone interface in non-cemented THR. It can also address the stress-bypass scenario, which develops in the bone around the femoral stem, when proximal load transfer is bypassed in favour of distal load transfer. This can lead to a situation where the proximal femur is under-stressed and becomes subject to strain-adaptive resorption. Geometric analysis can allow all these factors to be legislated for immediately before the operation.

When fit-and-fill analysis is carried out at the prosthesis design stage a well-established procedure exists. The design prototype is placed in a series

[1]The cement mantle is the cement between the prosthesis and the bone.

of specimen femoral bones, and the gaps between implant and bone—or the thickness profile of the cement mantel for cemented THR—are evaluated [58]. The data from experimentation performed by the manufacturers in the design stage are hardly ever published and therefore are not available to the surgeon performing the operation. For these and other reasons, it is very desirable for fit-and-fill analysis to be carried out again just prior to the operation.

The fit-and-fill process must be adapted to make it appropriate for use in the short time period available before the operation. This necessitates using and manipulating computer images of the prosthesis and the femur as opposed to the physical items. Instead of a three-dimensional model of a generic femur it is necessary to use the dimensions of the actual femur. The femoral dimensions required to construct a computer diagrammatic representation of the patient's femur (see Figure 4.2) are specified in Table 4.1. These dimensions are read from the relevant patient's radiographs and inputted into the patient database. The image of the patient's femur is then constructed from these dimensions and superimposed on images of different prostheses, selected from the range contained in the prosthesis database. The surgeon manipulates the image of the prosthesis over the image of the patient's femur to attain the best fit for that prosthesis.

Two measures of fit-and-fill are calculated for all the prostheses in the range for the patient's femur. These measures correspond to the antero-posterior and lateral views and are combined to give an overall measure of fit-and-fill. The superimposed images of the prostheses and the patient's femur can then be used to perform FEA of the competing patient/prosthesis combinations.

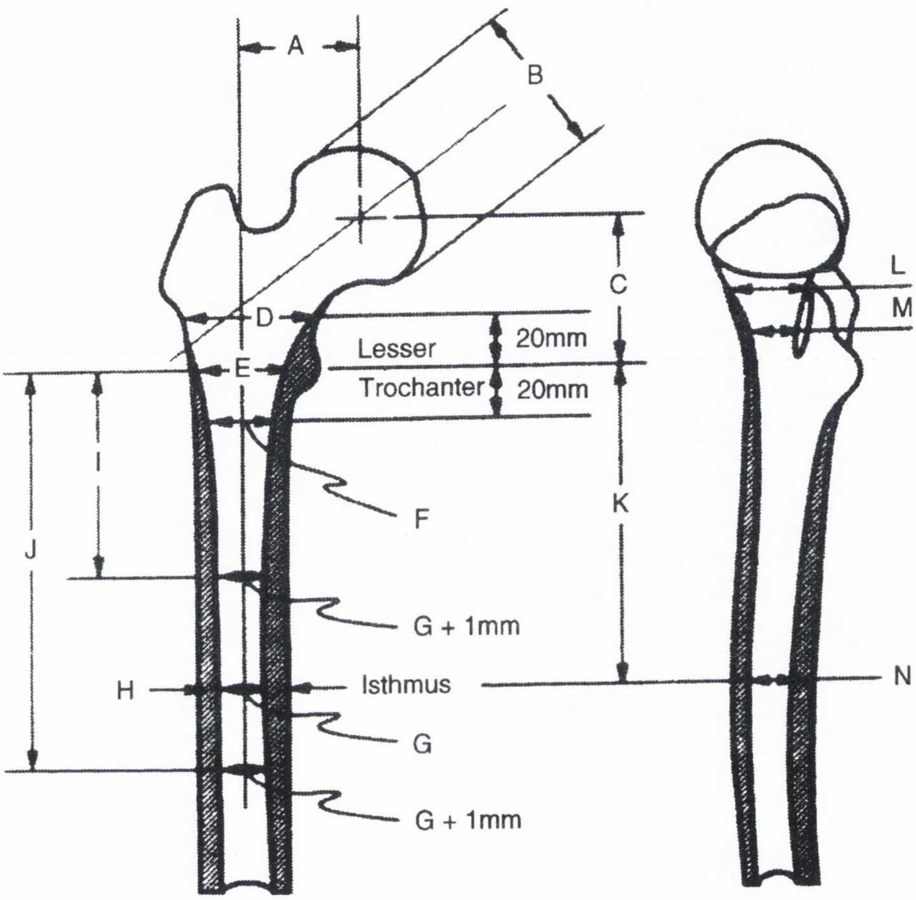


Figure 4.2: Diagrammatic representation of the standard dimensions of the femur in the anteroposterior and lateral views.

The fit-and-fill analysis just described can be automated and made much more user-friendly. Computed axial tomography (CAT scans) could provide three-dimensional images of both the prosthesis and the patient's femur. An algorithm could be written to automate the fit-and-fill calculation and the fitting of the prosthesis to the femur could be simulated and graphically validated. These enhancements are feasible and desirable but outside the scope of the current research.

Dimension	No. of Specimens	Average	Standard Deviation	Minimum	Maximum
Femoral head offset (A)	200	43.0mm	6.8mm	23.6mm	61.0mm
Femoral head diameter (B)	200	46.1mm	4.8mm	35.0mm	58.0mm
Femoral head position (C)	200	51.6mm	7.1mm	32.8mm	74.3mm
Canal width (lesser trochanter +20) (D)	200	45.4mm	5.3mm	31.0mm	60.0mm
Canal width (lesser trochanter) (E)	200	29.4mm	4.6mm	17.0mm	41.9mm
Canal width (lesser trochanter −20) (F)	200	20.9mm	3.5mm	11.0mm	29.5mm
Isthmus width (mediolateral) (G)	200	12.3mm	2.3mm	8.0mm	18.5mm
Extracortical width (mediolateral) (H)	200	27.0mm	3.1mm	20.5mm	36.0mm
Proximal border of isthmus (I)	200	86.1mm	17.8mm	37.0mm	199.0mm
Distal border of isthmus (J)	200	145.0mm	19.4mm	92.0mm	205.0mm
Isthmus position (K)	200	113.4mm	16.4mm	63.0mm	157.0mm
Anteroposterior canal width (osteotomy level) (L)	99	24.1mm	3.1mm	15.5mm	31.0mm
Medial diameter of femoral neck (M)	99	16.5mm	2.9mm	10.0mm	22.5mm
Isthmus width (anteroposterior) (N)	196	16.9mm	3.5mm	10.0mm	27.0mm
Neck-shaft angle	200	124.7°	7.4°	105.7°	154.5°
Femoral length	132	436.8mm	35.3mm	353.0mm	523.0mm
Effective neck length	200	35.5mm	2.6mm	30.0mm	41.1mm
Age (years)	100	79.9	13.0	22.0	95.0

Table 4.1: Variables which characterise femoral specimens. The wide variation in variable values is evident [98].

4.1.1.2 Finite Element Analysis

FEA is a design evaluation and pre-clinical test method that is normally applied before prototypes of the THR components are actually made. But there is no reason why FEA should not be applied in the period immediately preceding the operation to evaluate a particular patient/prosthesis combination. Given information about shapes and dimensions of bones and implants, elastic properties of the materials, bonding conditions of the implants and external loads, an FEA computer model can predict the stresses and strains in the loaded THR [60]. In this way, the load-transfer phenomenon can be documented and compared for alternative designs. The mechanical design assumptions and performance objectives of alternative patient/prosthesis combinations can also be checked. In addition, stress values can be compared with strength data on implant materials, bone and interfaces, in order to estimate the probability of failure.

Constraints on time and data necessitate that the complexity of the FEA model used immediately preceding an operation be substantially less than that of models used at the design stage of a prosthesis. Initially, the FEA can be carried out on the two-dimensional superimposed images of the prostheses and of the patient's femur already constructed in the fit-and-fill module. However, with advances in computing power, in FEA algorithms, and in the learning experience of surgeons, this gap will be substantially eliminated and three-dimensional images will be used.

FEA of THR was achieved only after considerable effort was devoted to its development: in the early years, little was known about joint loads, bone properties and damage accumulation processes in the materials concerned. FEA of biological structures has, however, made enormous progress during the past decade. The features of FEA methods themselves and the capacity of computers are now sufficiently advanced for valid FEA models. Great progress has also been made in the analysis of hip-joint loads and bone properties. Three-dimensional FEA models are used to predict with reasonable accuracy, stresses, strains and even interface motions in THR structures. In this way, the probability of long-term failure due to accumulated damage, bone remodelling and absorption can be estimated pre-clinically.

The data that are required to enable the computer application to carry out FEA, fall into two categories: first, the dimensions of, and positional relationship between, the prosthesis and the patient's femur (taken from the fit-and-fill analysis); and secondly the forces (estimated from the patients weight, height and activity level) operating on, and the material properties of, the prosthesis and the patient's femur. All these data can be assembled for a particular patient/prosthesis combination and then formatted as an input file to facilitate the execution of the required FEA.

4.1.1.3 Alternative Pre-clinical Methods

There are a number of other pre-clinical evaluation methods, which are not easily adapted for use in the short time-frame before an operation. The special nature of these tests and the reasons why they are not included in the proposed computer application are touched upon briefly.

Mechanical testing using static and dynamic loading can test the mechanical behaviour of prosthetic implants in the laboratory. Different load axes are selected when testing strength and endurance. The laboratory bench test also measures fatigue. Endurance tests are also carried out for wear on the connections in modular prosthetic components [62] but this test is only needed in the design stage: the properties of an individual patient's femur would not affect the outcome.

Tribological testing of the articulating surfaces is possible and several test benches are now on the market. Dynamic hip simulators simulate the motions as well as the forces on the hip joint in order to measure articular wear. Accurate measures of wear rates and wear particles can be made pre-clinically. Tribological testing, however, is not an option in the short time-frame before an operation: this is a destructive test that is done on samples and takes a long time. In any case, since the wear propensity is not a function of variations in the properties of individual femurs, the test is not relevant at this stage.

Finally, animal experimental models can be used to pre-clinically test devices for all failure scenarios. However, animal total hip replacement differs in many respects from its human counterpart—such as shape, dimensions and properties of bones, as well as loads—so the actual devices to be tested in animals cannot be used in humans. Nevertheless, particular innovative aspects of prosthetic components can be tested in this way. Examples of this are new materials for prostheses or fixation methods. In recent years, there has been much progress in the development of standardised models for this purpose [129]. Animal experiments bear no relation to individual patient femurs and are not an option just prior to an operation.

4.1.2 Clinical Methods (Hip Scores)

Turning now to clinical methods that can be adapted for use in the short period available to the surgeon before the operation. Clinical evaluation is predicated upon evidence and inferences from patient data, accumulated over extended time periods. The data collection must be uniform and capable of being replicated.

The outcome of total hip replacement is measured clinically either by survival analysis or hip scoring methods. Survival analysis assesses only one attribute of hip replacement namely the proportion of THRs each year that do not fail, whereas hip scores assess many different criteria aimed at achieving some composite measure of how well a hip replacement is functioning.

The hip scoring technique which is essentially a method of capturing all the richness and lessons of past experience, is something which, either formally or informally, must be part of all THR evaluation methodology. Hip scores may be based on clinical or radiological criteria, or both [94]. Hip scoring must be conducted at regular intervals in order to chart the progress of the hip over time and in this way meaningful analysis can be subsequently carried out on the hips in the database. Before the operation, a hip score indicates the reasons for THR, and after the operation, the success of the THR is chronicled.

There are a large number of different hip-scoring systems, and most of them have several variations. In essence they are designed to produce an objective score of how a particular hip has performed post-operatively. An attempt is made to quantify a number of different factors, which may be clinical (pain, mobility, function) or radiological. These are then added together to give a composite score. The great advantage of hip scores is that a number is generated which can be manipulated statistically. However, as the scores do not produce normal statistical distribution curves the validity of using statistical methods to analyse them is open to question. Arbitrary levels are set, so as to define hips that are considered to be excellent, good or poor. As the different hip scoring systems may produce very different results, comparing results using different scoring methods can be problematic. Repeatability can also be a problem, even using the same hip scoring method for similar patients; different authors get different results.

4.1.2.1 Hip Scores Based on Clinical Criteria

A number of options were considered before finally choosing a hip scoring system for the computer application. Merle d'Aubigne and Postel [25] proposed a 0–6 scale for registration of each of the following attributes: pain, walking ability, walking aids and motion. This system was later modified by Charnley [18], who pointed out the importance of a classification system for describing the degree of walking impairment. Harris [50] presented a different scoring system for evaluating the results of hip replacement which enabled the status of the hip to be described with a single number in the range 0–100. The factors assessed in this method are pain (total score 40), function (total score 47), range of motion (total score 5) and absence of deformity (total score 8). The Harris hip score has been widely used and because of its widespread acceptance it was chosen for the research model (see Figure 4.3).

I. Pain (44 possible)
  A. None or ignores it (44)
  B. Slight, occasional, no compromise in activities (40)
  C. Mild pain, no effect on average activities, rarely moderate pain with unusual activity, may take aspirin (30)
  D. Moderate pain, tolerable but makes concessions to pain. Some limitation of ordinary activity or work. May require occasional pain medicine stronger than aspirin (20)
  E. Marked pain, serious limitation of activities (10)
  F. Totally disabled, crippled, pain in bed, bedridden (0)
II. Function (47 possible)
  A. Gait (33 possible)
    1. Limp
      a. None (11) b. Slight (8)
      c. Moderate (5) d. Severe (0)
    2. Support
      a. None (11) b. Cane for long walks (7)
      c. Cane most of the time (5) d. One crutch (3)
      e. Two canes (2)
    3. Distance that can be walked
      a. Unlimited (11) b. Six blocks (8)
      c. Two or three blocks (5) d. Indoors only (2)
      e. Bed and chair (0)
  B. Activities (14 possible)
    1. Stairs
      a. Normally without using a railing (4) b. Normally using a railing (2)
      c. In any manner (1) d. Unable to do stairs (0)
    2. Shoes and socks
      a. With ease (4) b. With difficulty (2)
      c. Unable (0)
    3. Sitting
      a. Comfortably in ordinary chair one hour (5)
      b. On a high chair for one-half hour (3)
      c. Unable to sit comfortably in any chair (0)
    4. Enter public transportation (1)
III. Absence of deformity points (4) are given if the patient demonstrates:
  A. Less than 30 degrees fixed flexion contracture
  B. Less than 10 degrees fixed adduction
  C. Less than 10 degrees fixed internal rotation in extension
  D. Limb-length discrepancy less than 3.2 centimeters
IV. Range of motion (index values are determined by multiplying the degrees of motion possible in each arc by the appropriate index)
  A. Flexion 0-45 degrees * 1.0 C. External rotation in ext. 0-15 degrees * 0.4
    45-90 degrees * 0.6 over 15 degrees * 0
    90-110 degrees * 0.3
  B. Abduction 0-15 degrees * 0.8 D. Internal rotation in extension any * 0
    15-20 degrees * 0.3 E. Adduction 0-15 degrees * 0.2
    over 20 degrees * 0

To determine the over-all rating for range of motion, multiply the sum of the index values * 0.05. Record Trendelenburg test as positive, level, or neutral.

Figure 4.3: Synopsis of the Harris hip score evaluation system [96].

4.1.2.2 Hip Scores Based on Radiographic Evaluation

A second clinical technique, which is also an integral part of the pre-clinical fit-and-fill analysis, is radiographic evaluation. As well as diagnosing the condition of a patient's femur, an important use of radiographs before the operation is to provide the patient's femoral dimensions (see Table 4.1), essential for fit-and-fill and FEA analyses. Measurements from the radiographs of prosthesis loosening in the patient's femur, taken at regular intervals after the operation, in conjunction with a clinical hip score system, are an essential element in producing an overall measure of success for a particular patient/prosthesis combination (see Figure 4.5). Hip scores based on radiographic evaluation were not incorporated into the research model at this initial stage.

The standardisation of positioning of the patient and the harmonisation of film-focus distance and exposure rates are critical to obtaining a correct interpretation of serial radiographs. A close teamwork between the radiologist and the orthopaedic surgeon is necessary to reach this goal.

Conventional Radiographic Analysis Gruen et al. [44] are the authors of an established method of radiographic analysis for loosening of the femoral component. They divided the proximal femur into seven zones. This division was carried out by dividing the femoral stem into thirds (see Figure 4.4). Lateral to the proximal third is Zone I. Lateral to the middle third is Zone II. Lateral to the distal third is Zone III. Distal to the prosthesis is Zone IV. Medial to the distal third is Zone V. Medial to the middle third is Zone VI and medial to the proximal third is Zone VII. These areas are analysed

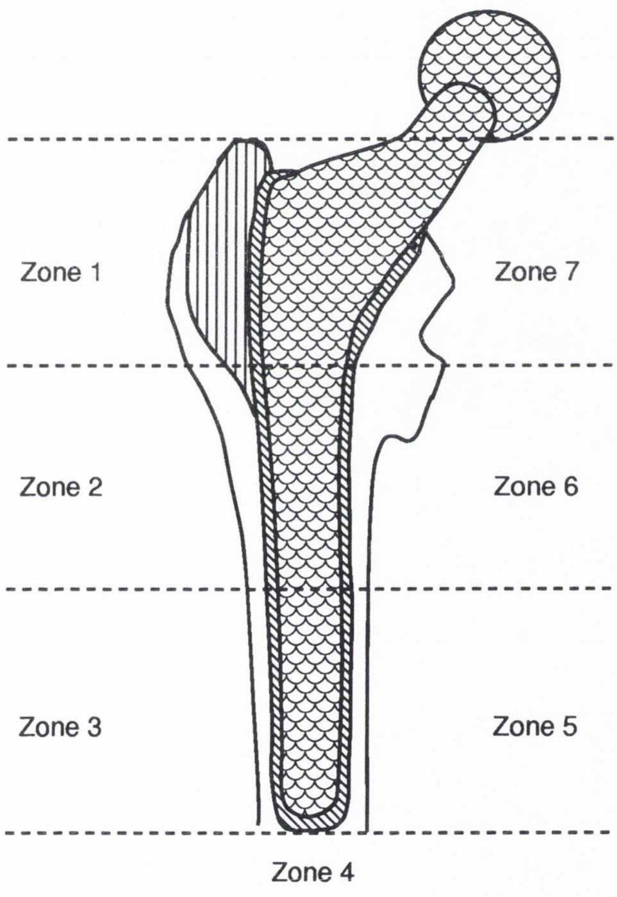


Figure 4.4: The regions used in the analysis of the stem-interfaces.

for acrylic cement fracture and a radiolucent zone at the stem-cement and cement-bone interface. Radiographs are evaluated chronologically to assess loosening as manifested by progressive changes in the width or length of the radiolucent zones; appearance of sclerotic bone reaction; widening of the acrylic cement fracture gap; and fragmentation of the cement and gross movement of the femoral component. This method of analysis is widely used for recording femoral loosening.

Clinical assessment of hip replacement is difficult and radiological assessment is even more so. The Gruen classification is probably the best but loose hips frequently demonstrate more than one failure mode. As failure is a dynamic and continuous process, a single radiological snapshot may well be misleading.

Another area of contention is the behaviour of certain geometric designs of hip replacement. The double tapered polished collar-less Exeter hip replacement undergoes Gruen Type 1A failure in 70% of cases, yet the aseptic loosening rate after 18–20 years is approximately 2%. American authorities, particularly W.H. Harris, consider the Exeter hip replacement to be a failure because of its migration. Clinically, however, the results of the Exeter stem are amongst the best reported in the literature [96].

The systematic recording and evaluation of femoral loosening in patient/prosthesis combinations is important when trying to assess the propensity for loosening in proposed combinations. However, to be successful, it requires the availability of a large database of cases from which to search for comparable examples. Such a database does not yet exist in Ireland. Its inclusion in Figure 4.5 is by way of a proposal on how such a radiograph evaluation may be incorporated into the process of THR evaluation. If the radiological data were to be collated from many different centres, this would expedite the process of setting up the database.

Alternative Radiographic Evaluation Methods There are many new developments in radiographic evaluation [119, 30, 37]. Because of the complexity and expense of these methods, they are not yet widely available and thus inappropriate for inclusion in an application that seeks to have widespread appeal. However, as these developments achieve more general acceptance, their inclusion in the application may become viable.

4.1.2.3 Survival Analysis

It is important for the surgeon to know and utilise historical information about the life span of particular prosthesis designs for various categories of patient. Hip survival analysis for a variety of prostheses has now been carried out for several decades by the Swedish National Hip Arthroplasty Register [1, 82, 83].

Survival analysis measures the proportion of THRs each year that do not fail. For each year these proportions are cumulated. The cumulated proportions of those surviving are calculated and plotted against years since operation. These calculations are done with corrections for the irregular nature of the sampling [28]. There are a number of problems with survival analysis. The three main problems relate to patients who are lost to follow-up, analysis with small numbers and the definition of failure.

Each year a number of patients with THR are withdrawn from the trial.

These include both patients that have died and those patients who have been lost to follow-up. A fundamental assumption in survival analysis is that the group of withdrawals has the same failure rate as the group that has not been withdrawn. This is probably a valid assumption for the patients that die. It is less likely to be valid for those patients who are not followed up although there is some evidence to support this assumption [31]. It is essential that the number of patients lost to follow-up is included with the survival analysis. This may be shown numerically or graphically.

In survival analysis only very few patients tend to be followed up for long periods. Therefore, the confidence limits[2] are likely to be very large at the end of the follow-up. Survival curves may, therefore, give a false impression unless they include standard errors or confidence limits. It is important that some indication of errors is included and these are probably best calculated with an error equation [105].

A fixed end-point is chosen for the analysis. This is usually revision of the prosthesis. Another end-point is the development of a particular radiological sign. The decision to revise a prosthesis depends on many factors. These include the fitness of the patient, the length of the waiting list and how aggressive the surgeon is. A revision is therefore not necessarily a good criterion on which to base survival analysis, even though it is easy to measure. Some authors have attempted to overcome this by including patients in severe pain; this is difficult to quantify. Survival analysis does not take into account

[2]Confidence limits are the end points of a confidence interval. A confidence interval is a group of continuous or discrete adjacent values that is used to estimate a statistical parameter (as a mean or variance) and that tends to include the true value of the parameter a predetermined proportion of the time if the process of finding the group of values is repeated a number of times.

how well the prosthesis is functioning. For example, two prostheses with the same revision rate would, apparently, perform equally well when investigated with survival analysis, even though one had a higher incidence of thigh pain than the other.

The statistical approach to survival analysis which is currently used requires a statistician and is not adaptable for use in the period just prior to the operation. Chapter 7 looks at the possibility of using neural networks to perform survival analysis for selected patient/prosthesis combinations just prior to the operation.

4.1.2.4 An Alternative Clinical Method

Gait analysis has now been developed to the point where it can be applied routinely to provide objective information about hip function and indications of prosthesis loosening, whereas earlier there were only subjective indications which could be used [2]. A loose implant will provide evidence of instability in the force patterns at the hip joint during gait, even before the patient has developed clear signs of pain. Gait analysis is not widespread and thus data on the outcome of this technique are not readily available. Gait analysis is not practicable for use as an evaluation method just prior to an operation.

4.2 COMPUTERISING THE SELECTED EVALUATION METHODS

A European Commission Report [39], quoted at the outset of this thesis asserted that there is great potential for the application of the information technologies to the health sector. There appears to be an obvious opportunity to design a computer application to automate at least some of the evaluation techniques discussed in the previous section. This application would be placed at the disposal of the orthopaedic surgeon to improve the selection of a prosthesis to suit a particular patient's input variables. The surgeon would have the capacity to evaluate competing patient/prosthesis combinations prior to carrying out the operation. A very important by-product would be the automatic collection of data on the hip replacement operation, something that is not routinely carried out in Ireland.

At the outset the application consisted of five modules: FEA, fit-and-fill analysis, clinical hip score, radiograph analysis, and survival analysis. A weighted selection method is used to arrive at a prosthesis rating appropriate for the patient. Other evaluation methods can be incorporated at a later date. The structure of this application is represented in Figure 4.5.

The application prototype was developed using *Visual Basic™2.0*. The clinical hip score and the fit-and-fill analysis components were an integral part of the application. The radiographic analysis component was not fully implemented: an inadequate database of radiographic records meant no feedback was possible. A similar situation pertained to the survival analysis component. Survival analysis data is not recorded in Ireland. The FEA component

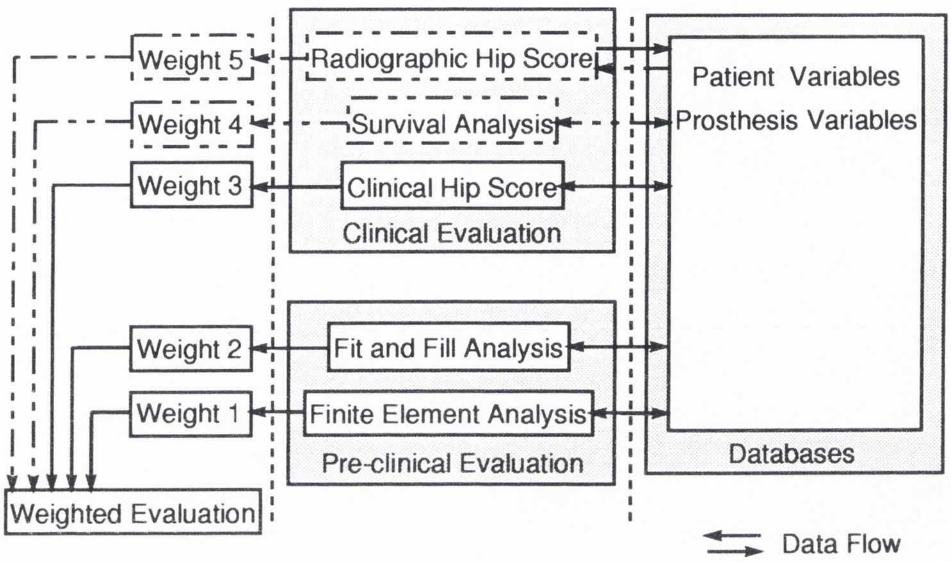


Figure 4.5: Flow of knowledge in patient/prosthesis combination evaluation application.

was separate from the application. The application supplies the appropriate input file to the FEA package (see Section 4.1.1.2). The FEA package carries out the analysis and produces an output file containing the calculated stress values. This output file is imported back into the application, which in turn produces a representation of the stress patterns set up by the particular patient/prosthesis combination. The FEA capability, initially stand-alone, can be incorporated into the application at a later stage. A decision was also made to use two-dimensional images and modelling and to confine the analysis to the femoral component.

4.2.1 Application Interface

The application presents the surgeon with a graphical user interface (GUI). This GUI prompts the user to enter data into the various databases. Data is entered into the databases in one of two ways: directly, when keying in the femoral dimensions of the patient and indirectly, when the results of a FEA for a particular patient/prosthesis combination are automatically entered into the database.

4.2.2 Evaluation Modules

4.2.2.1 Fit-and-Fill Analysis

The application initially measured fit-and-fill by comparing a two-dimensional image of different prostheses with a two-dimensional image of the patient's femur. The two views used were the anteroposterior and the lateral views (see Figure 4.2). These correspond to roentgenograms obtained using a standardised technique that provides views parallel and perpendicular to the plane of the femoral neck. The technique is capable of expansion to three-dimensional images but this adds enormously to the complexity of the model. The surgeon chooses the prosthesis to be evaluated from a database of prostheses and displays two-dimensional anteroposterior and lateral views of the prosthesis chosen. This is then compared with the corresponding view of the patient's femur that has been constructed, based on femoral measurements inputted in the patient database. The femoral measurements (see Table 4.1) used for this exercise are those proposed by Nobel et al. [98] for characterising femoral specimens. The surgeon moves the image of the patient's femur over the image of the prosthesis, rotating it as required, to gauge the fit-and-fill of the prosthesis. After the surgeon has successfully oriented the prosthesis in the femur, the application calculates measures of fit-and-fill for the patient/prosthesis combination. Prostheses are then ranked according to this measure. Although fit-and-fill is of less importance in the case of cemented prostheses, the surgeon can still usefully judge the thickness of the cement mantle. The surgeon also gains important insights into the best method of insertion of the prosthesis into the patient's femur.

This fit-and-fill process is predicated upon the surgeon's expertise rather than on a rigid algorithm. In all cases, however, the prosthesis is checked first for overall compatibility, conformity of crucial measurements, and approximate fit. This can substantially reduce the number of prostheses to be examined by the surgeon.

4.2.2.2 Finite Element Analysis

When the image of the patient's femur has been satisfactorily placed on top of the image of the prosthesis, the resulting superimposed image is then used as the basis for FEA. An appropriate input file for the FEA package is prepared by the application; this includes the property values of the different materials used in the prosthesis, taken from the prosthesis database, as well as dimensions from the superimposed image.

When the FEA is completed, its output, in the form of a file, is analysed and interpreted by the computer application to inform the surgeon about the levels of stress in the proposed patient/prosthesis combination. The surgeon then makes a judgement on whether this is acceptable.

Most surgeons are not, however, competent at interpreting FEA. This competence will only be achieved through a learning process involving dialogue between FEA experts and surgeons which will eventually lead to the calibration of rules/rule-based expressions to be incorporated into expert system. This issue forms the substance of Chapter 5.

4.2.2.3 Clinical Hip Score Comparisons

A key component of the computer application is its ability to record clinical hip scores (see Section 4.1.2.1). A comprehensive database of clinical hip scores collected over an extended time period is essential if meaningful conclusions, based on historical experience, are to be drawn.

The surgeon uses the inherent resources of the clinical hip score database to highlight similar patients in the past and to compare their outcomes with their treatments. The tacit knowledge which is at the basis of the surgeons ability to carry out this exercise and to make judgements on the selection of the correct prosthesis for his current patient, remains to be unearthed. The elicitation of this expert knowledge in the form of rules is the subject of Chapter 5. For the clinical hip score component of the application to successfully work, the database needs to be populated with clinical hip score data which have been collected in a uniform, repeatable manner. This means strict adherence to standardised procedures.

4.2.2.4 Weighting

The initial approach to achieve an overall evaluation for a patient/prosthesis combination was to assign to each of the evaluation output values a predetermined weight and hence to arrive at a composite evaluation. The patient/prosthesis combination would be accepted or rejected on the basis of this aggregated value. This approach proved too simplistic as this is not a linear problem.

4.2.3 Databases

The construction and maintenance of up-to-date databases is critical to the success of the application. The surgeon needs to have confidence in the data contained in these databases; the data must be reliable and capable of verification.

There are two very different databases in use (see Figure 4.6): one for data on prostheses and the other for patient data. The prosthesis database, which

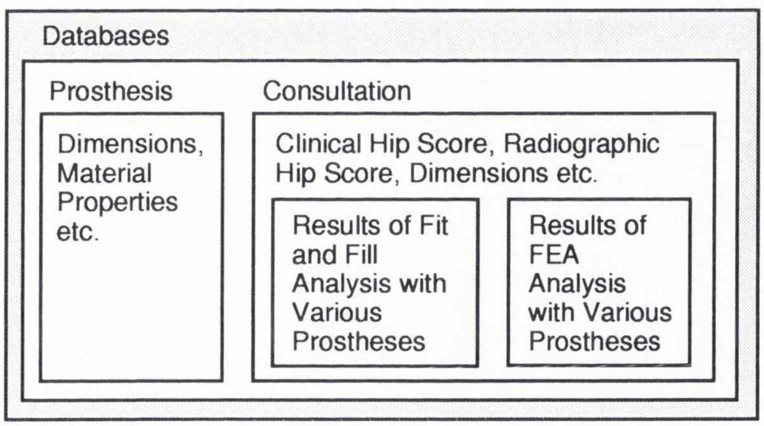


Figure 4.6: Databases in the application.

contains information relating to the different types of prosthesis available, is relatively static, in that once the information about a particular prosthesis is keyed in there should be no need to change it. Entries are made into the prosthesis database only when new prostheses are added to the range. The prosthesis database holds all information about prostheses, materials, as well as their geometry. Different sizes of the same generic prosthesis have a different entry.

The second database is the patient database. This database records the patient details every time the patient has a consultation with the surgeon: it is updated with clinical hip scores and measurements from radiography in order to track the state of the patient's hip.

The patient database is divided into two components. The first component contains the measurements of the femur highlighted in Table 4.1. These measurements are taken directly from anteroposterior and lateral roentgenograms. This measurement process is difficult to automate. Discerning the boundary of the endosteal surface and allowing for necessary reaming are factors which must be left to surgical experience. There is an added difficulty of taking reproducible roentgenograms.

The second component contains the patient clinical hip score. The Harris hip score described in Section 4.1.2.1 was used because of its widespread acceptance [47]. This database is updated whenever the patient returns for a consultation. Both components of the patient database are time dependent and are updated when the patient comes for consultation (hopefully at regular intervals).

4.3 CONCLUSIONS

- Significantly more emphasis should be given to the evaluation imperative at the time when the prosthesis/patient selection decision is made i.e. just before the operation. It is only then that the full facts become known about the patient, the appropriate surgical procedure and prosthesis availability.

- A portfolio of evaluation techniques, clinical and pre-clinical, is now available and these are capable of being put at the service of the surgeon, using the power of modern computing.

- A computer-based decision model was designed and constructed to bring together the outputs of three key evaluation techniques, all of which utilise and operate on shared data contained in patient and prosthesis databases. The evaluation techniques are: fit-and-fill analysis; finite element analysis; and clinical hip score. This portfolio can be expanded to include radiographic hip score and survival analysis.

- This computer application is at a relatively early stage of development and a number of the separate evaluation techniques require further refinement and automation, to make them user-friendly and attractive to surgeons working under severe time-constraints.

- Surgeons, on the other hand, will require training and familiarisation with the quantitative evaluation techniques, particularly those such as FEA that have their origins and widest application outside the medical domain.

- A major limitation of the prosthesis/patient evaluation computer application, as proposed in this chapter, is its failure to address the issue of how to weight the separate evaluation outputs to arrive at a composite evaluation score. The alternatives are to give fixed weights to the three evaluation outputs (based on the judgement of some "superior" expert) or to give each surgeon discretion in the allocation of weights. The first alternative is unacceptably rigid and fails to capture the nuances inherent in surgical decision-making and the second defeats one of the main purposes of the application which is to move the THR process towards a measure of standardised best practice.

- A rule based expert system is proposed in the next chapter to deal with this deficiency.

Chapter 5

THR Rule-Based Expert System

We are drowning in information, but starving for knowledge.

John Nasibett

This chapter describes a prototype rule-based expert system that attempts to address the limitations of the approach to evaluation described in the previous chapter. The program uses ordinal scales[1] to represent the qualitative reasoning of experts. The prototype was tested on the data of half a dozen patient/prosthesis combinations. The results were encouraging but some limitations were also uncovered.

[1]Variables measured on an ordinal scale have order relationships. The values that can be taken on by an ordinal variable can be placed in a unique order. "Greater than" and "less than" have meaning. But the actual numerical values of ordinal variables ***do not convey any information whatsoever beyond the order itself.***

5.1 BACKGROUND

Major potential benefits accrue when a rule-based expert system is added to the computer application:

- The insights and judgements of the expert are added to and built into the decision-making process.
- The time spent by experts on trivial and straightforward cases is reduced thus making scarce expert time available for more challenging work.
- The service to the patient is improved by providing a better patient/prosthesis match and shortening the time required for a thorough evaluation.
- A rule-based expert system is a training tool for surgeons who can ask it to justify its decisions.

Rule-based systems are costly and time-consuming to develop. However, because of the scale of money involved in THR, even a modest improvement in the decision-making process can have considerable financial benefits and can also significantly reduce human suffering.

The following section describes the rule-based expert system prototype, outlining its major components and the programming environment. The descriptions are in terms of actual Prolog code to emphasise the close relation between the representation of knowledge in the program and THR concepts themselves. The reader unfamiliar with Prolog is referred to [19] for details of the language. However, the features of the program can be understood without a knowledge of the programming language.

5.1.1 Prolog as a Development Language for Rule-Based Expert Systems

An extensive search was made to identify an appropriate programming environment for the prototype. The possibilities included Lisp, Prolog, one of the available "off-the-shelf" expert system shells or a "conventional" programming language such as Algol, Pascal or C. Eventually the choice fell on Prolog which was deemed to provide the fastest and cheapest way to write the prototype.

The choice was very fortunate. Prolog is a very powerful tool for symbolic computation and can deal efficiently with the necessary arithmetic. It is relatively easy, even for a new user of Prolog, to formulate the rules in a natural way and it is not necessary to deal with many syntax problems. Complex rules can be expressed quite easily in the language. Any computer-aware researcher can quickly use Prolog to write an interesting initiatory expert system without prior or special training.

With Prolog, the user is free to concentrate on issues of knowledge representation and acquisition rather than on search strategies. This allows rapid feedback to be presented to the experts, maintaining their interest and commitment.

Another factor contributing to rapid feedback is the natural program development style of the language. One can naturally develop a top down program. This feature is very important when writing rule-based expert systems in so far as rule-based expert systems cannot be defined in advance. Typically the expert does not know *a priori* how he thinks, and the system inevitably develops very much on a trial and error basis.

Another important feature of Prolog is an easy ability to incorporate meta-programming: it can be used to define any search strategy the user wishes and to write explanation facilities. This flexibility comes at a price. Meta-interpreters slow down the computation speed. Prolog still lacks the rich programming environment of Lisp, but for present purposes, its exceptional power in building rule-based expert systems compensates for these shortcomings.

The knowledge contained in expert systems can, to a significant extent, be expressed as rules. Prolog, whose basic statements are rules, is thus a natural language for implementing expert systems. Prolog is a logic programming language. Logic provides a precise language for the explicit expression of goals, knowledge, and assumptions. Thus, Prolog facilitates the construction of expert systems that explicitly represent the knowledge of human experts.

5.2 CONSTRUCTING EXPERT STATEMENTS FOR SELECTED EVALUATION FACTORS

A prototype expert system for the evaluation of hip prosthesis designs was developed to combine the three evaluation methods outlined in Chapter 4 namely fit-and-fill, FEA, and clinical hip score. This was achieved by embedding the appropriate Prolog logic-base, which will now be described, in the *Visual Basic*^{TM}*2.0* application which was described in the previous chapter.

Received wisdom [126] dictates that, when developing an expert system, knowledge engineers consult a single evaluation expert. The use of multiple experts often leads to discrepancies in the rules. In the field of hip prosthesis design, however, fit-and-fill, FEA, and clinical hip score have evolved into distinctive domains of expertise. Hence the researcher had to relax the constraint on using more than one expert. This clearly has implications for the integrity of the system.

The first important evaluation factor considered was the fit-and-fill of the prosthesis in the patient's femur. The dimensions that constitute fit-and-fill (shown in Table 4.1) are divided into three categories to reflect their relative importance when assessing fit-and-fill. In the first category are the primary dimensions, such as femoral head offset (A), where the margin of error is critical and where there is very little tolerance. Next are the secondary dimensions, such as canal width, measured at the isthmus (G), where there is more leeway in the margin of error. Finally, tertiary dimensions, as instanced by the canal width, taken at $\pm$20mm vertically from the lesser trochanter (D and F), have the largest tolerance. Table 5.1 shows the breakdown of the dimensions which were first introduced in Table 4.1 and Figure 4.2.

The second evaluation factor considered is the FEA analysis output, which is supplied, in the desired form, to the Prolog logic-base. In practice, the surgeon supplies a value which equates to his evaluation of the FEA

Primary	Secondary	Tertiary
A	G	D
C	L	F
E	M	$G + 1\text{mm}$

Table 5.1: Breakdown of dimensions by the importance of their margin of error. (The legend for these letters is contained in Table 4.1 and Figure 4.2.)

for that particular patient/prosthesis combination. The rules for ascertaining the FEA rating from the FEA stress data could eventually be incorporated into an additional Prolog module.

The third evaluation factor is the clinical hip score analysis value. This is ascertained by taking the salient variables of the patient (age, sex, primary diagnosis, activity level and weight), and looking at the resultant two year post-operative clinical hip scores of patients with like variable values, and who used the same prosthesis. For a knowledge engineer with an understanding of the field of THR, no further explanation of such concepts is necessary. An understanding of the subject domain by the knowledge engineer is a prerequisite for communicating with the domain expert. Furthermore, the knowledge engineer encodes domain concepts in the expert system in a form that is understandable by the domain expert.

Experts use qualitative terms in considering and speaking about all three evaluation factors, the fit-and-fill, FEA and clinical hip score. "The prosthesis produced a low level of resorption, or a tight fit." "The FEA shows an excellent stress profile," etc. Even concepts that could be determined quantitatively are usually represented in qualitative terms. The evaluation of hip prostheses is never likely to be reduced to simple numbers and ratios. When making judgements, experts are more comfortable with qualitative terms. To echo expert reasoning, it is imperative to model qualitative reasoning.

5.3 TRANSLATING EXPERT STATEMENTS INTO A RULE-BASED SYSTEM

On talking to the experts, it became clear that a significant amount of the relevant expert knowledge could be expressed as a mixture of procedures and rules. The rules for determining the quality of fit-and-fill and the clinical hip score analysis rating of a particular prosthesis, involve considerable calculations even after an initial screening is done to assess if the prosthesis is at all suitable.

The information described in Section 5.2 is sufficient to build a prototype. The judgements and observations based on conversations with the experts will now be translated into a rule-based system. The initial screening to determine if a particular prosthesis is worth considering is performed by the predicate[2] `ok_profile(Prosthesis)`. The top-level basic relation is `suitability(Prosthesis,Answer)`, where `Answer` is the reply to the question as to the suitability of the `Prosthesis`. The code has three modules—`fit_rating`, `hip_score_analysis`, and `fea_analysis`—corresponding to the

[2]In Prolog, the simplest kind of statement is called a *fact*. Facts are a means of stating that a relation holds between objects. An example is `father(Joe,John)`. This fact says that Joe is the father of John, or that the relation `father` holds between the individuals named `Joe` and `John`. Another term for a relation is a ***predicate***.

three important evaluation factors. There is a preliminary check that no essential data are missing (if so the system prompts the user to enter the data). The answer `Answer` is then determined with the predicate `evaluate(Profile,Answer)`, which evaluates the `Profile` built by the three modules.

Knowledge engineers stress the importance of top-level formulation; this means modular programming is essential. Each of the modules can be developed independently without affecting the rest of the system. There is no commitment to any particular data structure, i.e. data abstraction is used. For this example, a structure `profile(FitRating,HipScoreAnalysis,FEA)` represents the profile of the fit rating, the clinical hip score analysis rating and the FEA rating of a particular patient/prosthesis combination. However, nothing fundamental depends on this decision, and it would be easy to change the sequence. Consideration of the features of these modular pieces follows.

5.4 SPECIFIC EVALUATION MODULES

The essential features of the fit-and-fill rating module (see Program 5.1), which determines a rating for a particular patient/prosthesis combination's fit-and-fill, are first examined. The initial step is to determine an appropriate profile. This is done with the predicate `fit_profile`, which classifies the patient/prosthesis combination dimensions as `primary_dimensions`, `secondary_dimensions`, or `tertiary_dimensions` and gives the percentage of the overall dimensional error which is attributable to each in the particular

Program 5.1 The fit rating module.

```
/* The Fit Rating Module */
fit_rating(Prosthesis,Rating) :-
      Rating is a qualitative description assessing
      the fit offered by prosthesis to cover the request
      for suitability. */
fit_rating(Prosthesis,Rating) :-
      fit_profile(Prosthesis,PrimaryDimensions,
            SecondaryDimensions,TertiaryDimensions),
      fit_evaluation(PrimaryDimensions,SecondaryDimensions,
            TertiaryDimensions,Rating).
fit_profile(Prosthesis,PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions) :-
      requested_suitability(Prosthesis,Suitability),
      fit_percent(primary_dimensions,Prosthesis,Suitability,
            PrimaryDimensions),
      fit_percent(secondary_dimensions,Prosthesis,Suitability,
            SecondaryDimensions),
      fit_percent(tertiary_dimensions,Prosthesis,Suitability,
            TertiaryDimensions).
fit_percent(Type,Prosthesis,Total,Value) :-
      findall(X,(fit(Fit,Type),
amount(Fit,Prosthesis,X)),Xs),
      sumlist(Xs,Sum),
      Value is Sum*100/Total.
```

patient/prosthesis combination. The relation draws upon the information in the database concerning both the prosthesis and the patient. Program 5.2, which uses a hypothetical situation, indicates that the isthmus is a primary dimension and hence demands the most exacting fit.

The profile is evaluated to give a rating by `fit_evaluation`. It uses rules of thumb to give a qualitative rating of the fit: excellent, good, etc. (see Program 5.2). The first `fit_evaluation` rule, for example, reads: "The rating is *excellent* if the dimensional error of the primary dimensions in the patient/prosthesis combination accounts for less than 40 percent, primary and secondary dimensions combined for less that 70 percent, and primary,

Program 5.2 Fit module's evaluation rules and arthroplasty data.

```
/* Evaluation Rules */
fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,excellent) :-
      PrimaryDimensions < 40,
      PrimaryDimensions + SecondaryDimensions < 70,
      PrimaryDimensions + SecondaryDimensions +
            TertiaryDimensions =< 100.
fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,excellent) :-
      PrimaryDimensions < 70,
      PrimaryDimensions + SecondaryDimensions =< 100.
fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,good) :-
      PrimaryDimensions =< 100.

/* Arthroplasty Data - Classification of Fit */
fit(isthmus,primary_dimensions).
fit(femoral_head_offset,primary_dimensions).
fit(canal_width,secondary_dimensions).
fit(canal_width_plus_20,tertiary_dimensions).
fit(canal_width_minus_20,tertiary_dimensions).
```

secondary, and tertiary dimensions combined for less than or equal to 100 percent of the total error."

Two features of the code need to be explained. First, the terminology used in the program is the terminology of the expert. This makes the program (almost) self-documenting to the expert and means that he can modify it with little help from the knowledge engineer. Allowing people to think in concepts, which are specific to the domain, facilitates debugging and assists in using a domain-independent explanation facility. Second, the apparent simplicity of the evaluation rules may be deceptive. Considerable knowledge and experience are hidden behind these simple numbers. Choosing poor values for these numbers may mean severely misjudging the fit.

The clinical hip score evaluation module determines the historical performance of the particular prosthesis in patients with similar characteristics. It uses information taken from the clinical hip score database. This rating is also qualitative: a weighted sum of clinical hip score analysis factors is calculated by `score` and used by `calibrate` to determine the qualitative class. The code which implements this is shown in Program 5.3.

It should be noted that the modules both for the fit rating and, to a lesser extent, for the clinical hip score analysis rating, reflect the point of view and style of a particular expert, rather than an objective standard. Within the discipline there is not a consensus on the subject. Some experts tend to be conservative and some are prepared to take considerable risks.

Program 5.3 Hip score analysis rating module.

```
/* Hip Score Analysis Rating Module
hip_score_analysis(Prosthesis,Rating) :-
      Rating is a qualitative description assessing
      the clinical hip score analysis record offered by a
      Prosthesis to support the request for suitability. */
hip_score_analysis(Prosthesis,Rating) :-
hip_score_factors(Factors),
score(Factors,Prosthesis,0,Score),
calibrate(Score,Rating).

/* Hip Score Evaluation Rules */
calibrate(Score,bad) :- Score =< -500.
calibrate(Score,medium) :- -500 < Score, Score < 150.
calibrate(Score,good) :- 150 =< Score, Score < 1000.
calibrate(Score,excellent) :- Score >= 1000.

/* Arthroplasty Data - Weighting Factors */
hip_score_factors([(age,10),
(sex,2),
(cause,5),
(activity_level,5),
(weight,2)  ]).

score([(Factor,Weight)|Factors],Prosthesis,Acc,Score) :-
value(Factor,Prosthesis,Value),
Acc1 is Acc + Weight*Value,
score(Factors,Prosthesis,Acc1,Score).
score([],Prosthesis,Score,Score).
```

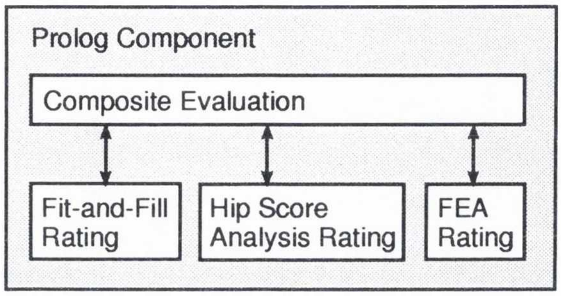


Figure 5.1: Structure of the Prolog component.

5.5 COMPOSITE EVALUATION OF PROSTHESIS/PATIENT COMBINATIONS

Programming the code for determining the fit-and-fill and clinical hip score analysis ratings was straightforward in both these cases, in that the knowledge provided by the experts was translated into program rules. These rules act on the raw input data to produce qualitative evaluations of the respective modules (just as an expert would). Similar rules are not yet established for eliciting an FEA output value. In this instance direct expert intervention is still required.

Constructing a module for the overall evaluation of the patient/prosthesis combination was, however, more challenging. The major difficulty was formulating the relevant expert knowledge. Experts were less forthcoming with general rules for overall evaluation than, for example, with rules for rating the fit-and-fill results, the code for which is shown in Program 5.2. They were happy to discuss the profiles of particular patient/prosthesis combinations, but were reluctant to generalise. They preferred to react to suggestions rather than volunteer rules.

This forced a critical re-appraisal of the exact problem to be solved. There were three possible answers that the system could give: approve the patient/prosthesis combination, refuse it, or ask for advice. There were also three evaluation factors to be considered. Each evaluation factor had a qualitative value that was a point on an ordinal scale representing a small set of possibilities. For example, the clinical hip score analysis rating could be *bad*, *medium*, *good*, or *excellent*. Further, the possible values were ranked on an ordinal scale.

The problem faced can be generalised: find an outcome from an unknown ordinal scale, based on the qualitative results of several ordinal scales. Rules to solve the problem would give a conclusion based on the output values of the evaluation modules (see Figure 5.1). When the experts were pressed with this formulation, they offered several rules of which the following is typical: "If the patient/prosthesis combination's clinical hip score rating is excellent, its clinical hip score analysis rating good (or better), and its FEA analysis at least reasonable, then approve the operation."

An immediate translation of the spoken rule above is

```
evaluate(profile(excellent,good,reasonable),give_suitability).
```

But this misses many cases covered by the expert's rule, for example, the case where the patient/prosthesis combination's profile is `(excellent,good,excellent)` would be overlooked. The expression only deals with the specific profile presented to it `(excellent,good,reasonable)`. All the cases for a given rule can, of course, be listed but it is obviously better to build a more general tool to evaluate rules expressed in terms of qualitative values from ordinal scales.

There is potentially a problem with using ordinal scales because of the large number of individual cases that may need to be specified. If each of the N modules have M possible outcomes, there are N^M cases to be considered. In general, it is not feasible to have a separate rule for each possibility. Not only is it a problem to enumerate so many rules but the search involved in finding the correct rule may be computationally prohibitive. So instead, a small set of *ad hoc* rules was defined. It was hoped that the rules selected, which covered many possibilities, would be sufficient to cover the prostheses the surgeons normally used. The structure chosen for the rules was `rule(Conditions,Conclusion)` where `Conditions` is a list of conditions under which the rule applied, and `Conclusion` is the rule's conclusion. A condition has the form `condition(Factor,Relation,Rating)`, insisting that the rating from the factor named by `Factor` bears the relation named by `Relation` to the rating given by `Rating`.

The relation is represented by the standard relation operators: <, =, >, etc. The previously mentioned rule is represented by

```
rule([condition(fit,'=>',excellent),
      condition(success,'=>',good),
      condition(analysis,'=>',reasonable)],give_suitability).
```

Another rule reads "If both the fit rating and clinical hip score analysis rating are good, and the FEA analysis is at least reasonable, then consult your superior." This is translated to

```
rule([condition(fit,'=',good),
      condition(success,'=',good),
      condition(analysis,'=>',reasonable)],consult_expert).
```

Factors can be mentioned twice to indicate that they lie in a certain range or a factor might not be mentioned at all. For example, the rule

```
rule([condition(fit,'=<',moderate),
      condition(success,'=<',medium)],
      refuse_suitability).
```

states that a patient/prosthesis combination should be refused approval if the fit rating is no better than moderate and the clinical hip score analysis rating is at best medium. The FEA analysis is not relevant and so is not mentioned.

The interpreter[3] for the rules is written non-deterministically (see Appendix A). The procedure is: "Find a rule and verify that its conditions apply," as defined by `evaluate`. The predicate `verify(Conditions,Profile)` checks that the relation between the corresponding symbols in the rule and the ones that are associated with the `Profile` of the patient/prosthesis combination is as specified by `Conditions`. For each `Type` that can appear, a scale is necessary to give the order of values which the scale can take. Examples of scale facts in the database are `scale(fit, [excellent,good,moderate])`

[3]An interpreter is the mechanism by which the code is broken down and executed.

Program 5.4 Test data for the suitability evaluation system.

```
/* Prosthesis Data */
fea_analysis(charnley_medium,excellent).
requested_suitability(charnley_medium,40).

amount(isthmus,charnley_medium,5).
amount(femoral_head_offset,charnley_medium,4).
amount(canal_width,charnley_medium,2).
amount(canal_width_plus_20,charnley_medium,10).
amount(canal_width_minus_20,charnley_medium,10).

value(age,charnley_medium,20).
value(sex,charnley_medium,10).
value(cause,charnley_medium,49).
value(activity_level,charnley_medium,9).
value(weight,charnley_medium,9).

ok_profile(charnley_medium).
```

and `scale(success, [excellent,good,medium,bad])`. The predicate `select_value` returns the appropriate symbol of the evaluation factor under the ordinality test that is performed by `compare`. It is an access predicate, and consequently the only predicate dependent on the choice of data structure for the profile.

At this stage, the prototype program was tested. Data from real patient/prosthesis combinations were used, and the answers given by the system tested against the corresponding expert opinions. The data for `charnley_medium` is given is Program 5.4. The reply to the query `suitability(charnley_medium,X)` is `X = give_suitability`.

Finally, in order to help the expert to define a consistent set of rules, the ordinal model was considered in a broader context. A consistency check was added to verify the integrity of the system. Any specific rule is a boundary condition of the problem and it has to be consistent with its predecessors. The following two meta-rules were used for consistency, and each rule must satisfy them both.

Consistency Rule 1: If all of patient A's factors are better than or equal to patient B's factors, then the outcome of patient A must be better than, or equal, to that of patient B.

Consistency Rule 2: If all of patient A's factors are equal to or between the corresponding factors of patients B and C, then the outcome of A must be equal to or between, those of B and C.

All aspects of a prototype THR decision support system are now in place. The system was tested and proved to function in a plausible fashion and now awaits site implementation and greater acceptance among orthopaedic surgeons.

This prototype expert system is a composite of styles and methods—not just a backward chaining system. Heuristic rules of thumb were used to determine the fit rating; an algorithm, albeit a simple one, was used to determine the clinical hip score analysis rating; and there is a rule language, with an interpreter, for expressing outcomes in terms of values from discrete ordinal scales. The rule interpreter proceeded forward from conditions to conclusion rather than backward as in Prolog. Expert systems must become such composites if they are to exploit the different forms of knowledge that already exist.

5.6 CONCLUSIONS

- The case can now be made that a meaningful, empirically-based rule-based expert system prototype has been developed for THR, which can contribute significantly to improving the survivability of hip implants.

- Prolog has proved to be an excellent tool for developing the prototype. Prolog, whose basic statements are rules, is a natural language for implementing expert systems. The developer is free to concentrate on issues of knowledge representation and acquisition rather than on search strategies: this insured rapid feedback to the surgeon experts.

- The structure of the program was tested and validated on a range of common prostheses applied to a selection of patients. The data input for a representative patient/prosthesis combination, using a Charnley, is shown in Program 5.4. The data for the clinical hip score analysis were constructed from the results of published papers which used the Harris hip score [47, 80, 73]. The decision output of the program (`give_suitability`, `consult_expert` and `refuse_suitability`) for all the input cases listed gave answers compatible with expert opinion.

- It is not claimed that the program as it stands is ready to go into production. But with an expansion in the number of rules, additional modules and a fuller database, there is every confidence that it can deal comprehensively with the diverse patient/prosthesis combinations that might be presented to it in practice. However, the refinement of the system requires that these enhancements take place *in situ* which is not something the present orthopaedic climate facilities.

- The rule structure of the individual fit-and-fill and clinical hip score modules and of the overall evaluation component were shown to be practical and soundly based. A rule structure for the FEA evaluation module must await fuller automation and can then be incorporated. Expert system rules, however, can only be definitively established when the system enters production. Even at that stage, continuous fine tuning is of the essence as the system moves down its learning curve.

- The inherent capability to explain its decisions was a powerful and necessary dimension of the system as developed. An expert system tries to capture the tacit knowledge and expertise of eminent practitioners but this may require compromise in the articulation of rules. Transparency is therefore necessary if the rules elicited are to prove persuasive and to achieve general acceptance.

- Excellent domain experts are the major key to the success of any expert system. But experts guard this expertise which is their distinguishing characteristic. A rule-based expert system is likely to be resisted to the extent that it attempts to codify proprietary expertise. There is no doubt that a more recent variation on expert systems, which uses fuzzy logic, is closer to the way experts express their thinking. It is conceivable, therefore, that surgeons may be more comfortable with the fuzzy logic approach.

- This refinement is pursued in the next chapter, not to challenge the validity and integrity of the expert system as developed in this chapter, but merely to see if it can be made more user-friendly.

Chapter 6

THR Fuzzy Expert System

> As the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.
>
> *Lotfi Zadeh*

The operation and synthesis of the evaluation techniques used in the rule-based expert system discussed in Chapter 5 is impaired by the imprecise nature of the way surgeons express their assessment of particular situations. The experts were much more likely to express their judgement of a particular evaluation factor linguistically rather than numerically: a method by which this limitation of rule-based systems may be overcome is elaborated on below. The inability to formulate precise statements to describe particular situations has been addressed by proponents of "fuzzy expert systems."

An additional opportunity to exploit fuzzy rules occurs in cases of incomplete data, for example where the source data is deficient and decisions must be made based on uncertain, incomplete and even contradictory information. The fuzzy logic expert system can cope with this situation also.

The essence of the problem with rule-based systems is that the criteria for measuring the different evaluation factors (fit-and-fill, FEA and clinical hip score analysis) result in the representation of the expert's decision in terms of discrete values: excellent, good, moderate, medium, bad and poor. In reality, the experts, rather than expressing the evaluation factors for a patient/prosthesis combination in these terms, preferred to use vaguer expressions such as "If the fit-and-fill is loose then the risk of failure is increased." The problem then becomes how to specify an uncertain term like loose which can range from "slightly loose" to "very loose." This uncertainty can be included in a fuzzy model as well as uncertainty associated with data limitations, for example incomplete clinical hip score data.

The uncertainty estimate for each evaluation factor is quantified by projection onto a risk curve (consequent fuzzy set) to get an associated risk of failure. The individual evaluation factor risks are aggregated to give an overall risk of failure for each competing combination. With the fuzzy approach, risk of failure becomes the selection criterion rather than finding an outcome from an unknown ordinal scale, based on the qualitative results of several ordinal scales (see Section 5.5). The risk of failure allows patient/prosthesis combinations to be classified as acceptable or not.

The fuzzy approach used in this chapter seeks to emulate a risk function. The risk for competing patient/prosthesis combinations is calculated.

6.1 REPRESENTING UNCERTAINTY: THE PATH LEADING TO FUZZY LOGIC

In the early 1960s, researchers in applied logic were convinced that theorem provers were powerful and general enough to solve practical, real-life problems. In particular, the introduction of the *resolution principle* to logic reasoning by J.A. Robinson lead to this conviction [114]. It subsequently became apparent that the appropriateness of mathematical logic (as typified by the Prolog rule-based expert system constructed in Chapter 5) for solving practical problems was overrated. One of the complications with real-life situations is that the facts and experience necessary for solving a problem are surrounded by a degree of uncertainty; moreover, the available information is frequently imprecise and insufficient. Yet human experts must form judgements and make decisions from uncertain, incomplete and at times contradictory information. To be useful in an environment characterised by such imprecise knowledge, an expert system has to capture and exploit not only the highly specialised expert knowledge, but also the uncertainties that go with such knowledge. This observation has led to the introduction of models for handling uncertain information in expert systems. Research into the representation and manipulation of uncertainty has grown into a major research area called *inexact reasoning* or *plausible reasoning*.

Probability theory was the first point of departure for the development of models for handling uncertain information in rule-based expert systems.

It was found, however, that probability theory could not be applied in this particular context without some qualifications. Two *quasi-probabilistic models* were proposed: the *subjective Bayesian method*, which was developed for application in the expert system PROSPECTOR and the *certainty factor model*, which was designed by E.H. Shortliffe and B.G. Buchanan to deal with uncertain information in the expert system MYCIN [78].

The appropriateness of probability theory for handling uncertain information in a knowledge-based context has itself been questioned [78]. The twin convictions that probability theory has little to say about inherently imprecise notions, such as the quantifiers "most" and "few," and that the application of probability theory requires "too much data," led to the search for alternative formalisms for expert systems applications. *Belief functions* and *fuzzy logic* are two such formalisms.

Belief functions are also called the *Dempster-Shafer theory of evidence.* This theory may be viewed as a generalisation of probability theory. The development of the Dempster-Shafer theory has been motivated by the observation that probability theory is not able to distinguish between *uncertainty* and *ignorance* as in when there is incomplete information. In probability theory, probabilities have to be associated with individual elemental hypotheses. Only if these probabilities are known are we able to compute other probabilities of interest. In the Dempster-Shafer theory, however, it is possible to associate measures of uncertainty with *sets* of hypotheses, interpreted as disjunctions, instead of just with the individual hypotheses, and nevertheless to be able to make statements concerning the uncertainty of other sets of hypotheses. In this way, the theory is able to distinguish between uncertainty and ignorance.

However, the Dempster-Shafer theory cannot be applied in an expert system without modification. Two basic problems preventing the use of the model in rule-based systems are first, its computational complexity and secondly, the lack of several combination functions for propagating uncertain hypotheses. With respect to the second problem, various *ad hoc* solutions have been proposed, none of which is really satisfactory [78].

Many of the theorems and methods needed when using probabilities in expert systems require the expert to estimate probabilities, sometimes without recourse to relative frequencies. There is evidence that many people do not have an intuitive understanding of the laws of probability, and so estimates are likely to be very inaccurate. The research of Kahneman and Tversky [66] showed that people are apt to discount prior odds and accord more weight to recently presented evidence. Other research suggests that people are over-confident in their own judgements [65]. Furthermore, simplification of formulae can cause gross errors in computation [51].

6.1.1 Fuzzy Logic

The need for an alternative formalism to probability theory for representing and manipulating uncertainty led to a second point of departure and a major new line of research in plausible reasoning, namely *fuzzy set theory* and *fuzzy logic.* The great advantage of fuzzy logic over probability theory for expert systems appears to be the compositionality of its logical operators [64]. The detailed workings of fuzzy logic are fully explained in Appendices B and C.

Part of the attraction of fuzzy logic is that it is grounded on the use of language. At the centre of fuzzy modelling techniques is the idea of a linguistic variable. Fuzzy sets are functions that map a value, that might be a member of a set, to a number between zero and one, indicating its degree of membership of that set. The fuzzy set LOOSE is a simple linguistic variable (see Figure 6.1) used in a rule to make a decision based on the looseness of

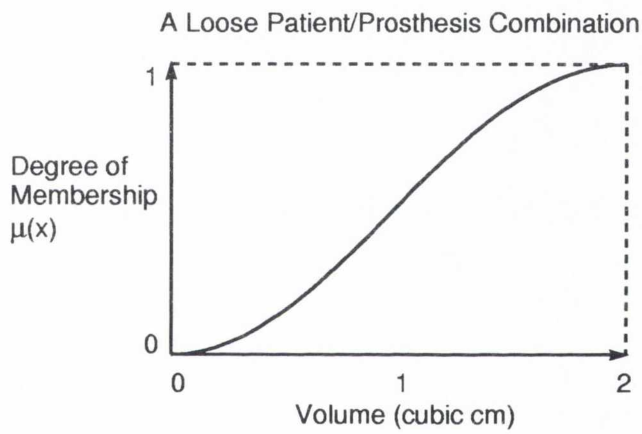


Figure 6.1: Idea of a loose patient/prosthesis combination.[1]

a particular patient/prosthesis combination:

```
if fit_and_fill_level is loose
    then operation.risk is increased.
```

Similar rules are formulated for the other evaluation techniques and this series of fuzzy rules is then aggregated to give an overall evaluation. In a fuzzy system all rules are fired or activated all the time. They fire in *parallel* and all rules fire to some degree. The result is a fuzzy weighted average.

[1]The x axis value is a measure of the space between the prosthesis and the femur.

6.1.1.1 Approaches to Aggregating Fuzzy Rules

Fuzzy models that have a large number of rules with the same consequent fuzzy set (the `INCREASED.RISK` set) have a persistent problem: namely that the solution fuzzy set quickly becomes saturated. This saturation problem is a feature of both the more established fuzzy models i.e. the min/max inference technique and the additive inference technique.

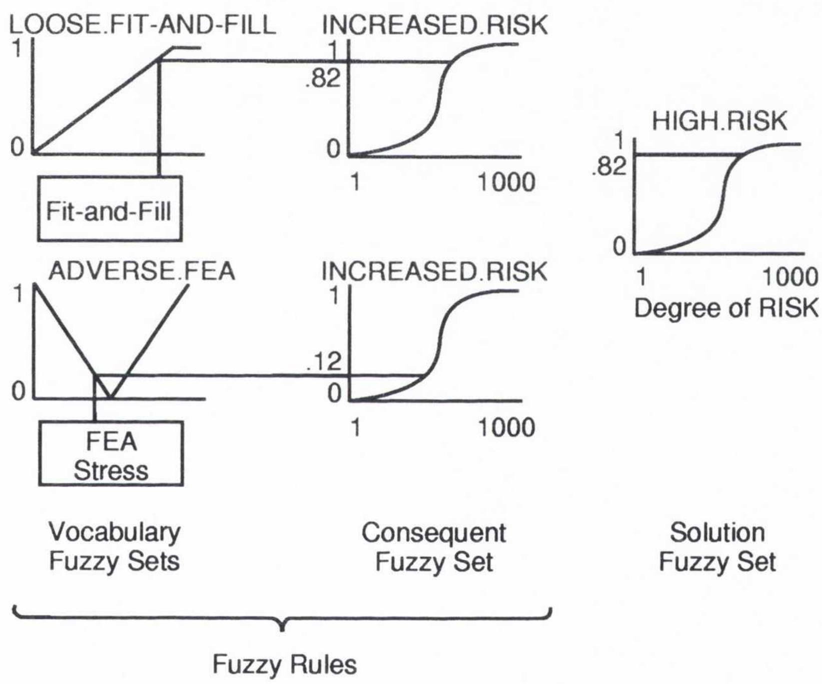


Figure 6.2: The min/max technique.

Min/max inference technique In the min/max inference technique, the solution fuzzy set is updated (for conditional `IF ... THEN` rules) by taking the maximum value of the consequent fuzzy set. This means that after a few rules, the "high water mark" in the solution set will move quickly toward 1. It also means that rules whose truth is less than the current truth level of the solution fuzzy region, will not contribute to the solution. In risk assessment models this is unsatisfactory. It is important to accumulate as much evidence as possible. If rule `[R1]` has a predicate truth of .82 then the solution fuzzy set `HIGH.RISK` contains the consequent fuzzy set `INCREASED.RISK` to a degree of .82 (see Figure 6.2). If none of the other rules has a truth greater than .82, none of them will affect the `HIGH.RISK` solution space. *Defuzzification* is the process by which a number representing the risk is obtained from the `HIGH.RISK` solution fuzzy set. The risk of the project is determined by a single rule and this does not seem reasonable. The cumulative effect of the other rules should influence the determination of risk: if any of these rules also has a significant truth (say .80, .68, or .52) then the evident risk of accepting this particular patient/prosthesis combination is much higher than indicated by the state of `HIGH.RISK`.

Additive inference technique The alternative to the min/max inference technique is the additive inference technique but this is equally unsatisfactory with respect to solving the risk estimation problem. After executing two rules with predicate truths of .62, .48, the solution fuzzy set is now saturated because the maximum truth of a fuzzy set is bounded at 1 (see Figure 6.3). This is a problem associated with the semantic way `INCREASED.RISK` is used in the case of a large number of rules, each of which maps a consequent fuzzy set to the fuzzy solution set. The defuzzification procedure is the same as for the min/max technique. Hence an alternative inferencing technique is

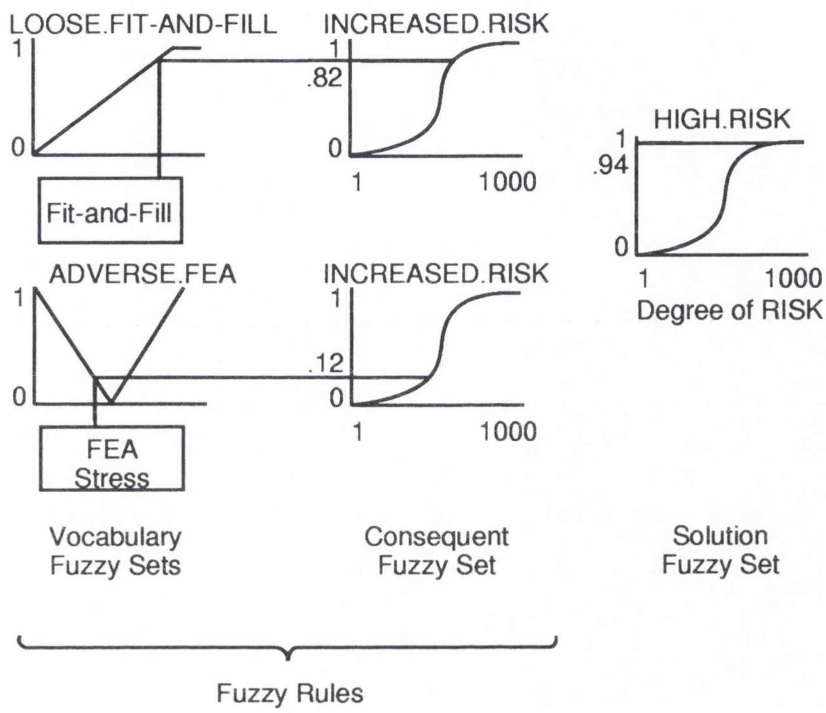


Figure 6.3: The additive technique.

required. Chained monotonic scaling [24] is such a method.

Chained monotonic scaling technique In the chained monotonic scaling technique, two major new ideas are introduced. First, the idea of determining the risk associated with each evaluation factor, based on a vocabulary fuzzy set, and the subsequent mapping of this risk onto a risk measuring fuzzy set (see Figure 6.4). Since the evaluation factor fuzzy sets can have different profiles, this results in a highly non-linear model. Secondly, risks are accumulated in this model by simple addition without the problem of saturation, although it may be preferable to weight them by grade of membership of

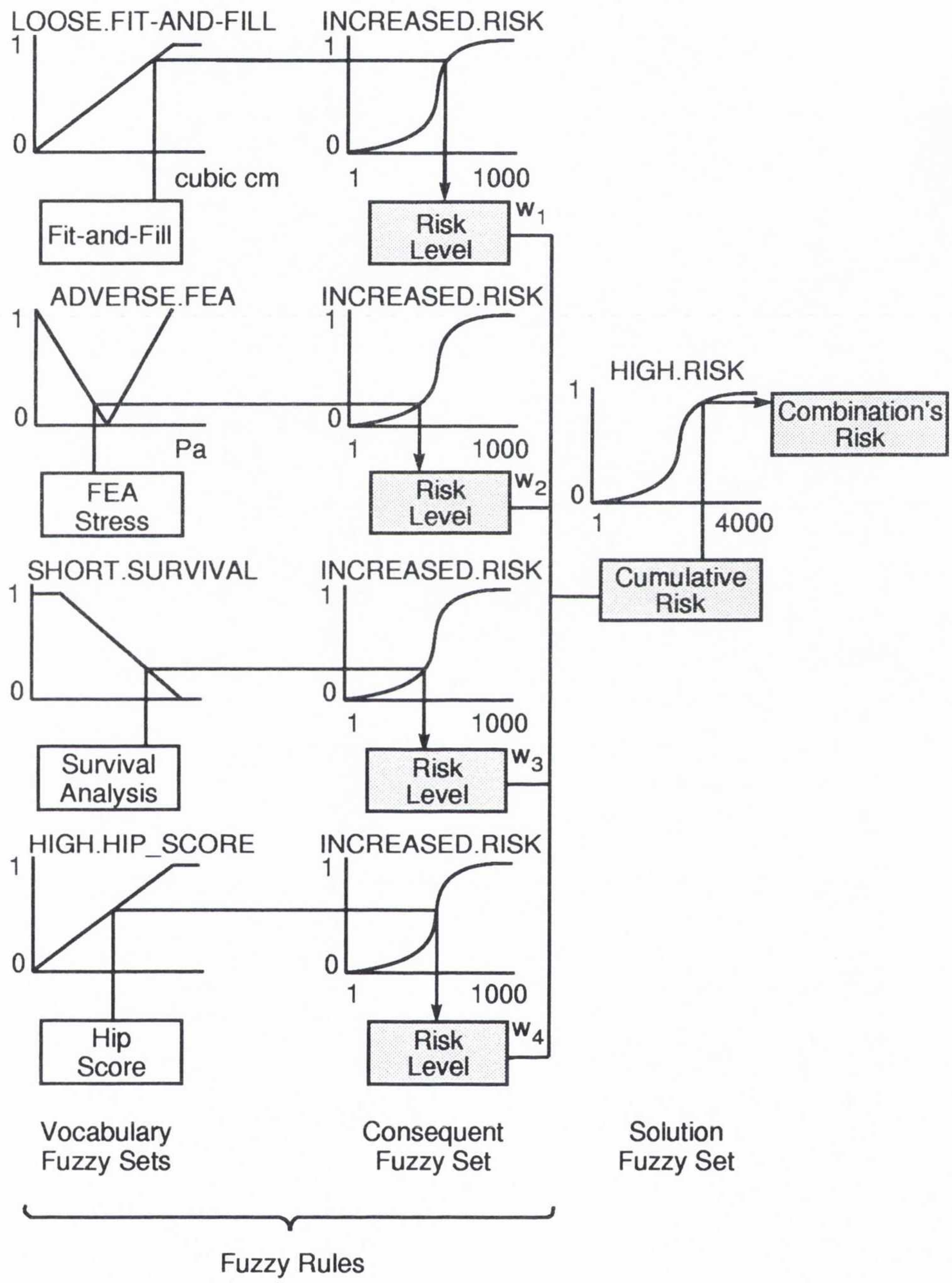


Figure 6.4: The monotonic chaining technique.

the consequent fuzzy set. The scalable monotonic fuzzy model can handle a large number of rules and still maintain the important relationships between the underlying rules and the final risk assessment [6].

6.2 CONSTRUCTING A FUZZY EXPERT SYSTEM FOR SELECTED EVALUATION FACTORS

Developments in *scalable monotonic chaining* [24] and in particular its recent application to project risk analysis by Moody's investors service, raised the possibility of a breakthrough in the capability of fuzzy systems to contribute significantly to the hip prosthesis selection problem. There are sufficient parallels in the assessment of risk associated with competing alternatives, albeit in different spheres of activity, to make the new technique worthy of further investigation. The choice of prosthesis for a particular patient is still treated as a problem of competing alternatives but evaluation is now predicated on the risk of failure. The risk of failure of each patient/prosthesis combination is calculated and the combination which offers the lowest risk of failure is chosen.

6.2.1 Model Design

Figure 6.5 represents the architecture of the model that evaluates the critical factors underlying the risk of failure of a particular patient/prosthesis combination, thus enabling the alternative combinations to be placed in order of risk.

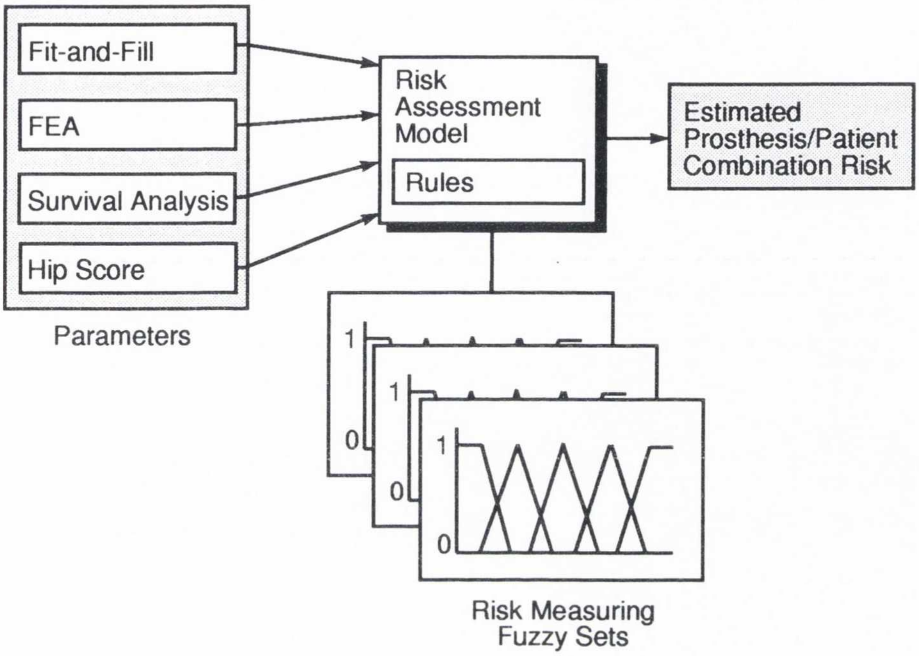


Figure 6.5: The structure of the patient/prosthesis combination assessment model.

The evaluation factors considered by the model developed in this chapter are the same as those represented in the rule-based system in Chapter 5, with the addition of the survival analysis factor. Other evaluation factors touched on in Chapter 4 could equally be included. In addition, a factor like clinical hip score could be deconstructed into its constituent parts and values for these components could be used in the model. For present purposes the core patient/prosthesis combination risk assessment model has four evaluation rules which are enumerated in Program 6.1.

Scaleable monotonic chaining fuzzy systems differ from conventional fuzzy systems which create and then defuzzify a solution fuzzy set. The risk spec-

Program 6.1 Production rules.

```
/* Production Rules (Fuzzy Model's Rules) */
if fit_and_fill_level is loose
      then operation_risk is increased.
if fea_stress_level is adverse
      then operation_risk is increased.
if survival_analysis_level is short
      then operation_risk is  increased.
if hip_score_level is high
      then operation_risk is increased.
```

ified in individual rules is mapped to an intermediate risk measuring fuzzy set (in this case, `INCREASED.RISK` shown in Figure 6.4). The result of this mapping is a scalar value from the domain of the risk measuring fuzzy set, indicating the degree of risk for this particular model factor. The monotonic reasoning results for each rule are added together to produce a final risk value. This value, a scalar, called the **`cummulative_risk`**, is used to find the actual patient/prosthesis combination risk. The **`cummulative_risk`**'s degree of membership in a controlling fuzzy set is found (the `HIGH.RISK` set in our model). This process scales the risk. The degree of membership of the "high-risk" fuzzy set is the overall patient/prosthesis combination risk. In this model, the truth function is multiplied by `1000` to produce a risk factor within the range of the `INCREASED.RISK` domain.

6.2.2 Model Execution

The scalable monotonic chaining technique requires considerable trial and error once the rules (see Program 6.1) have been determined. There are two distinct aspects of the model which are amenable to adjustment in this way.

They are the vocabulary fuzzy sets and the weights applied to the corresponding INCREASED.RISK fuzzy sets. The weights reflect the importance given to a particular rule. The weighting process is straightforward but tedious. The construction of the vocabulary fuzzy sets which implements the

Program 6.2 Fuzzy sets.

```
/* Fuzzy Sets Definitions
Note: The Maximum Domain Limit acts as a
Weighted Measure for its corresponding set. */
fuzzy_set(fit_and_fill_level, loose, as,
      0.0, 1.5, 3.0, 0.0).
fuzzy_set(fea_stress_level, adverse, at,
      0.0, 24.0, 0.0, 0.0).
fuzzy_set(survival_analysis_level, short, dt,
      0.0, 100.0, 0.0, 0.0). /* Greatest Weight */
fuzzy_set(hip_score_level, high, dt,
      0.0, 10.0, 0.0, 0.0 ). /* Lowest Weight */

/* Output Fuzzy Sets */
fuzzy_set(combination_risk, increased, as,
      0.0, 1000.0, 0.0, 0.0).
fuzzy_set(combination_risk, high_risk, at,
      0.0, 5000.0, 0.0, 0.0).
```

fuzzy rules is more involved.

The shapes of the different vocabulary fuzzy sets can be seen in Figure 6.4. The code in Program 6.2 shows how these fuzzy sets are defined.[2] These curves are indicative of a general approach rather than a working prototype.

Once the shape of the vocabulary fuzzy set is decided, its other characteristics such as the curve slope are subject to trial and error adjustment.

[2]This approach was adapted from the Prolog implementation of scalable monotonic chaining proposed by Pacheco (1997) [122]. The variables dt, at, tp and as indicate linear decreasing, linear increasing, trapezoidal or triangular, and increasing s-curve fuzzy set types respectively (see Appendix D).

Program 6.3 Inputs and outputs for low and high risk assessment models.

```
/* Input for a low risk assessment model
fit_and_fill_level: 0.8000000
fea_stress_level: 12.000000
survival_analysis_level: 40.000000
hip_score_level: 3.000000

Output for a low risk assessment model
combination_risk: amount = 142.222244 membership = 0.142222
combination_risk: amount = 500.000000 membership = 0.500000
combination_risk: amount = 600.000000 membership = 0.600000
combination_risk: amount = 700.000000 membership = 0.700000
Cummulative Risk is 1942.222168
combination_risk = 388.444427

Input for a high risk assessment model
fit_and_fill_level: 25.000000
fea_stress_level: 19.000000
survival_analysis_level: 10.000000
hip_score_level: 1.000000

Output for a high risk assessment model
combination_risk: amount = 944.444397 membership = 0.944444
combination_risk: amount = 791.666687 membership = 0.791667
combination_risk: amount = 900.000000 membership = 0.900000
combination_risk: amount = 900.000000 membership = 0.900000
Cummulative Risk is 3536.111084
combination_risk = 707.222229 */
```

The prototype program was tested with data which were obtained in conservations with surgeons and finite element analysis experts. These data values, though hypothetical, are therefore agreed and acceptable estimates. The inputs, together with the corresponding outputs, are shown in Program 6.3. Two samples are given, one refers to a low risk patient/prosthesis combination, the other to a high risk patient/prosthesis combination. The results validate the working of the model but they also confirm the necessity to expand the number of rules and to refine the associated fuzzy sets. Such a process would require a far greater acceptance of the use of fuzzy rules, before the committal of resources to a lengthy dialogue with surgeons could be justified.

6.3 SUMMARY AND CONCLUSIONS

- In order to address the reality that the expert knowledge of surgeons is difficult to represent by precise rules (as required by rule-based expert systems) this chapter explored the possibility of adapting the expert system of the previous chapter in such a way as to legislate for the linguistic and imprecise manner in which surgeons like to describe their professional assessments.

- Fuzzy logic is a formalism that allows this imprecision and uncertainty to be captured. The fuzzy approach is an unorthodox one and is not amenable to classical mathematical analysis and justification.

- A working prototype which exploits the versatility of Prolog has been developed and this is offered as a template for a more in-depth exploration of the fuzzy technique, as applied to patient/prosthesis combination evaluation. At present such an exploration is curtailed by the lack of sample data with which to fine tune the program. Fine tuning is achieved by adapting the shape of the fuzzy sets until the output of the program reflects expert judgements.

- Only four rules were used in the prototype and this number would have to be greatly expanded in a production program. Each of the rules suggested can be decomposed into a number of more specific rules but this places a premium on having data to test and verify the system.
- The fuzzy system approach, to the extent that it is developed here, can only claim to demonstrate that conventional expert systems can be adapted to reflect the way surgeons think and express professional judgement.
- The rule-based expert system developed in Chapter 5 remains the most highly developed solution and the one most likely to be accepted for implementation in the short term and is also the foundation from which fuzzy systems will ultimately evolve.

Chapter 7

THR Feed-Forward Neural Network

Where is the wisdom,
That we have lost in knowledge;
Where is the knowledge,
We have lost in information.
T.S. Elliot

Chapters 4, 5 and 6 have addressed the problem faced by the surgeon: "How do you evaluate a particular patient/prosthesis combination at the clinical stage?" The techniques used in previous chapters come within the generic approach of pattern recognition or data mining. A pattern can be thought of as an instance of a model i.e. $f(x) = 3x_2 + x$ is a pattern whereas $f(x) = \alpha x_2 + \beta x$ is a model. Data mining involves fitting models to, or determining patterns in observed data [41]. So far, the approaches have concentrated on eliciting from surgeons rules that may help in identifying patterns in the data, with a view to classifying patient/prosthesis combinations as acceptable or not. The prototypes of these approaches proved viable and hold much potential once the rules have been stabilised.

There is however another approach which bypasses the need for surgeons to explicitly enumerate rules at the clinical stage. This is the neural network approach were the rules are implicitly learnt and contained in the network. Neural networks are analogous to curve-fitting in so far as they recognise patterns in historical data. Neural networks can be applied to the prosthesis, patient and other relevant data available to the surgeon immediately prior to the operation; and can also be factored into the rule-based expert system developed in Chapter 5. But the real power and potential of neural networks is realised through having access to national and international hip prosthesis databases. These data can be used to train a neural network to recognise patterns and classify patient/prosthesis combinations.

For systems that have no learning or have supervised learning,[1] neural networks and curve-fitting are pretty much the same. Both distil functions from data, thereby allowing interpolation, extrapolation and generalisation. The main difference between the two is in the processing of input data. The learning techniques of neural networks mean that the function can be automatically updated, while most curve-fitting techniques are "batch."[2] In curve-fitting a relatively complex system of linear equations is solved and when an additional data point is added the old curve must be replaced.

[1] Unsupervised learning techniques for neural networks (reinforcement learning) do not have any clear analogues in traditional curve-fitting.

[2] A batch system processes data in discrete groups of previously-scheduled operations rather than interactively or in real time.

Neural network techniques are best suited to situations where the number of data points is large and continual learning from the addition of new points is involved (as is the case in THR analysis). With neural networks, however, the problem of proving that the solution found is optimal remains unresolved.

In this chapter it is proposed to first outline the background and potential of neural networks. The underlying theory is outlined in Appendix E while some notable examples are described in Appendix F. The fact that neural network design is critically dependent upon the availability of reliable and consistent data is highlighted by describing the process involved in neural network training and testing. This leads to a discussion on the inadequacies of the current situation with regard to national and international hip registers.

A detailed description and analysis of the recorded information contained in the Swedish Hip Register, which provides the data for this research, follows. This researcher did not have access to the original Swedish Hip Register data but to data which was extracted from graphs published in a series of research papers on the Swedish Hip Register [1, 82, 83]. These graphs show the effect of selected variables on survivability. Data on patient and implant factors are available separately, but not in combination. More limited data are available on other variables, such as surgical procedures and type of hospital, which influence prosthesis survivability. This inhibited a fuller exploration of the value of neural networks (and resulted in the construction of two more limited neural networks where a single comprehensive one would have been more appropriate), but it did not diminish the validity of the substantive point of the thesis that neural networks hold great potential in the field.

Then follows the most substantive part of the chapter which describes the design, training and testing of two separate neural networks one with patient-related data sets and the other with implant-related data sets. In the conclusion, the positive outcomes, and the limitations, of the neural network approach are highlighted.

7.1 BACKGROUND

Neural networks are a new phenomenon as evidenced by the fact that many of the techniques used in this thesis were developed and first propagated within the past decade. The enormous potential of neural networks is expected to increase as neural network techniques are further developed. Neural network techniques are permeating every sphere of computer use, and have made possible applications in such intractable areas as voice and handwriting recognition.

Artificial Neural Networks (ANNs) are simplified models of the human central nervous system. They are networks of highly interconnected neural computing elements that have the ability to respond to input stimuli and to learn to adapt to the environment. It is believed by many researchers in the field that neural network models offer the most promising unified approach to building truly intelligent computer systems; and, that the use of distributed, parallel computations, as performed in ANNs, is the best way to overcome the combinatorial explosion associated with symbolic serial computations when using von Neumann computer architectures [101].

The human neural network system provides a strong stimulus for emulating its behaviour in ANNs. Biological networks are able to process millions of input stimuli in milliseconds, even though the processes are electrochemical in nature, and therefore propagate at relatively slow millisecond rates. This rate is several orders of magnitude slower than the high-speed picosecond operations performed in conventional serial digital computers. In spite of this wide divergence in signal propagation and unit processing speed, conventional state-of-the-art computer systems, such as vision systems, fall far short of the performance exhibited by biological systems in their processing ability. ANNs have been shown to be effective as computational processors for various tasks including pattern recognition (e.g. speech and visual image recognition), associative recall, classification, data compression, modelling and forecasting, combinatorial problem solving, adaptive control, multi-sensor data fusion and noise filtering. They exhibit a number of desirable properties not found in conventional symbolic computation systems including robust performance when dealing with noisy or incomplete input patterns, a high degree of fault tolerance, high parallel computation rates, the ability to generalise and adaptive learning.

7.2 ACQUIRING DATA SETS

The successful operation of a neural network is dependent on the availability of a comprehensive set of training and testing data. This section describes and classifies the data, pertinent to the field of total hip replacement, used for training and testing data sets.

7.2.1 Purpose of Data Sets

Data sets are required to train and test or validate a neural network.

7.2.1.1 Training

A neural network is usually trained in one of two ways. The most common (and the method used in this thesis) is *supervised* training. Many samples are collected to serve as exemplars. Each sample in this *training set* completely specifies all inputs, as well as the outputs that are desired when those inputs are presented. Then we choose a subset of the training set and present the samples in that subset to the network one at a time. For each sample, we compare the outputs obtained by the network with the outputs we would like it to obtain. After the entire subset of training samples has been processed, we update the weights that connect the neurons in the network. This updating is done in such a way as to reduce the error in the network's results. A single pass through the subset of training samples, along with an updating of the network's weights is called an *epoch.* The number of samples in the subset is called the *epoch size.* When the epoch size is less than the entire training set, it is important that the subset be selected randomly each time, or troublesome oscillations may occur.

The other principal training method is *unsupervised* training. As in supervised training, we have a collection of sample inputs. But we do not provide the network with outputs for those samples. We typically assume that each input arises from one of several classes, and the network's output is an identification of the class to which its input belongs. The process of training the network consists of letting it discover salient features of the training set, and using these features to group the inputs into classes that it (the network) finds distinct.

7.2.1.2 Testing

After training the competence of the network must be tested before it is put into service. This process is called *validation.* The usual procedure is to separate the known cases into two disjoint sets. One is the training set, which is used to train the network. The other is the testing set, which is used to test the trained network (the training of the hip replacement survival neural networks is described in Section 7.3.2 and the testing of the networks is described in Section 7.3.3).

In many respects, proper testing is more important than proper training. A very small error on the training set may mistakenly be taken as an indication that all is well with the network. An examination of Figure 7.1 shows why a low training-set error can be misleading. If the model has too many free parameters relative to the number of cases in the training set, it can *over-fit* the data. Rather than learning the basic structure of the data, enabling it to generalise well, it learns irrelevant details of the individual cases. Naturally, we can expect the error on the testing set to exceed slightly that on the training set. But if the difference is large, we must suspect that one of the two sets is not representative of the same population, or the model has been over-fitted. In either case, disparity in the errors is a warning sign that must not be ignored.

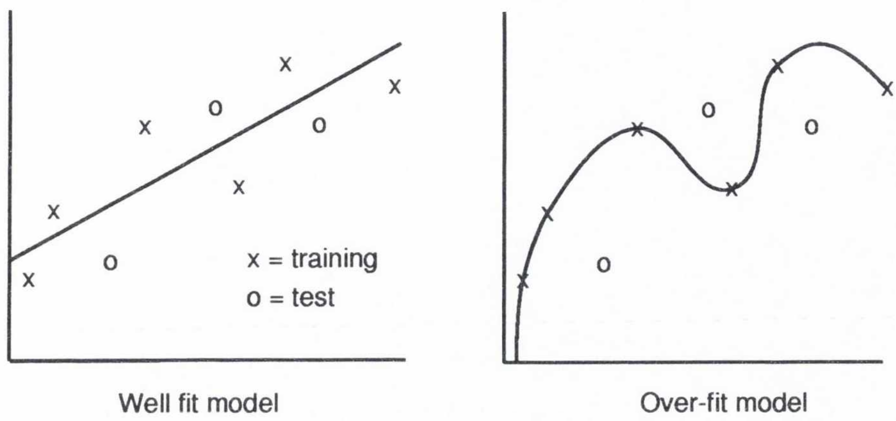


Figure 7.1: Low training error need not imply good performance.

7.2.2 Search for Consistent Data Sets

An enormous amount of research, across a wide range of countries, has been undertaken into total hip replacement, but there is little consistency in the data on which that research is based because there is no *international* body with responsibility for keeping records on hip replacements and for monitoring how measurements are taken and how data are collected.

This situation is exasperated by the use of many alternative hip scoring systems which produce a variety of objective scores characterising the outcome of hip replacements [96]. The difficulty of taking reliable and repeatable measurements from x-rays, even within the same hospital, further complicates the task of pooling the data. The harmonisation of such scoring systems and the standardisation of measurements requires an international hip register body and agreed standards guiding the introduction of new implants into clinical use.

7.2.2.1 International Hip Register

The lack of an international hip register is put down to the extent of the resources required to set up such a body. However, in light of the cost of the research that is being pursued and which is rendered less effective because there are no overall terms of reference, this position is difficult to sustain. Furthermore, the economic and social cost of sub-optimal prostheses is immense in the medium term. For example, the general failure of the Christiansen prosthesis (see Figure 7.11) in Sweden is calculated to have cost US$ 20 million (€17.14 million) in total [1]. The cost of one revision due to aseptic loosening is put at US$ 12,500 (€10,715). This takes no account of the individual patient's suffering. The number of people presenting for difficult and costly revision surgery is increasing rapidly.

7.2.2.2 Ordered Introduction of New Implant Technology

Closely allied to the use of an international hip register is the importance of developing and agreeing models and standards for pre-clinical and clinical testing *prior* to the introduction of new implants into clinical praxis. Only the United States and very few other countries have legal and regulatory processes guiding the introduction of new implants into clinical use. New implants and fixation techniques are, as a consequence, usually introduced into clinical practice without scientific validation, thereby exposing many patients to health hazards. The fight for "market share" amongst manufacturers has resulted in various implant design modifications (multi-modularity, new plastics, different surface finishes) each claiming superior clinical performance, but lacking scientific validation. Designers have been frustrated by the existence of incompatible design goals which can create enormous complications [62].

Many flawed prostheses have been introduced which, with hindsight, could have been discovered prior to introduction had all such innovations been examined in an ordered fashion [81]. For example, pre-clinical tribologic testing would have pointed out the risks associated with the heat treatment of polyethylene; and fatigue testing of experimental designs could have detected the proclivity to aseptic loosening when using low viscosity cements.

Neural networks, applied to hip replacement survival prediction, have enormous potential to further illuminate the hidden complexities of total hip replacement as part of an ordered introduction of new implants, as well as predicting the success of particular patient/prosthesis combinations. However, for the full potential of the neural network model to be realised, a comprehensive training data set from an authoritative hip register is an essential requirement.

7.2.3 Data Sets for this Research

The national hip register for hip replacement data, which has existed in Sweden since 1979, provided the data set for this thesis. Although the data contained in the hip register are not as comprehensive as would be desired, they are sufficient to train and test the potential of neural networks and thereby to advance total hip replacement research. The information contained in the Swedish national hip and knee replacement registers has led to documented changes of THR medical practice and to the establishment of no less than 19 registers in Sweden covering different medical areas. The data used in this thesis were extracted from the findings of the Swedish Hip Register [1, 82, 83].

In the Swedish Hip Register, the definition and end-point for failure of a hip implant is *revision*, which is defined as the exchange of one or both prosthetic components or the permanent removal of the prosthesis. The Swedish study started on January 1, 1979. All orthopaedic departments in Sweden were included. Complete copies of the hospital records of all revised THRs were assembled and these data were computerised. Each patient was allocated a unique identification number thus bringing together individual data, such as age, sex, diagnosis and type of prosthesis used.

More than 130,000 primary THR operations were performed in Sweden in the period 1978–1994. In the period 1979–1994 a total of 9,965 revisions were performed, about one thousand of these were on patients who had received

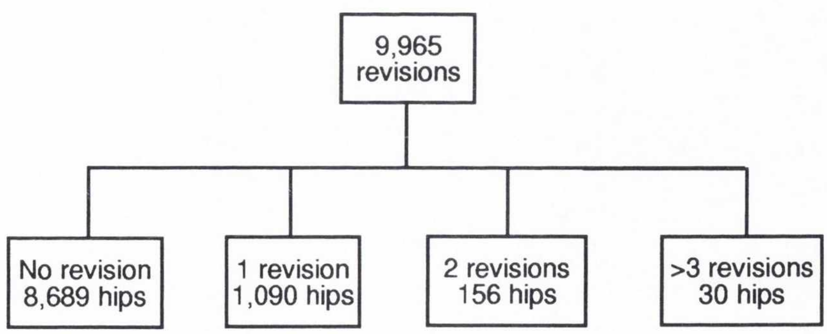


Figure 7.2: THR revision 1979–1994.

at least one previous revision. These latter cases were excluded from the data set used for the current study and hence the analysis is based on the balance of 8,689 hip revisions. The primary THR operation for these hips was performed in the period 1978–1994.

All surgical units performing THR in Sweden have provided a detailed report of their surgical techniques year-by-year since 1978. The techniques used include a femoral plug, cleaning of the bone bed and pressurisation by means of a proximal plug. Furthermore, type of cement, cèment mixing and application techniques, including vacuum mixing, were described. This information makes it possible to make an evaluation of the relevance of *surgical technique*.

Reason	Number	Percentage
Aseptic loosening	6,368	79.3
Infection	633	7.3
Technical error	339	3.9
Dislocation	294	3.4
Bony fracture	91	1
Pain	33	0.4
Miscellaneous	777	8.9

Table 7.1: Reasons for revision 1979–1994.

Prophylactic measures taken against deep infection by means of parenteral and local antibiotics were also recorded. The type of operative environment including laminar air flow and body exhaust gowns were documented year-by-year for every department and in this way the relative effectiveness of different prophylactic measures against *deep infection* was monitored.

Aseptic loosening has emerged as the main problem and the immediate reason for revision (Table 7.1). Deep infection and technical problems are becoming less significant as surgical techniques improve. Sections 7.2.3.1 and 7.2.3.2 describe the origin of the two groups of graphs from which two separate data sets were obtained. Each of the data sets was subsequently used to separately train and test two distinct neural networks.

7.2.3.1 Patient-Related Parameters

Important patient-related factors that influence total hip replacement survivability are age, sex and primary diagnosis. The graphs in Figures 7.3, 7.4, 7.5 and 7.6 are representative of the results from the Swedish Hip Register showing patient-related factors which influence survivability. In the register a total of 24 graphs, divided into three groups of eight, represent the three different primary diagnoses (osteoarthrosis, rheumatoid arthritis and hip fracture). Each group of eight is comprised of four graphs representing the different age groups (< 55, 55–64, 65–74, > 75) for male and female patients respectively.

Training and testing set data were constructed from these graphs and the data used to train and test a neural network to predict patient survivability (see Section 7.3.2.1). The conclusions to be drawn from these graphs are not always obvious. For example, the lower survival rate in men with the primary diagnosis osteoarthrosis of the age group 55–64 (see Figure 7.4) compared with the age group less than 55 (see Figure 7.3) is an unexpected outcome.[3] The distinctive capability of a neural network is a capacity to learn from historical data and to adjust correctly for a range of different influences.

[3]It is intuitively expected that prostheses in younger patients, who lead more active lives, would be more likely to fail.

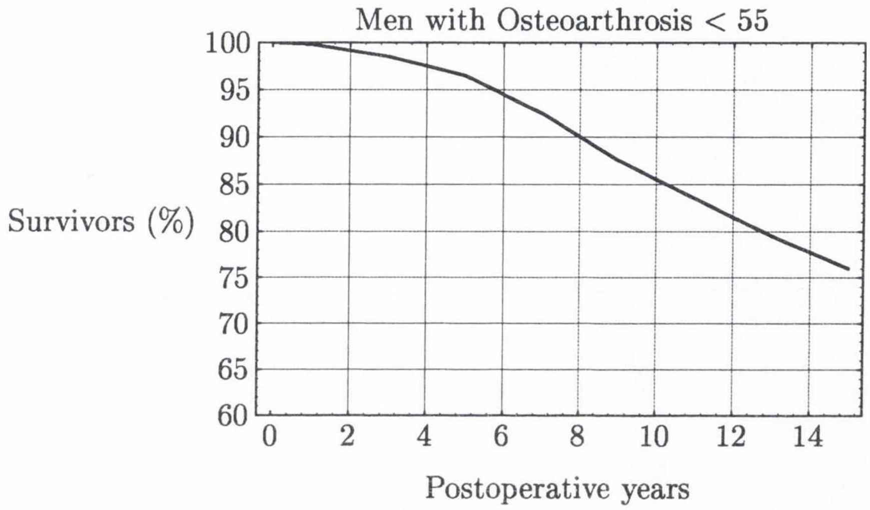


Figure 7.3: Survival rate for men with the primary diagnosis osteoarthrosis in the age group less than fifty five.

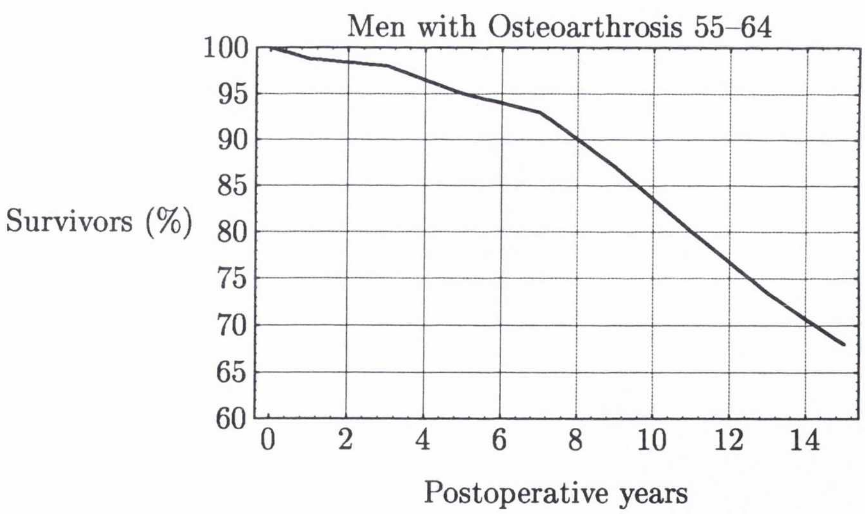


Figure 7.4: Survival rate for men with the primary diagnosis osteoarthrosis in the age group fifty five to sixty four.

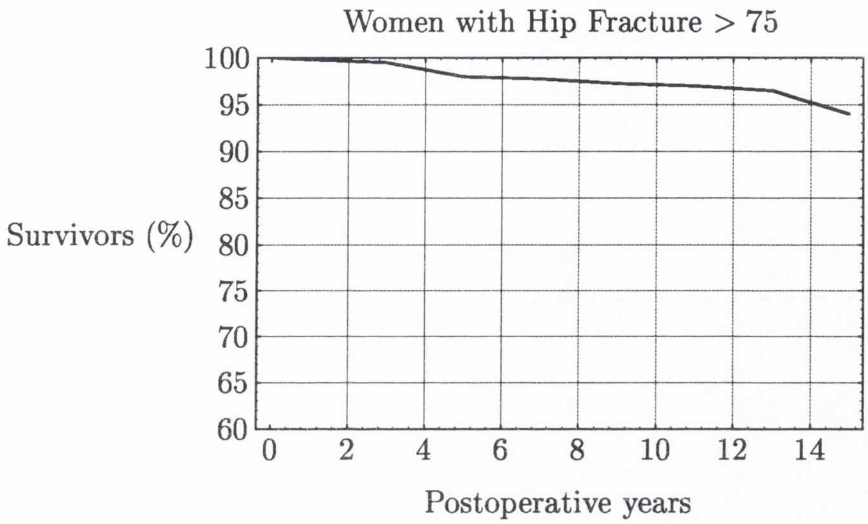


Figure 7.5: Survival rate for women with the primary diagnosis hip fracture in the age group greater than seventy five.

An additional attribute of neural networks is their ability to predict outcomes. The graph in Figure 7.6 is truncated due to insufficient data. A neural network, based on adequate training and testing data, was used to predict values for the missing data.

The available research papers on the Swedish Hip Register limit their analysis of the patient characteristics that affect the outcome of THR to age, gender and disease type. Other patient characteristics that can affect outcome are weight, socio-economic status, education, ethnicity, preoperative function status, activity level and co-morbidity[4] [138]. Thus the potential of a

[4]Co-morbid conditions may have a significant impact on patient outcome after THR. For example, a patient who is unable to walk up a flight of stairs before THR due to hip disease may not show improvement after surgery due to secondary restrictions, such as ischemic heart disease.

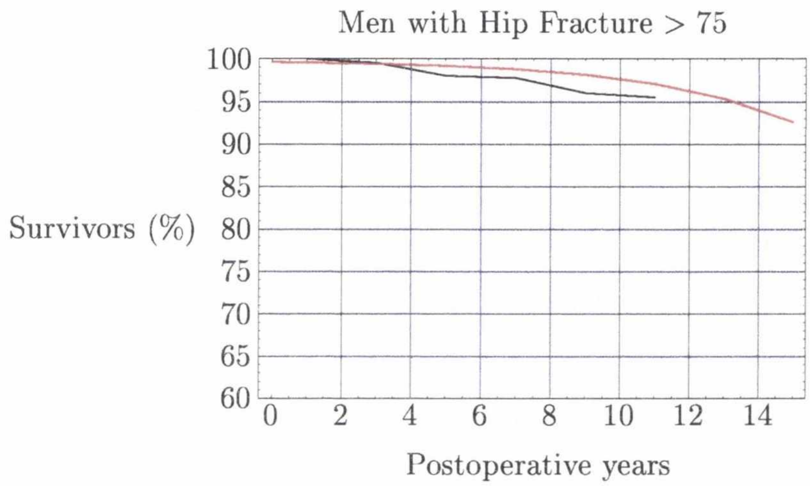


Figure 7.6: Survival rate for men with the primary diagnosis hip fracture in the age group greater than seventy five: the neural network predicted survival rate is in red.

patient-related neural network to predict survivability is capable of significant expansion.

7.2.3.2 Implant-Related Parameters

We now turn to non-patient-related factors that influence total hip replacement survivability. Critical non-patient-related parameters include type of implant, type of hospital and choice of surgeon. The type of implant is the only parameter for which graphs, based on data in the Swedish Hip Register, were published, and the second neural network is therefore implant-related.

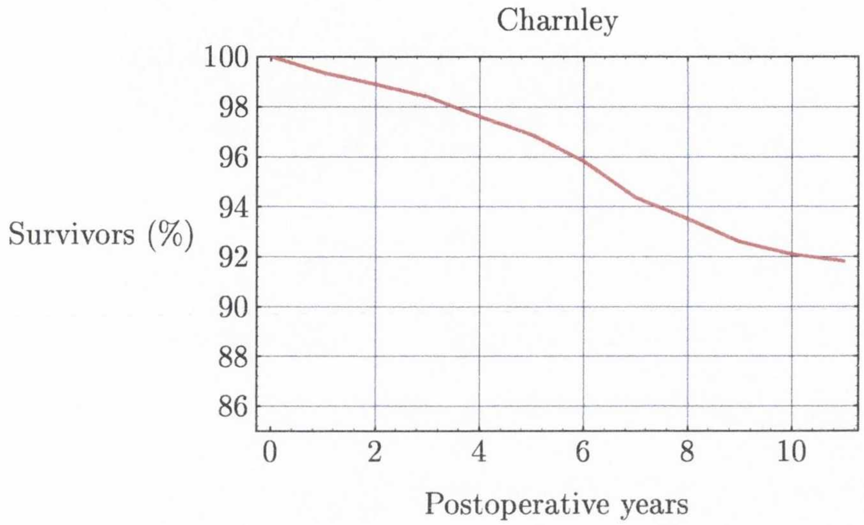


Figure 7.7: Survival rate for aseptic loosening using the Charnley prosthesis.

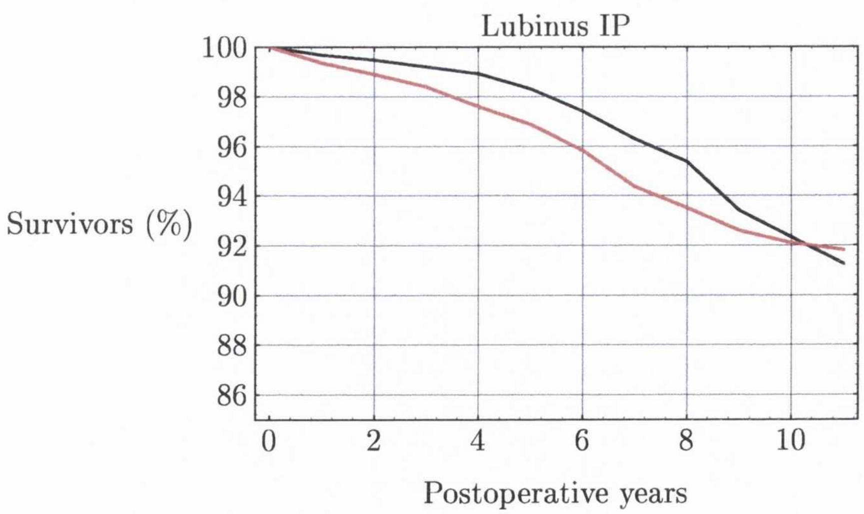


Figure 7.8: Survival rate for aseptic loosening using the Lubinus prosthesis compared with Charnley in red.

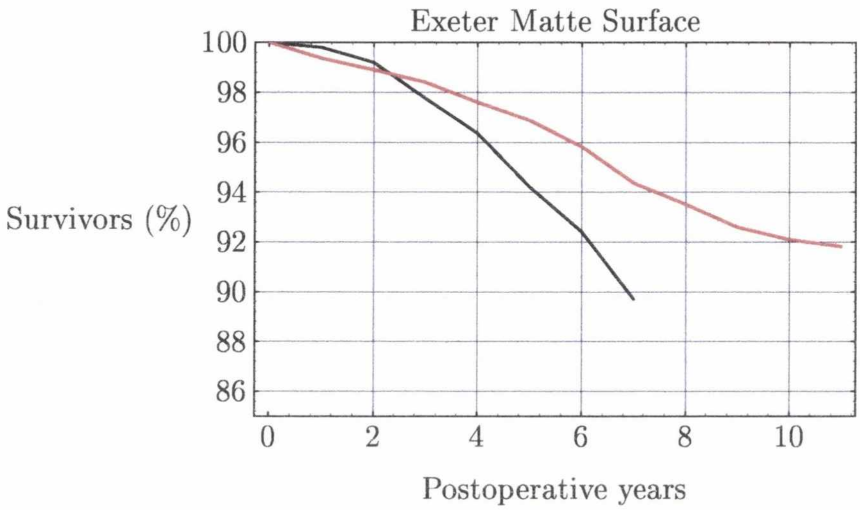


Figure 7.9: Survival rate for aseptic loosening using the Exeter matte surface prosthesis compared with Charnley in red.

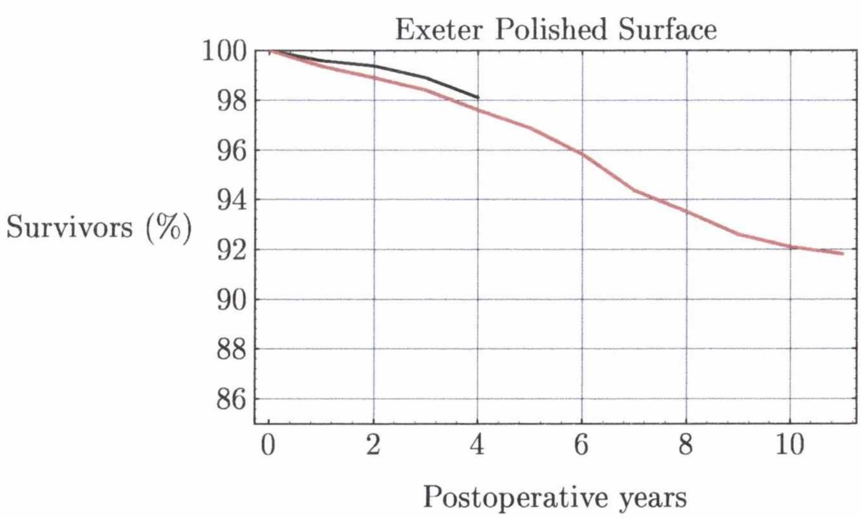


Figure 7.10: Survival rate for aseptic loosening using the Exeter polished surface prosthesis compared with Charnley in red.

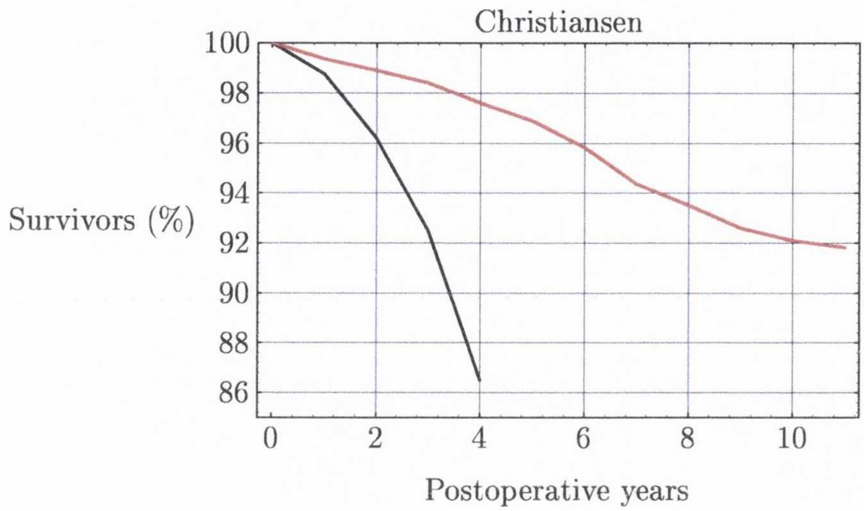


Figure 7.11: Survival rate for aseptic loosening using the Christiansen prosthesis compared with Charnley in red.

Type of Implant The Swedish Hip Register highlights significant differences between various implants with respect to revision for aseptic loosening due to osteoarthrosis. The graphs in Figures 7.7, 7.8, 7.9, 7.10 and 7.11 are representative of the results from the Swedish Hip Register showing the effect of implant type on survivability for aseptic loosening in osteoarthrosis. The Charnley survival curve is plotted in red in all of the graphs to enable comparison. There were no differences between prosthetic designs when failure leading to revision for infection was analysed.

The Charnley, Lubinus IP and CAD implants have been used throughout the full observation period. The performance of these implants has been good and no difference in survival rate is noted between any of them at the 10 year follow-up stage. The Exeter matte surface and Müller curved prostheses are in an intermediate group with a higher rate of revision. The worst performance is observed with the Christiansen prosthesis. More recent implants, however, developed in the middle of the 1980s (Spectron, Lubinus SP, and Scan Hip), have a very low revision rate. Improved implant design, cement pressurisation, femoral plugging and vacuum mixing of cement were increasingly used in combination with these newer implants. A significant improvement has been observed, leading to a reduced rate of revision for aseptic loosening for the Spectron as opposed to the Charnley design. One explanation may be that the Charnley procedure still uses the second generation of cementing technique whereas the newer implants use the most recent techniques.

7.2.3.3 Other Non-patient-Related Parameters

In addition to patient-related and implant-related factors there are other variables which influence prosthesis survivability. Two such variables are type of hospital and skill/experience of the surgeon. The Swedish Hip Register contains some hospital and surgeon-related information, but this information was not graphed by patient group or prosthesis type and thus it was not possible to include it in the construction of the neural networks. A short description of the other parameters is included only for completeness.

Type of Hospital Records of revision rates for aseptic loosening and deep infection exist for different hospitals. Statistical differences were found between three different types of hospital: university hospitals (tertiary hospitals), regional hospitals (secondary hospitals), and community hospitals (primary hospitals).

University hospitals reported more infections than the other two groups. The reason could be that these hospitals attracted special problem cases that demanded long-term and extensive procedures.

When comparing aseptic loosening among all the prostheses the regional hospitals outperformed the others in all cases except for the Christiansen prosthesis and the surface replacements.

Surgeon Related Factors Information on the number and type of primary prostheses which 37 individual surgeons had implanted each year was available. These 37 orthopaedic surgeons worked mainly at one of three hospitals: Malmö General Hospital, Sahlgren Hospital or East Hospital in Gothenburg. Information on some surgeons working at community hospitals was also available. All the surgeons were categorised as experienced surgeons and performed 30 to 70 hip replacements annually.[5]

There were some statistical differences between the surgeons for aseptic loosening. Two of 33 surgeons had fewer complications for aseptic loosening than the others and one surgeon had more re-operations than any of the others. Most surgeons performed to standard.

The basis of the statistical differences was not related to the number of primary prosthetic operations performed per year: all of these surgeons were well trained and had long experience. This was reinforced by the fact that the number of dislocations (normally associated with surgeon error) was, with one exception, very small. This latter surgeon had a higher incidence

[5]Due to their high failure rate, double-cups (surface replacements) and Christiansen prostheses were excluded, because these prostheses could distort the analysis.

of dislocation after primary hip replacement even though he performed more than 200 primary operations a year.

Finally, a recent study has found that the survivability of hip implants for Sweden does not match the performance of centres of excellence using modern cementing techniques [94]. This is explained by the scope of the Swedish Hip Register that includes a great number of individual surgeons with different surgical skills and backup. However, comprehensive national statistics, such as those assembled in Sweden, more accurately reflect the average quality of medical care to the public. Average surgical results are predictably less impressive than outcomes from specialised orthopaedic centres.

7.3 MODEL DESIGN, TRAINING AND TESTING

Two distinct neural networks have been designed, trained and tested with the patient-related and the implant-related data sets. This is the subject of this section.

The process of designing, training and testing a hip prosthesis neural network can now be initiated. In practice two separate networks will be built because the manner in which the data became available to the researcher did not facilitate the linking of the patient and the prosthesis input variables.

The neural networks were initially constructed in *Mathematica*™. This made it possible to get a visual understanding of the problem, allowed experimentation with variants of neural network architecture, and facilitated network performance analysis. The results were then reprogrammed in C++ in order to speed up calculations.

7.3.1 Design of Feed-Forward Neural Networks

This section addresses the difficult problem faced in choosing a structure of feed-forward network.

7.3.1.1 Number of Hidden Layers

The first decision faced (in designing the networks) related to the number of hidden layers. There is no theoretical reason *ever* to use more than two hidden layers (see Section E.4). Furthermore, for the vast majority of practical problems, there is no reason to use more than one hidden layer. Those problems that require two hidden layers are rarely encountered in real-life situations. But the question arises as to whether the *theoretical* requirement reconciles with the *practical* requirement i.e. are there problems where learning is facilitated by having more than the minimum theoretically required number of hidden layers?

Experience has shown that the use of more than one hidden layer is almost never beneficial [84]. The problem is that training tends to slow dramatically when additional hidden layers are used. This is due to two effects:

1. Additional layers, through which errors must be back-propagated make the gradient more unstable. The success of any gradient-directed optimisation algorithm is dependent on the degree to which the gradient remains unchanged as the parameters (weights in a neural network)

vary.

2. The number of false minima is likely to increase dramatically. This means that there is a higher probability that after many time-consuming iterations, the network's calculations will be stuck in a local minimum and have to escape or start over.

7.3.1.2 Number of Hidden Neurons

Choosing an appropriate number of hidden neurons was the next critical decision. Using too few starves the network of the resources it needs to solve the problem. Using too many increases the training time, perhaps to the extent that it becomes impossible to train the network adequately in a reasonable period of time. Also, an excessive number of hidden neurons may result in over-fitting. The network will have so much information processing capability that it will learn insignificant aspects of the training set, aspects that are irrelevant to the general population. The performance of the network will be excellent when it is evaluated with the training set but, when the network is called upon to work with the general population, it will perform poorly. This is because it will consider trivial features, unique to training set members, as well as important general features, and become confused (see Figure 7.1). Thus, it is imperative to use the absolute minimum number of hidden neurons that will make the network perform adequately.

The decision on choosing the number of hidden neurons was facilitated by the existence of the *geometric pyramid* rule. This states that for many practical networks, the number of neurons follows a pyramid shape, with the number decreasing from the input towards the output. This is illustrated in Figure 7.12. The numbers of neurons in each layer follow a geometric

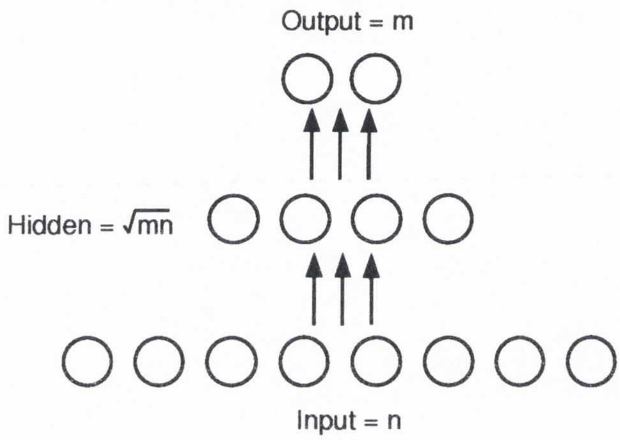


Figure 7.12: Typical three-layer network.

progression. Thus, if we have a three-layer network with n input neurons and m output neurons, the hidden layer would have $\sqrt{mn}$ neurons. A similar rule applies to four-layer networks. In this case, computation of the number of hidden-layer neurons is slightly more complex

$$r = \sqrt[3]{\frac{n}{m}} \tag{7.1}$$
$$N_{hidden1} = mr^2$$
$$N_{hidden2} = mr.$$

The above formulas should be taken as rough approximations. If there are very few inputs and outputs, and the problem is complex, the formulas may underestimate the number required. For example, approximating a complicated function of one variable involves just one input and one output neuron, but may require a dozen or more hidden neurons. On the other hand, in the case of a simple problem with many inputs and outputs, fewer neurons will often suffice.

Finding the optimal number of hidden neurons was time-consuming. The procedure was to start with a number of neurons which was definitely too small. Choose an appropriate criterion for evaluating the performance of the network (see Section 7.3.3.1). Then slightly increase the number of hidden neurons, and train and test again. This was to be repeated until the error was acceptably small, or no significant improvement was noted, whichever comes first. When testing sets are easily obtained, neurons may be added past the point of acceptable results, as the testing procedure can be relied on to warn of over-fitting. Otherwise, it is advisable to use the minimum number of hidden neurons necessary to achieve acceptable performance. Increasing the number beyond that minimum may cause deterioration in generalisation ability. The training flowchart used is shown in Figure 7.13.

It is tempting to preserve the learned weights for the next test. In other words, suppose that a network having five hidden neurons has been trained. When we add a sixth, keep the same weights for the first five. Initialise the weights for the new, sixth neuron to small random numbers, and continue training from there. The rationalisation is that we already have learned a lot. This is a mistake unless you are willing also to try totally random initialisation. Although there are some problems for which this will work, there are many situations in which this will rapidly lead to a false minimum. The optimal weights for n hidden neurons rarely are even close to being a

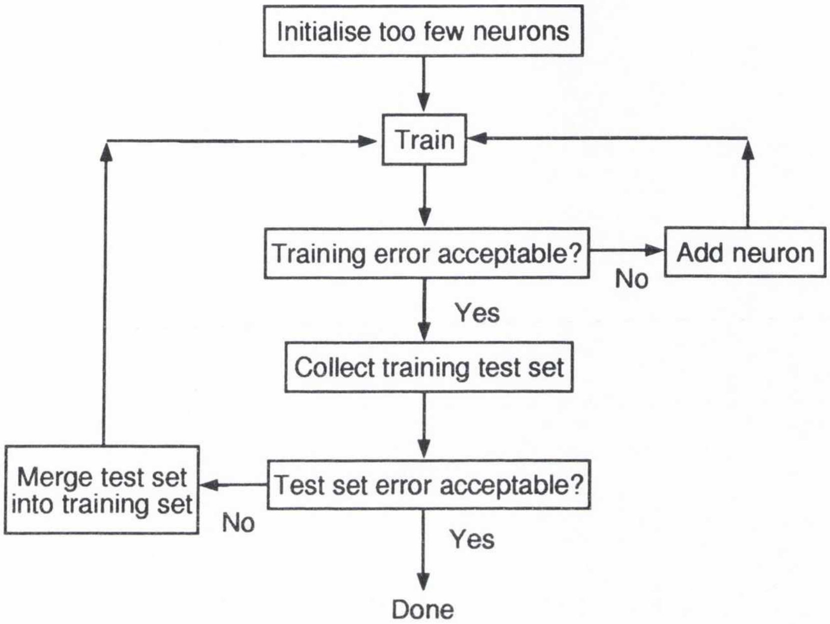


Figure 7.13: Training flowchart when known cases are easily obtained.

subset of those for $n+1$ hidden neurons. If neural networks were linear, this would be an excellent procedure, with each addition taking a bite out of the error subspace left behind by its predecessors. But the profound non-linearity of these networks prevents this from happening. When a new neuron is added, the augmented reasoning capability usually means that the network should have an entirely different approach to the solution.

A lot of effort is currently being devoted to the design of *self-pruning* networks. A wide variety of methods have been proposed, but most have in common the simultaneous minimisation of output error and minimisation of the number of hidden neurons. Many of these methods define an auxiliary criterion based on the number of hidden neurons, or on the size of the weights connecting hidden neurons to other layers, or on hidden-neuron activation distributions. The function that is optimised is a composite of the output error with the hidden-neuron economisation criterion. These methods can be easily abused [84]. They do not directly address the issue of over-fitting. Naturally, this issue is indirectly addressed, in that using fewer hidden neurons decreases the likelihood of over-fitting. However, that fact may give a false sense of security. Furthermore, the choice of how to weigh economy versus training accuracy is highly arbitrary. Self-pruning algorithms are in still their infancy. Developments are coming; in particular Fakhr proposes what seems to be a promising approach [40].

One last point should be made. It is often surprising how few hidden neurons are required. The tendency is invariably to overestimate the requirement. It is not at all unusual to have a problem with hundreds of input neurons and several output neurons, which only requires five or so hidden neurons. The best approach is start low and work up as needed. It is unfortunate that using fewer hidden neurons often increases the likelihood of the learning algorithm becoming trapped in a local minimum. Additional weights can create new channels through which gradient descent is able to pursue a global minimum. In practice, though, the trade-off is rarely worthwhile. Stick with the minimum number necessary to solve the problem, and emphasise thorough training.

7.3.1.3 Preparing Input Data

Measured variables are categorised according to the type and amount of information they contain. Identifying the correct type of each variable in an experiment is vital to the performance of the network. The principal categories are defined as nominal variables, ordinal variables and interval variables.

Nominal Variables Nominal variables are associated with named items. They do not have a numerical value. Measured values of a nominal variable do not have relationships like "greater than" or "less than." The only mathematical relationships relevant to nominal variables are equality and inequality.

Sex type is the obvious nominal variable. Numerical data can also be nominal. Postal codes and telephone area codes are classic examples.

To determine whether or not a variable is nominal, it is must established if an ordered relationship is to be implied by means of the values of the variable. If this is the case the variable belongs to a higher information category.

Nominal variables were presented to the network by using as many neurons as there are values that the variable can take on. Exactly one of the neurons will be turned on according to the value of the variable. All of the other neurons will be turned off. This is called *one-of-n encoding.* The only exception to this rule is if the variable is binary, taking on one of only two possible values. In this case, one neuron is used. It is turned on for one value, and off for the other.

Ordinal Variables Ordinal variables are above nominal variables in the "amount of information" hierarchy. Variables measured on an ordinal scale have order relationships. The values that can be taken on by an ordinal variable can be placed in a unique order. "Greater than" and "less than" have meaning. But the actual numerical values of ordinal variables *do not convey any information whatsoever beyond the order itself.* For example, suppose that we measure an ordinal variable for three subjects. The measured values are 3, 4, 250, respectively. All we know is that the second subject measures greater than the first subject on this variable, and the third subject measures greater than both of the others. We cannot say that the third subject is greater than the second to a larger degree than the second is greater than the first. The ability to do so would imply a higher type of variable than just ordinal.

When rank data is used, we cannot compare differences. In a class of 200, suppose person A ranks 20, person B ranks 25, and person C ranks 70. It might be that the grade point averages of persons B and C are very close, but their widely differing ranks are due to a cluster of people with approximately that average. We simply do not know. All we can say is that A is less than B, and B is less that C.

For the purposes of the present research the "percentile transformation" method was used for presenting rank data. For each measured value in the ranked group, count the number of cases in the group whose values are less than or equal to the case being transformed. Divide by the total number of cases. This gives a fraction ranging from nearly zero to one. This fraction is usually expressed as a percentage. In this way, we remove the effect of the group size, standardising the measurements to a uniform range of zero to one.

Ordinal variables may also be presented to neural network inputs as *ther-*

mometers. As many neurons are used as there are values for the ordinal variable, minus one. If that is too many, as would be the case when using percentiles, which can take on a great number of values, we use an "appropriate" number of neurons and quantify the variable. We might, for instance, use 9 neurons, breaking the percentiles into "less than 10," "10 to 19," etc. Arbitrarily assign an order to the neurons and map each neuron to an ordinal value other than the smallest. Then, for a given input value, turn on the corresponding neuron *and all neurons less than it.* If the input is the smallest value, leave all the neurons off. In the above case of 10 neurons representing a percentile score, an input value of 35% would be represented by turning on the first 3 neurons.

A thermometer representation of product grades is shown in Figure 7.14. The four product grades require three neurons. The lowest grade, D, turns off all three neurons. Higher grades turn on successive neurons.

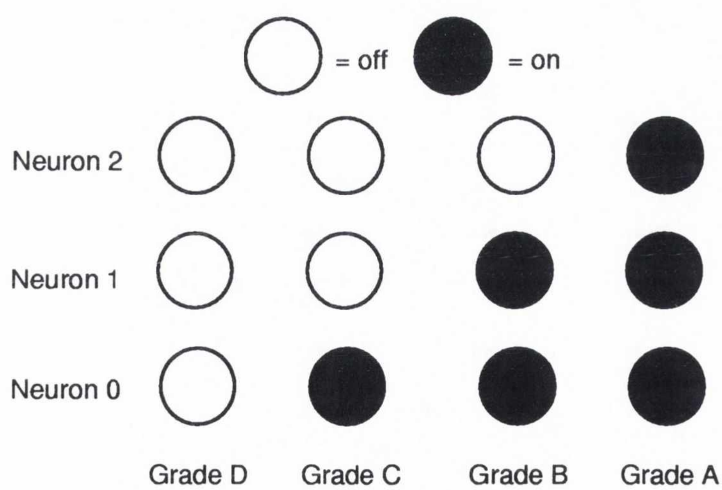


Figure 7.14: Thermometer representation of ordinal data.

There is no theoretical need for turning on neurons below the one designating the variable's value. Turning on exactly one neuron could convey the same information. However, experience has shown that training of networks is usually faster with this method. It more closely corresponds to the physical reality of ordinal data. Each neuron makes a contribution to a decision, and larger values retain the contributions of small values.

Interval Variables A higher order of measured data is achieved with *interval* variables also referred to as interval scales. These measurements share the property of order relationship with ordinal variables. If we have three interval measurements, A, B, and C, it makes sense to rank them as $A < B < C$ or some other ordering. But with interval data, we can go one step further: we can also rank *differences* between measurements. We can say things like $(C - B) > (B - A)$. Now, if $A = 4$, $B = 5$, and $C = 250$, we can genuinely say that C is very much larger than B, while B is only a little larger than A. This was not possible with ordinal data.

Variables measured on an interval scale are almost always presented to a neural network using exactly one neuron. The variables must be scaled in such a way as to be commensurate with the model's neuron-activation limits. Care must be taken that the scaling is done so that the data used in training will be commensurate with that used in testing. It would be incorrect, for example, to scale the training data based on the minimum and maximum values in the training set, then scale a batch of test data based on the extremes in the test set. The scaling must be done in a universally applicable way.

Probably the most common scaling method employed is simple linear mapping of the variable's practical extremes to the network's practical extremes. In the unusual case of a measured value going beyond the limit, the value would be truncated to that limit. Let the variable's maximum and minimum values expected in normal use be designated V_{max} and V_{min} respectively. Let the network's practical limits be A_{max} and A_{min}. For a feed-forward network with logistic activation functions, output activation limits would typically be 0.9 and 0.1, respectively. Inputs, of course, have no theoretical limits, but stability is usually improved by using comparable limits. An observed value V is scaled to a presentable value A with a simple formula

$$\begin{aligned} A &= r(V - V_{min}) + A_{min} \\ &= rV + (A_{min} - rV_{min}) \\ r &= \frac{A_{max} - A_{min}}{V_{max} - V_{min}}. \end{aligned} \tag{7.2}$$

If the variable was used to train an output neuron, it is necessary to unscale activation levels to obtain meaningful values for the variable when the network is used. This is done trivially by inverting the preceding formula

$$\begin{aligned} V &= \frac{(A - A_{min})}{r} + V_{min} \\ &= \frac{A}{r} + \left(V_{min} - \frac{A_{min}}{r}\right). \end{aligned} \tag{7.3}$$

7.3.2 Training of the Neural Networks

Two data sets were extracted from the research material outlined in Section 7.2.3 to train two separate neural networks. They reflect the effects on survivability respectively of patient variables and of implant variables (see Sections 7.2.3.1 and 7.2.3.2). Ideally these two data sets should be combined in a single data set where the effect of both patient and implant variables combined would be interpreted.

7.3.2.1 Patient-Related Neural Network

Purpose The aim of the patient-related neural network was to capture the richness of the information contained in the graphs described in Section 7.2.3.1. This was done by extracting a data set from the 24 graphs, which show the effect of patient-related parameters. This data set was then randomly divided into training and testing sets to be used in conjunction with the neural network. The patient-related neural network was able to use the inputted data to predict patient survivability outcomes. For example, to extrapolate the graph in Figure 7.6 which is truncated due to insufficient data. More generally it is applied to extrapolating the graphs past the length of time the experiment has been in progress.

In order to test the predictive ability of the patient-related neural network, data from one of the 14 graphs (that relating to women with the primary diagnosis hip fracture in the age range 55–64 years) was omitted and the neural network trained on data from the remaining graphs. The neural network was then asked to predict the missing case. The result, shown in Figure 7.15, is a convincing indication of the capability of neural networks in this respect.

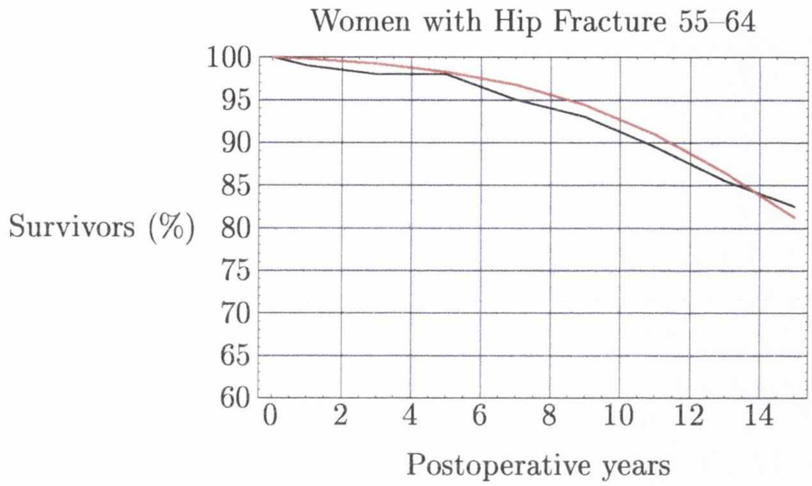


Figure 7.15: Survival rate for women with the primary diagnosis osteoarthrosis in the age group fifty five to sixty four: the survival rate predicted by the neural network is in red.

Construction The construction of the feed-forward neural networks was a matter of trial and error, guided by the discussion on hidden layers and hidden neurons in Section 7.3.1. The optimum structure of the patient-related neural network was found to have one hidden layer with three hidden neurons. The number of inputs in the input layer is six (see Figure 7.16). The first three inputs are nominal type variables and correspond to the variety of causes of total hip replacement (osteoarthrosis arthritis, rheumatoid arthritis and hip fracture) contained in the graphs outlined in Section 7.2.3.1. Only one of these three inputs can be activated at any one time. The fourth input corresponds to the sex of the patient and is a binary nominal variable: the patient is male or female. The fifth and sixth inputs are interval type

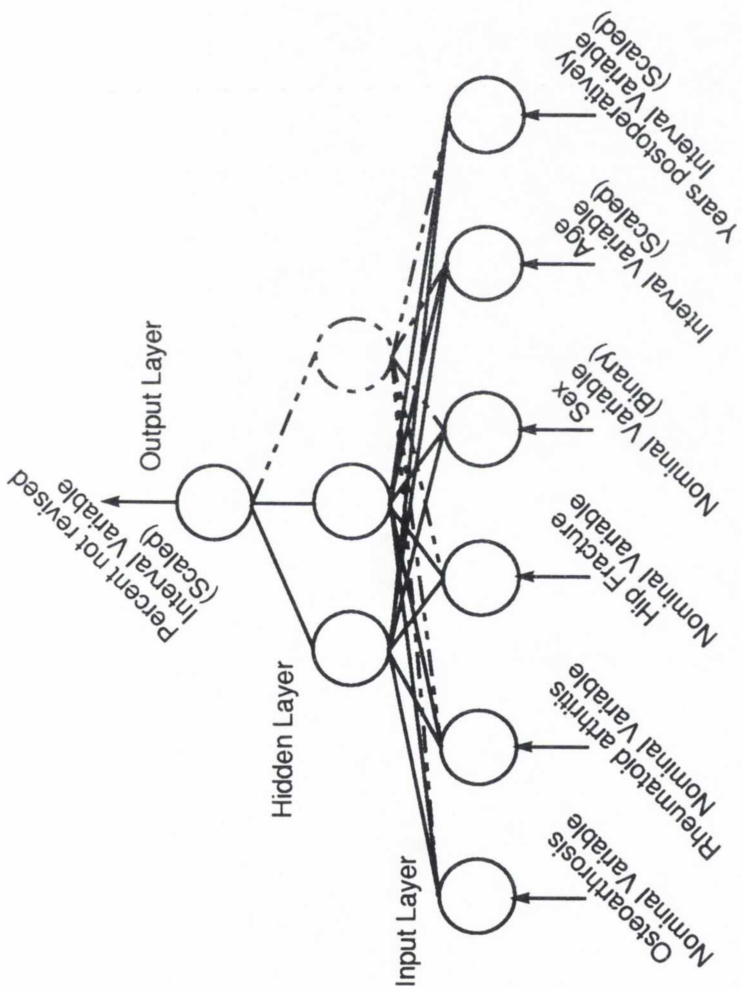


Figure 7.16: Patient-related feed-forward neural network structure.

variables and they represent the age of the patient and the number of years since the patient's operation. These input values are scaled to take on a value between 0 and 1. Finally, there is a single output which is an interval variable, representing the percentage of patients not revised. This output is scaled to take on a value between 0.1 and 0.9, and not between 0 and 1. This is to take account of the sigmoid or logistic activation function used by the feed-forward network.[6]

Operation With the structure of the patient-related neural network determined, the network was then trained to obtain a weights profile at the nodes that enabled the network to reproduce the characteristics of the patient-related graphs described in Section 7.2.3.1. The network is trained when the error, which is a function of the weights (see Section E.3.1), reaches stability and is below a certain value. However, it is also important to avoid over-training, an acceptable error was taken to be 0.01.

There are certain feed-forward neural network parameters that can be tuned to try and ensure convergence. Fortunately, the training of the patient-related neural network proved straightforward. The data set was readily amenable to being interpreted by a neural network model. Changing the parameters of the neural networks produces instructive changes in the rate of convergence to a solution. However, the networks converged and learnt the data set under most circumstances.

[6]The values 0 and 1 are not used in the input or output vectors. This is because the sigmoid function ($f(x) = \frac{1}{1+e^{-x}}$) asymptotically approaches the limits of 0 and 1 for infinite arguments. If we insisted that the actual network outputs attained the values of 0 and 1 , we could be iterating the weights forever, and they would grow to extremely large values (positive or negative). To avoid this problem, we let 0.1 represent the binary 0 state, and 0.9 the binary 1 state.

The graphs in Figures 7.17 and 7.18 represent the first and second training runs respectively: only a marginal increase in the rate of convergence was achieved when the number of hidden nodes was increased from two to three. Figure 7.19 represents the third training run and shows the effect of increasing the learning rate from .9 to 4. The larger value of learning rate results in a faster convergence.

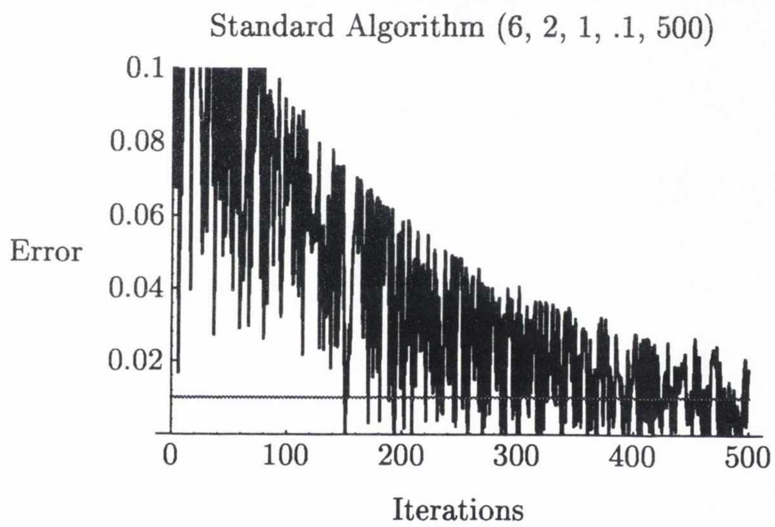


Figure 7.17: Output of first training run, patient-related network: 6 input nodes, 2 hidden layer nodes, 1 output node and learning-rate .1.

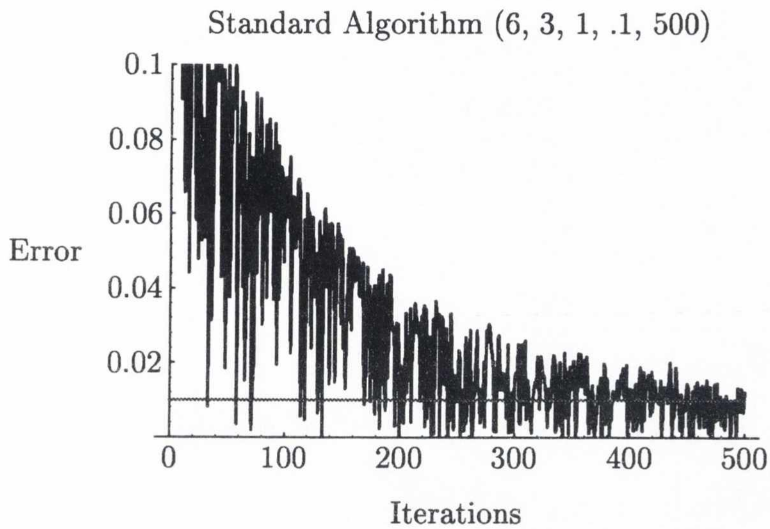


Figure 7.18: Output of second training run, patient-related network: 6 input nodes, 3 hidden layer nodes, 1 output node and learning-rate .1.

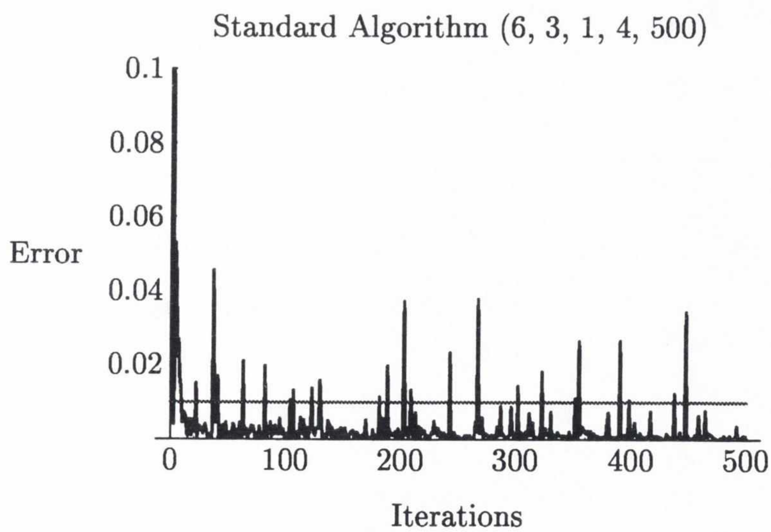


Figure 7.19: Output of third training run, patient-related network: 6 input nodes, 3 hidden layer nodes, 1 output node and learning-rate 4.

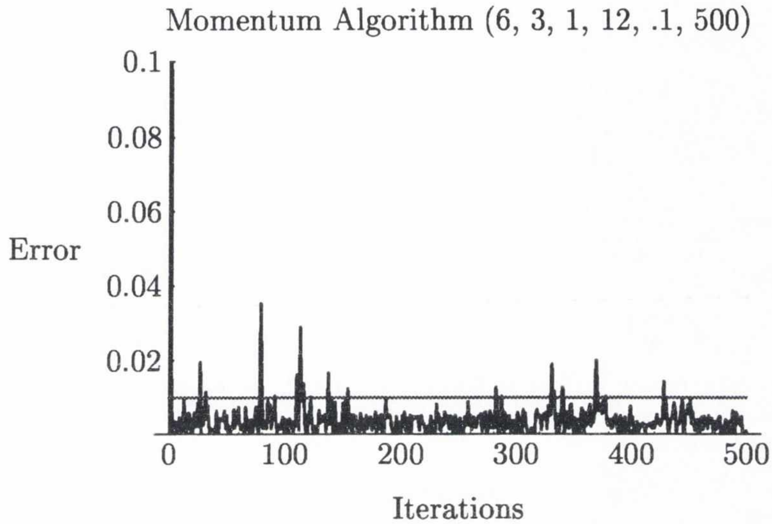


Figure 7.20: Output of fourth training run, patient-related network: 6 input nodes, 3 hidden layer nodes, 1 output node, momentum 12 and learning-rate .1.

The final graph, Figure 7.20, illustrates the effect of modifying the algorithm to add a momentum term to the weight-update equations. The mathematics of this are laid out in Section E.3.2.2. The rate of convergence was increased and no further modifications of the feed-forward neural network algorithm were deemed necessary.

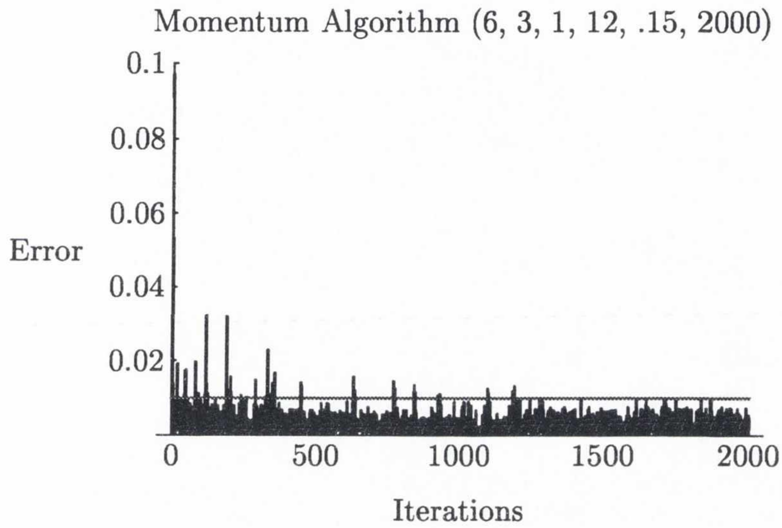


Figure 7.21: Output of fifth training run, patient-related network: 6 input nodes, 3 hidden layer nodes, 1 output node, momentum 12 and learning-rate .15.

Results The patient-related neural network had no difficulty in learning the intricacies of the patient data set as presented to it. This confirmed that the network would be capable of discerning much more complex patterns and more comprehensive data if and when these become available. The testing of this neural network is outlined in Section 7.3.3.

7.3.2.2 Implant-Related Neural Network

The implementation of the implant-related neural network has many parallels in that of the patient-related neural network. However, the uses and potential applications of the implant-related neural network are considered to be far more exciting.

Purpose As before, the aim of the implant-related neural network was to extract the fullness of information contained in the graphs described in Section 7.2.3.2. This was done by distilling a data set from the implant survivability graphs, which illustrate the effect of implant-related parameters on the life of the implant. This data set was then randomly divided into training and testing sets to be used in conjunction with the neural network. The implant-related neural network was able to manipulate the inputted data in two ways. The first was to extrapolate survivability outcomes for new implant designs that have only recently been fitted. The second was to postulate the survivability of proposed designs of implants. The implant-related neural network, when used in this way, has the capacity to gauge and anticipate the performance of proposed implants.

Construction The construction of the implant-related neural network proceeded in a trial and error manner adhering as far as possible to the guidelines outlined in Section 7.3.1. The optimum structure of the implant-related neural network had one hidden layer with two hidden neurons. The number of inputs and implant variables (the values of the input variables represent the characteristics of the prosthesis) are not as well defined as in the case of the patient-related neural network (see Figure 7.22) and may require adjustment and further experimentation as and if more and better data becomes available. For the purposes of this thesis, and the testing of this new application of neural networks, the number of inputs and implant variables was taken to be eight, five binary nominal variables and three interval variables (see Figure 7.22). These inputs correspond to the different classifications by which prostheses can be characterised as laid out by Cowley (1995) [23].

Cemented Metal femoral component, acetabular component polyethylene alone, or with metal backing. Components fixed into place with polymethyl methacrylate cement. Developed in the 1960s; improved cementing techniques in the late 1970s.

Ceramic Prostheses with ceramic heads and/or acetabular cups. Aim is to reduce wear. Cemented and cement-free types. Developed in the 1970s; more recent types have ceramic heads, polyethylene cups.

Press-fit Cement-free, metal femoral component, and metal-backed polyethylene acetabular cup; designed to achieve fixation by close geometric fit only. Developed in the late 1970s.

Porous-coated Cement-free, press-fit metal femoral component with metal-backed polyethylene acetabular cup, surfaces adjacent to bone coated with beads or mesh. The aim is to achieve fixation by bone in-growth into the pores in the surface coating. Developed in the early 1980s.

Hydroxyapatite-coated Cement-free, press-fit metal femoral component with metal-backed polyethylene acetabular cup, surfaces adjacent to bone coated with hydorxyapatite, believed to promote bone growth and form chemical bonds with the bone. Hydroxyapatite may be applied over a porous coating. Developed in the late 1980s; in the early stages of routine use.

Hybrid Cemented femoral component used with cement-less porous-coated acetabular cup. Aim is to reduce loosening of the acetabular compo-

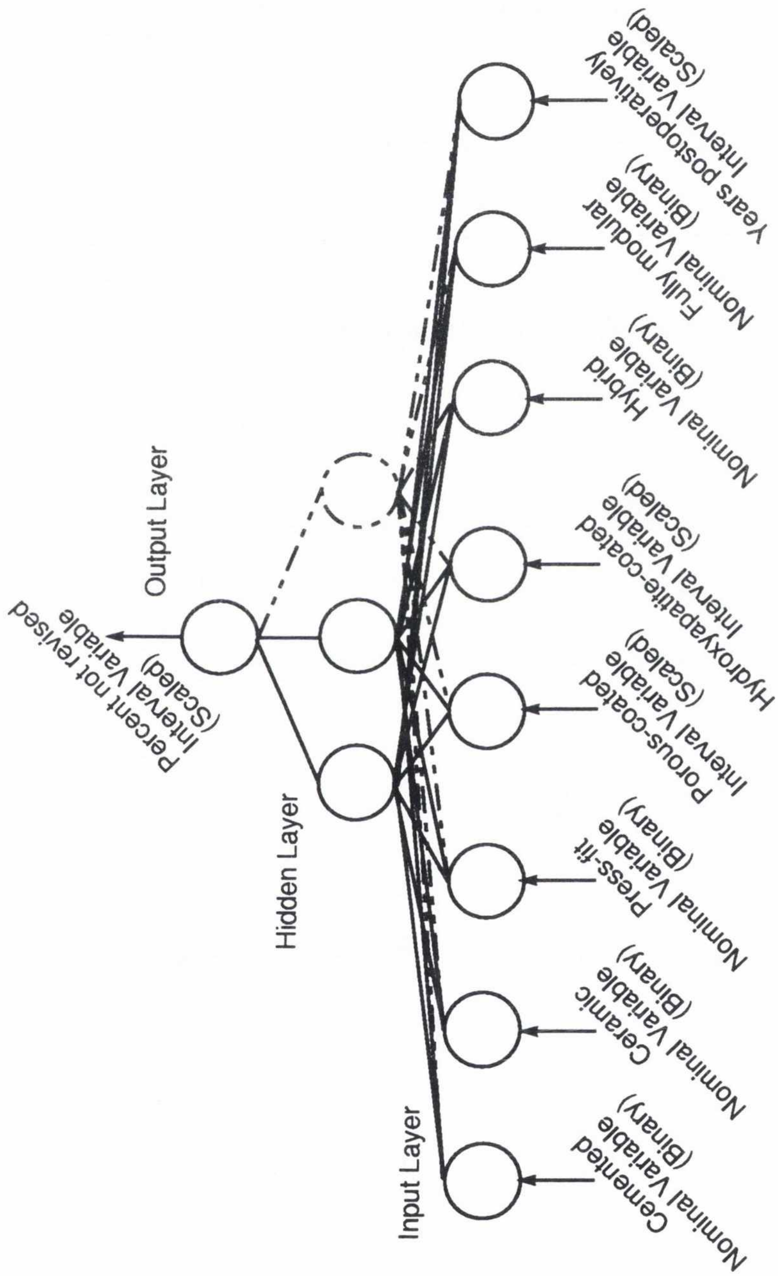


Figure 7.22: Implant-related feed-forward neural network structure.

Implant Type	Cemented	Ceramic	Press-fit	Porous-coated	Hydroxyapatite-coated	Hybrid	Fully modular
Charnley	✓	✗	✗	✗	✗	✓	✗
Lubinus IP	✓	✗	✗	✓	✗	✓	✗
CAD	✓	✓	✗	✗	✗	✗	✗
Christiansen	✓	✗	✗	✗	✗	✗	✗
Müller Curved	✓	✗	✗	✗	✗	✗	✓
Müller Straight	✓	✗	✗	✗	✗	✗	✓
Exeter Matte	✓	✗	✗	✗	✗	✗	✓
Exeter Polished	✓	✗	✗	✗	✗	✗	✓
Scan Hip Collar-Less	✓	✓	✗	✗	✗	✗	✓
Scan Hip Collar	✓	✓	✗	✗	✗	✗	✓
Spectron Metal-Backed	✓	✗	✗	✓	✗	✓	✓
Spectron All-Poly	✓	✗	✗	✓	✗	✓	✓
PCA	✗	✓	✗	✗	✓	✓	✓
Lubinus SP	✓	✗	✗	✗	✗	✓	✓

Table 7.2: Speculative implant variable profiles.

nent. Developed in the 1980s.

Fully modular Cemented or cement-free with porous or hydorxyapatite coatings. Modular components (such as stem, proximal sleeve, head) available in range of sizes and types, assemble inter-operatively to give best fit. Developed in the late 1980s.

Fourteen prosthesis survivability profiles (Charnley, Lubinus IP, CAD, Christiansen, Müller Curved, Müller Straight, Exeter Matte, Exeter Polished, Scan Hip Collar-less, Scan Hip Collar, Spectron Metal-Backed, Spectron All-Poly, PCA and Lubinus SP) are shown in Table 7.2. Eight input variables define each prosthesis, the seven listed in the table and an eighth, which was an interval variable, representing the number of years since the patient's operation. Finally, as in the patient-related neural network, there was a single output, of the interval variable (scaled) type, and this represented the percentage of patients that did not have revision surgery.

Operation Having decided on the implant-related neural network's structure, the network was next trained to obtain the weights profile at the nodes that would enable the network to reproduce the characteristics of the implant survivability graphs described in Section 7.2.3.2. A network is trained when the error, which is a function of the weights (see Section E.3.1), reaches stability and is below a certain predetermined value. An acceptable error was taken, as before, to be 0.01. The graphs in Figures 7.23, 7.24, 7.25 and 7.26 represent successive training runs for the implant-related network. They show the network error for illustrative combinations of network parameters.

The instability of the graph in Figure 7.25 is the result of setting the learning parameter too high. The graph in Figure 7.26 represents the same network with a learning parameter of 0.1. It can be seen that the network did not fully stabilise until after 300 iterations.

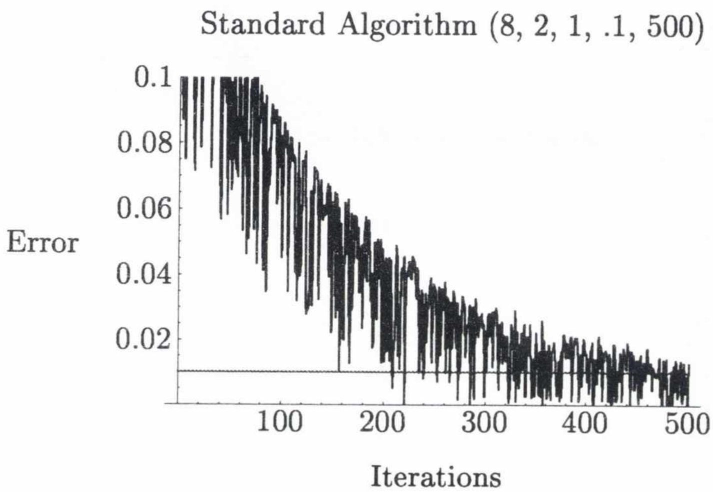


Figure 7.23: Output of first training run, implant-related network: 8 input nodes, 2 hidden layer nodes, 1 output node and learning-rate .1.

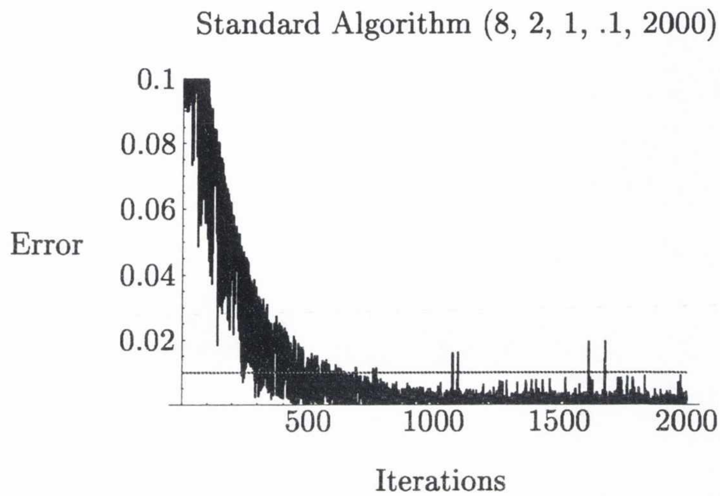


Figure 7.24: Output of second training run, implant-related network: 8 input nodes, 2 hidden layer nodes, 1 output node and learning-rate .1.

Results The implant-related neural network was successfully trained to recognise the characteristics of fourteen different prostheses. The analysis was based on eight input variables seven of which represented specific characteristics of the individual prosthesis. The remaining input variable and the output variable corresponded to the x and y axis values respectively, from the implant survivability graphs. Additional input variables can be introduced to the network and their effect on implant survivability measured. The testing of this implant-related neural network is outlined in Section 7.3.3.

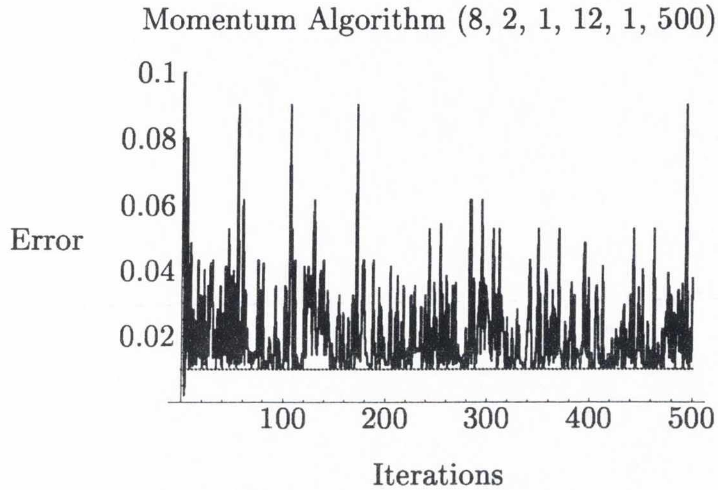


Figure 7.25: Output of third training run, implant-related network: 8 input nodes, 2 hidden layer nodes, 1 output node, momentum 12 and learning-rate 1.

Although the results were excellent, several reservations must be made about their interpretation. First, eight input variables were not enough to uniquely specify all fourteen prostheses. Four prostheses (Müller Curved, Müller Straight, Exeter Matte and Exeter Polished) had exactly the same implant variable profile and this does not reflect the differences between them. Similarly, both Scan Hip Collar-Less and Scan Hip Collar had the same profile. This raises the question: "Why not use more input variables to uniquely describe the implants?"

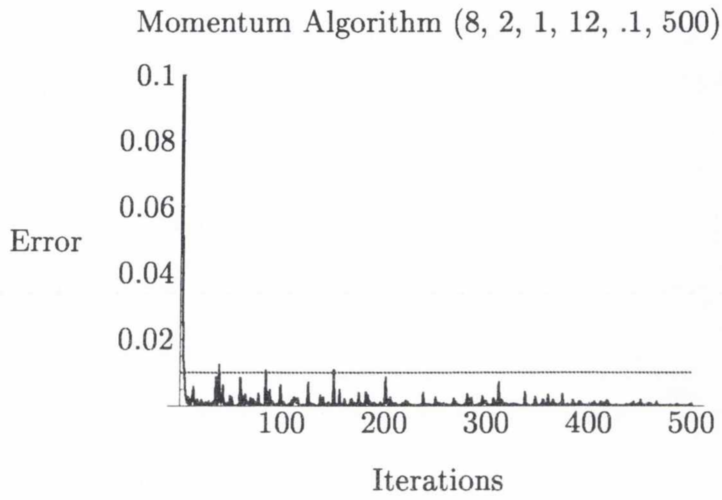


Figure 7.26: Output of fourth training run, implant-related network: 8 input nodes, 2 hidden layer nodes, 1 output node, momentum 12 and learning-rate .1.

This question brings to the fore the current practice of manufacturers who modify implants every few years, without changing their name [95]. This is a marketing issue which ensures continuity of brand recognition but invalidates the use of a prosthesis name. For example, in the Swedish Hip Register, the Charnley, Lubinus IP, Spectron Metal-Backed, Spectron All-Poly, and Lubinus SP prostheses come as both cemented and hybrid models. The PCA prosthesis is available in porous-coated and hydroxyapatite-coated models.

The variables in Table 7.2 are, therefore, only best estimates for the majority of implants listed. Furthermore, the success of the more recently developed implants might also be attributable to improved cementing techniques including cement pressurisation and a careful bone-bed preparation [83]. Thus operation-related variables should also be included. In view of this situation, further analysis of the effects of specific prosthesis characteristics would be superfluous.

Finally, a point on the relevance of adapting of this neural network for use by the surgeon in the period just prior to the operation. Ideally, this neural network needs to be trained on a data set whose records comprise the patient variables and the precise name of the implant. This would be sufficient to identify the best implant for a particular patient. However, since implant names are currently imprecise and do not uniquely specify the implant used, additional implant variables would have to be included to help distinguish between implants.

7.3.3 Testing Model Performance

7.3.3.1 Testing Neural Networks Performance

The mean square error of the outputs is the most universally accepted measure of performance for neural networks. Optimisation of this measure is the criterion against which the success of neural network training is normally judged. This is not surprising because the mean square error indicator can claim both practical and theoretical advantages.

The performance of a neural network is invariably tested with a different data set to the one on which it was trained. Statistical techniques that estimate only a few parameters relative to the number of training samples can often be exempted from this restriction. Although the strictest rules dictate that performance measures based on the training set are unfairly biased toward false optimism, in practice this may not be a serious problem for many traditional statistical techniques. However, the rule with regard to separate training and testing data sets for neural networks is rarely breached. The relatively large number of parameters involved (many weights) means that it would be all too easy for the network to concentrate on unique characteristics of the training samples, rather than generalising on the properties of the population.

Mean Square Error Many statistical techniques use mean square error as their basic measure of performance. It is easily computed by summing the squared differences between what a predicted variable should be and what it actually is, and then dividing by the number of components that went into that sum. Such a measurement has great intuitive appeal probably because it emphasises large errors more than small errors, a frequently valuable property.

More importantly, for models (statistical, neural, or otherwise) that are mathematically defined, the derivative of the mean square error can be far more easily computed than most other performance measures. This means that direct methods of optimising performance, such as linear regression, can often be easily done when the optimisation criterion is mean square error. Even in non-linear cases, such as feed-forward networks with non-linear activation functions, indirect methods are feasible. Optimisation of performance measures for which a derivative cannot be found is a far more expensive proposition. Finally, the mean square error lies close to the heart of the normal distribution. If errors can be assumed to be normally distributed, minimising the mean square error often corresponds to other very desirable optimisations. For these reasons, nearly all training algorithms for feed-forward networks (and many other models as well) rely solely on the mean square error as the object of their optimisation efforts.

The mean square error for neural networks, where we are concerned only with the output neurons, is relatively easily defined. For any input, the output neurons take on an activation level determined by the input and by the network. For that input there is a desired set of output activation. Suppose that we are processing case p. Let the correct (target) activation of output neuron j be designated as t_{pj}, and let the observed activation be o_{pj}. If there are n output neurons, the error for that single presentation is

$$E_p = \frac{1}{n}\sum_{j=0}^{n-1}(t_{pj} - o_{pj})^2. \tag{7.4}$$

If there are m presentations in the epoch, the error for that epoch is

$$E = \frac{1}{m}\sum_{p=0}^{m-1} E_p. \tag{7.5}$$

7.3.3.2 Calculation of Patient and Implant-Related Neural Network Error

The data sets for both the patient and the implant-related neural networks were randomly divided into training and testing sets. The testing sets were used in the calculation of the mean squared error (see Table 7.3) for the respective networks. The mean squared error was small for each network.

This proved the ability of the networks to learn the characteristics of the survivability graphs.

Neural Network	Mean Square Error
Patient-related	0.014905
Implant-related	0.019283

Table 7.3: Neural network mean square error.

In spite of the attainment of a low mean square error by a network, the mean square error is still only an indication of how the network is likely to perform. The ultimate test for a particular network is how it performs in service. The portents for success are certainly evident for the patient and implant-related neural networks, but it is only over time that their value can be confirmed.

7.4 CONCLUSIONS

The neural networks which have been built are clearly capable of discerning the complex non-linear effect of changes in input variables on survivability. The patient-related neural network successfully interpreted the effect of changes in patient profiles (age, sex, pathology) while the prosthesis-related neural network facilitated experimentation with implant variables to optimise survivability. The value of these capabilities to the surgeon (if and when confirmed in surgical practice) is of great significance.

The construction of a single neural network, however, which combines both patient and prosthesis survivability variables (and where possible other relevant variables) would be of considerably greater value. The constraints on achieving this objective are inherent in data limitations rather than in any problem intrinsic to the training and testing of neural networks. The researcher did not have direct access to the raw data of the Swedish Hip Register, but had to rely on indirect access through two discrete sets of graphs, one relating to patients and the other to implants.

A second limitation of the available data further inhibited the capacity to build a neural network to adequately exploit the obvious potential of this medium. Neural networks have achieved outstanding success with problems which are characterised by a large number of interdependent variables. The implant selection process with its very large number of input variables, is prototypical of such a problem. Nevertheless, only eight variables could be reliably identified from the data available, for the design of the implant neural network.

This is because of incomplete specification of design variations of individual prosthesis types. Generic designs, such as the Charnley, manifest many modifications and refinements over time and these are not, as a rule, captured in the Hip Register. An example of a slight modification in design having far reaching results is the Exeter stem: the change from matte to polished stem resulted in dramatically different long term results. Therefore, it is vital that the exact specification of the prosthesis used should be communicated to the neural network, if the highly non-linear relationship that exists between the patient/prosthesis variables is to be uncovered.

Chapter 8

Discussion, Conclusions and Recommendations

For every complex problem there is a solution which is neat, simple and wrong.
H.C. Menken

There is a time in the life of every problem when it is big enough to see, yet small enough to solve.
Mike Leavitt

8.1 DISCUSSION

8.1.1 Background

The surgical operation of THR is, after 30 years, a well established medical procedure. Yet, paradoxically, it is becoming less satisfactory with the passage of time as the expectations of patients are increasingly being frustrated. The average life of an artificial hip has remained at about 15 years over the last decade whereas, during the same time period, the average life expectancy and quality of life of the elderly has improved greatly, leading to widespread demand for hip replacements which last longer. More significantly, revision rates for different prosthesis designs, and for different surgical procedures, vary enormously and unpredictably. For example, in the UK there are some 62 primary THRs, ranging in price from UK£250 (€360) to UK£2000 (€2860), and manufactured by 19 separate companies. But there is no attempt to establish a correlation between the price of a THR and its survivability: 30 per cent of prostheses introduced in the past 5 years have not been evaluated in peer-review journals [97].

The case has been made in this thesis for a comprehensive scientific evaluation of potential patient/prosthesis combinations in the time frame immediately before the operation. Up to now, such evaluations have focused on finite element analysis and a variety of mechanical testing techniques carried out at the design stage of the prosthesis and, to a lesser extent, on statistically-based clinical evaluation after the operation. But a substantial amount of new information becomes available at the time of the operation itself, principally the patient's bio-data and pathology and the choice of surgical procedure available to the surgeon. Given recent advances in information technology and artificial intelligence, which enable all the different strands of data to be effectively co-ordinated and mined, it is difficult to defend a situation where a surgeon continues to select a prosthesis, largely on the basis of brand image and familiarity with a particular design.

The search for a more scientific, knowledge-based solution to the prosthesis selection problem, faced by the orthopaedic surgeon in the time frame just prior to the operation, has opened up some exciting vistas. Individually none of the instruments developed in this thesis can claim to be a neat, one-size-fits-all solution but, collectively, they are a significant advance and an optimistic basis for closer collaboration and co-operation between the medical and mechanical engineering professions. The hip prosthesis is a mechanical device performing an important biological function in the human body. Progress at this pre-eminent interface between the two professions can only benefit from a more intensive understanding and application of the analytical and computer-based techniques which are now the bedrock of engineering design.

8.1.2 Concept of Pattern Classification

This thesis proposed a framework for a systems solution to the prosthesis selection problem. All of the approaches examined fall into the data mining category of *classification.* The construction of a classification procedure from a set of data with known categories (e.g. good, satisfactory, poor) has been variously termed pattern recognition, discrimination, or supervised learning (in order to distinguish it from unsupervised learning or clustering where the classes are inferred from the data). The data associated with a particular patient/prosthesis combination are interrogated to identify patterns that would allow the combination to be put into a particular classification. That classification can range from a binary decision (to accept or to reject) to a multiple-class model that allows the combination to be placed in one of a number of classifications, as well as the possibility of rejection.

The three main historical strands of *classification* research are: machine learning, neural networks and statistics [86]. These have become the prerogative of different professional and academic groups, and as a result, the issues explored by each of the groups tend to diverge. All groups have, however, had some objectives in common. They have all attempted to derive procedures that would be able:

- to equal, if not exceed, a human decision-maker's behaviour, while having the advantage of consistency and, to some extent, explicitness,
- to handle a wide variety of problems and, given enough data, to be extremely general in their range of application,
- to be successfully used in a practical setting.

8.1.2.1 Machine Learning

Machine learning is generally taken to encompass automatic computing procedures based on logical or binary operations, that learn a task from a series of examples. Machine learning aims to generate classifying expressions simple enough to be easily understood. They must mimic human reasoning sufficiently to provide insight into the decision process. Like statistical approaches, background knowledge may be exploited in development, but operation without human intervention is assumed. After a period of steady growth, machine learning has reached practical maturity under two distinct headings: (a) as a means of constructing rule-based software (for example in "expert systems") from sample cases volunteered interactively and (b) as a method of data analysis whereby rule-structure classifiers, for predicting the classes of newly sampled cases, are obtained from a "training set" of pre-classified cases. This thesis confined itself to the subset (a) of machine learning (this comes down to the use of the symbol system of *predicate* as opposed to *propositional* logic). This is the construct underlying the chapters on rule-based (Chapter 5) and fuzzy expert systems (Chapter 6). The subset (b) of machine learning was not fully explored because it requires access to extensive training-set data as well as the enumeration of rules in conjunction with an expert. (Hence the neural network alternative is more appropriate for current purposes).

8.1.2.2 Neural Networks

The next category of classification procedure exploited in this research was neural networks. The versatility and sophistication of neural networks provide the capability to circumvent the limitations of machine learning systems, *which require explicit rules.* They can also be used, of course, to draw inferences from historical data thus replicating the function of statistical analysis.

Neural networks provide a key to unlock information embedded in extensive medical data-bases. They combine the complexity of some of the statistical techniques with the machine learning objective of imitating human intelligence: this is done, however, at a more "unconscious" level and hence there is no accompanying ability to make the concepts which have been learnt transparent to the user. Neural networks hold considerable promise in further unearthing the complex relationships that pertain between patient and prosthesis variables and their impact on the outcome of the THR operation.

8.1.2.3 Statistics

Statistical approaches to classification perform a valuable function in the analysis and the drawing of inferences from historical data as instanced in papers analysing the data in the Swedish Hip Register [1, 82, 83]. Statistical analysis, using advanced regression techniques, revealed the relative importance of the different factors which contribute to the success of the operation.

Statistical methods were not examined in this thesis, however, because they are not the most appropriate for use by the surgeon in the time frame before the operation. Statistical approaches are generally characterised by an explicit underlying probability model, which provides a probability of being in each class rather than a simple classification. In addition, it is usually assumed that the techniques will be used by statisticians and, hence, outside specialist intervention will be required with regard to variable selection and transformation and overall structuring of the problem [86].

8.1.3 Comparative Analysis of the Three Approaches

In this thesis, an expert system, a fuzzy expert system and a neural network were constructed to solve the same problem (see Figure 8.1): the matching of a hip prosthesis to an individual patient. Each of the procedures provided significant insights but success was ultimately circumscribed by the quality of the input data. The problem is a difficult one and is characterised by a multiplicity of factors some of which are poorly understood. The data

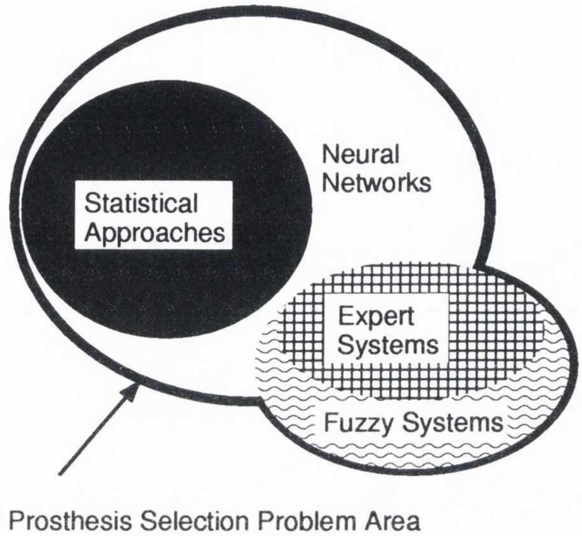


Figure 8.1: Hip Prosthesis Selection Problem Area.

available are incomplete and imprecise and the problem space is subject to continuous change (new and variant prostheses are being continually introduced). Whereas each of the three approaches has been successfully applied to better-structured problems, limitations became apparent when applied to the highly complex and ill-defined problem of hip prosthesis selection.

It is proposed to first highlight the positive features and the limitations of each approach, as well as its longer-term potential to contribute to the solution of the prosthesis selection problem. This will be followed by an appraisal of the relative merits of the machine learning as opposed to the neural network solution.

8.1.3.1 Expert and Fuzzy Expert Systems

The expert system approach raised great hopes initially. Its rules are readily comprehensible to surgeons and it is possible to start on a modest scale and expand as the procedures become more familiar to the participants. It attempts to emulate the human approach to problem-solving and has been successfully applied in medical diagnostics and in a number of other fields of human endeavour. Paradoxically, it was the inability to capture the richness and flexibility of surgeons' decision-making that contributed to the relative failure of the expert system developed in the course of this research.

1. It did not prove possible to build a modicum of *common sense* into the model to legislate for exceptional situations which can arise, such as no feasible solution (or a solution outside the expertise of the professionals directly involved).

2. The absence of a *learning dimension* makes expert systems unattractive in a discipline and an environment where change, adaptation and learning are of the essence.

3. The *unidimensional* character of expert systems and their reliance on symbolic inputs to the exclusion of sensory experiences and any element of creativity is at variance with the holistic culture of medical diagnostics.

What then is the prognosis for machine learning in the context of the patient/prosthesis selection problem? First, it is proposed to outline the minimum conditions necessary if the prosthesis selection problem is ever to come within the province of an explicit rule-based solution. A comment will be made on the extent to which these conditions are satisfied in the present instance. Next an attempt will be made to list some implications and assumptions inherent in a machine-based solution for a medical condition, with a view to highlighting the logical and emotional reactions which are likely to arise.

The conditions necessary for the development of the system are:

1. Representative surgeons and their patients available and willing to participate in the process and prepared to submit to a very intrusive ordeal which runs contrary to the cherished medical code of confidentiality.

2. Experts submitting to a series of hypothetical questions with a view to eliciting the reasoning logic once possessed but probably lost by them in the process of becoming experts.

3. A supportive clinical environment and a team of medics and paramedics actively co-operating with an experiment which may be perceived as presaging fundamental changes in work practices even to the extent of de-skilling.

It is not surprising that the laboratory conditions experienced by the researcher did not fully meet with this specification. The sample of surgeons who co-operated was relatively small and even this co-operation was in part a trade-off for earlier collaboration. Medical practitioners work under extreme pressure in a busy hospital environment where the urgent can soon sweep away the most genuine good intentions to participate.

It is now proposed to examine the implications and assumptions inherent in applying a rule-based expert systems to the prosthesis selection problem (and to any problem where the issue of human health, indeed of human life, is involved).

1. The solution of the expert system, though manifestly satisfactory, may fall somewhat short of what is capable of being achieved by human experts at the highest level.

2. There is an assumption, on the other hand, that human experts can at times make inordinately bad decisions and that at least this scenario is avoided when applying an expert system.

3. Ideally rate of advancement in a particular field of medicine is assumed to be such as to justify freezing best practice for some indeterminate period. This is not the case in hip prosthesis selection, which at the moment is in a state of flux.

4. An expert system does not know when a problem is outside its domain whereas a human expert knows to refer the problem on.

There are manifest causes for concern, for both patient and surgeon, in the implications of some of these assumptions. The human survival instinct is very strong and no avoidable risk, however small, is acceptable when it comes to matters of human health and quality of life. There is no doubt that expert systems have the capability to significantly improve the overall standard of decision-making in prosthesis selection. But this capability will take time and patience to reach fruition. The main challenge is to win the trust of patient and surgeon during the transition period.

In practice, expert systems have performed best where the depth of knowledge is greater than the breadth of knowledge, and where the content is specific and the knowledge well understood [131]. This can not be said to apply at the present time in the field of hip prosthesis selection. Even the addition of fuzzy rules, to try and legislate for the vagueness of expert statements, does not overcome the essentially poorly understood nature of the area.

8.1.3.2 Neural Networks

Neural networks are flexible tools pre-eminently suitable for a dynamic environment. They have the capacity to learn rapidly and change quickly as circumstances, data values and outcomes evolve. These attributes can be summarised as follows:

1. A capacity to *learn*. Where appropriate training data are available neural networks can learn from the patterns inherent in these data, thus obviating the need for programming by an analyst.

2. A capacity to *predict future behaviour* from patterns observed in historical data. Pattern recognition is a powerful technique for capturing the information in data and deriving conclusions from it.

3. *Flexibility* to respond to a changing environment. Although neural networks may take time to absorb a sudden or radical change, they are excellent at adjusting to constantly changing information.

4. The rapid growth of hospital and national medical data-bases opens up promising potential for the application of neural networks.

5. Neural networks are computationally intensive and originally were run on supercomputers but their routine is now optimised to the point where they can be run in reasonable time on the current generation of personal computers.

Inevitably there are also certain limitations associated with neural networks:

1. Not withstanding the fact that excellent results have been achieved in many applications neural network analysts are unable to satisfactorily explain the behaviour of their models. Experimentation with different parameters achieves some understanding of model behaviour and increases confidence in the results achieved.

2. Linked with this is the inability to explain, in the absence of explicit rules, the basis on which decisions are made. This has potentially serious ramifications in an era where patients can insist on their right to be informed about decisions which affect their lives and where medical outcomes are increasingly challenged in the courts.

3. Neural networks depend critically on the analyst's understanding of the problem being modelled and on having training data which accurately represent the problem. Incomplete data was a serious issue for the current research and this problem is also widely reported in the literature.

The results achieved with neural computing are only as good as the data used to train the system.

Comparison of Neural Networks and Rule-Based Methods

1. Rule-based systems have proved to be viable in medical applications where the problem can be presented in a structured way. Knowledge in the area of patient /prosthesis selection has not yet reached this degree of codification but this thesis offered some important advances in the formalisation of knowledge in the area.

2. The capacity of rule-based expert systems to explain their decisions is of immense importance in achieving recognition and acceptance by medical experts and in keeping medical decisions away from scrutiny by the law courts. It is the responsibility of the system designer, in conjunction with the expert, to elicit rules that are realistic and acceptable to the medical profession. This characteristic of rule based system contrasts with the lack of transparency that exists in the case of neural networks.

3. The chasm, however, between the impersonal, desiccated quality of a decision produced by a computer, relying on quantification and symbolic input, and the humane approach associated with personal intervention by a specialist is an inordinate gap to be bridged. In practice, however, a medic will always be required to impart and interpret the decision of an expert system.

4. Consequently, further development of rule-based expert systems is constrained by the difficulty of gaining the trust of medical professionals whose multifaceted expertise has to be encapsulated in explicit decision rules. Hip prosthesis selection expertise is not only predicated on a knowledge-base which spans the disciplines of medicine and mechanical engineering but also the surgeons' own operational dexterity. This may prove a step too far in the case of patient prosthesis selection.

5. Neural networks, on the other hand, are an extremely powerful and flexible instrument capable of mining immense quantities of historical data and reacting to changing environments. They perform statistical functions better than traditional statistical methods when the form of the data is unknown or non-linear, or when the problems are complex with highly inter-related relationships.

6. Neural networks are critically dependent on the availability of good quality, consistently compiled data. These data are especially important when training the system and, in the case of the current research, incomplete data were a major constraint.

7. The implementation of a neural network from Irish hip selection experience is not an option because of a dearth of data. The future for neural networks looks very promising, however, as many countries are moving quickly towards establishing national medical data-bases and comprehensive patient information systems, which cover all aspects of patient care and management. In Ireland such a system has yet to be realised.

8. The neural network approach offers the best option for the immediate future of carrying out a comprehensive evaluation of possible patient/prosthesis combinations just prior to the operation. It's main advantage, is that it obviates the need to interrogate experts about their choice of rules.

8.1.4 Summary

The way knowledge is captured and represented for a particular problem may well depend on the degree of knowledge formalisation and codification that has occurred in that field. This concept is illustrated in Figure 8.2 using the experience of this study as an example. It is a well-documented fact [133]

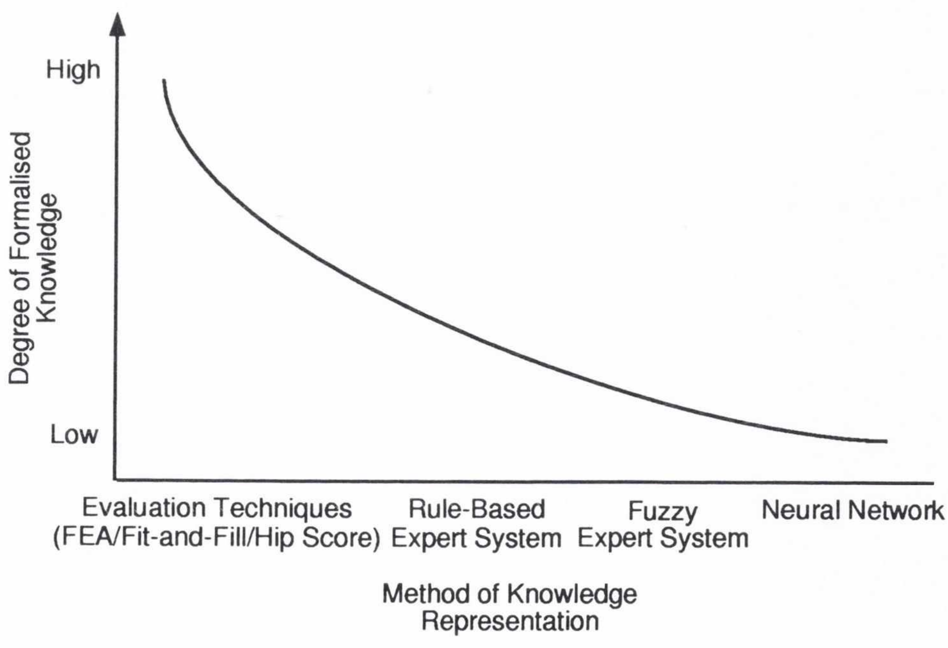


Figure 8.2: Knowledge refinement diagram.

that new technologies go through a period of flux with a variety of competing designs on the market before a "dominant design" eventually emerges (e.g. motor cars, computers, video recorders etc.). During this emerging phase there is a great deal of experimentation and many false starts precisely because the technology has not crystallised into a well-defined and mature state [132]. The proposition is that the field of total hip replacement has remained in this pre-paradigmatic condition for an inordinate time. There is reason to hope that continuing improvements in data mining capability, and particularly the recent progress in neural network techniques, will move prosthesis technology to a more mature and sophisticated plane.

8.2 CONCLUSIONS

- Comprehensive evaluation, just prior to the operation, of possible patient/prosthesis combinations, is essential to improve the success of the THR operation. A portfolio of quantitative evaluation techniques for this purpose now exists.

- THR evaluation methods differ fundamentally and further automation of some of the methods (e.g. fit-and-fill analysis) is feasible and necessary. Expertise on the evaluation methods tends to reside in different professional disciplines and this has impeded their harmonious development.

- The calculation of an overall evaluation, based on the outputs of the separate evaluation techniques, is not a trivial problem. A simple weighting of the evaluation output components, to arrive at an overall evaluation, did not capture the complexity of the medical decision-making process.

- A rule-based expert system offers the best means of achieving the correct overall evaluation in the immediate future. However, the lack of familiarity on the part of surgeons with the quantitative evaluations being used undermines medical confidence in the value of the system.

- The fuzzy expert system added some enhancements to the rule-based system by incorporating flexibility and allowing for an element of uncertainty in the rules elicited from the experts. But this advantage cannot be exploited until surgeons accept the intrinsic value and potential contribution of expert systems. The overall problem of gaining the surgeons' confidence remains.

- Neural networks, which focus on mining historical THR records, are manifestly a valid and useful tool for the evaluation of THR survivability. Survivability was taken to be the most objective measure of THR success. The problem was formulated so that the output of the objective function was the survivability of the THR. Four distinct groups of variables have an important influence on survivability: patient, implant, hospital and surgeon. This thesis dealt solely with patient and implant-related factors because these were the only areas for which data were available.

- The modelling of patient-related factors in isolation is not of great practical value. Patient factors cannot be changed at the time of the operation and the best that can be achieved is an indication of how similar patients performed in the past. This can equally be done by traditional curve fitting. Prediction of survivability for a given type of patient in cases where the available historical data is truncated was shown to be an area where neural networks can make a valuable contribution.

- The modelling of the impact of implant factors on hip survivability was of immense practical importance because this made it possible to evaluate a range of prosthesis designs. The neural network was able to predict the survivability of different designs when it was supplied with different combinations of inputs, delineating the characteristics of individual implants. The network was trained to gauge the effects of these characteristics, and combinations of these characteristics, on survivability. This is the type of complex task for which neural networks are pre-eminently suited. Neural networks are sensitive to subtle interactions that are not transparent to human observers.

- The network used eight variables to enable the neural network to differentiate between fourteen prosthesis types (see Section 7.3.2.2). The number of variables selected can be increased considerably and the neural network has no difficulty in factoring in the effects of the new variables.

- This opens up the exciting possibility of being able to estimate the survivability of new prostheses with speculative combinations of variables.

- The success of the separate patient and implant neural networks drew attention to the great potential of a neural network which could combine both patient and implant variables. The data required to implement such a network were not, at this stage, forthcoming to the researcher. Such a network would be particularly useful to the surgeon just prior to the operation because it could be constructed to provide information about the particular patient/prosthesis combinations being contemplated. The main limitation of such a network is an inability to explain its reasoning. However, neural networks in other fields, which have shown a proven ability to provide the correct answer, have been readily accepted by their users.

- One very obvious benefit for the patient to accrue from the use of neural networks is the ability to obtain reliable and timely information on the probable survival time of the particular prosthesis being recommended. This will empower the patient to make a decision on whether to proceed with the operation or not.

- The neural network does not replace the portfolio of other prosthesis evaluation methods (mechanical testing, tribological testing, finite element analysis etc.) but rather it provides a significant additional technique which enhances the capability to evaluate and select prostheses.

8.3 RECOMMENDATIONS

- The imperative of evaluating patient/prosthesis combinations in the timeframe just prior to the THR operation must be recognised. This is the stage where decisive new information (anatomic details, patient pathology, surgical options etc.) becomes available. Failure to analyse and evaluate all relevant data leads to sub-optimal decisions and is detrimental to the long-term advancement of THR as a field of medical expertise.

- It is recommended that the surgeon be empowered by putting at his disposal a range of evaluation techniques including FEA, fit-and-fill analysis, clinical hip scores analysis and radiological hip score analysis. At present the surgeon does not have the capacity or the time to organise or interpret the results of these evaluations because the expertise is distributed widely among selected engineering and paramedical sub-disciplines.

- The preferred way to empower the surgeon is the rule-based expert system which was constructed and validated in the course of this research (see Chapter 5). Initially this expert system can be fine-tuned and prototyped with the three most highly developed and automated evaluation techniques, namely FEA, fit-and-fill analysis and clinical hip score analysis. These are the techniques for which it is possible, given the current state of knowledge, to elicit rules which are realistic and acceptable to the profession.

- It is recommended that additional evaluation techniques should be progressively added to the model as these are improved and automated. All the evaluation components require further research and refinement. A neural network module can in due course be integrated into the rule-based expert system, but this is some way into the future since neural network evaluation, as of now, is predicated on a separate statistical population with its own characteristic surgical techniques (Swedish Hip Register).

- The concept of employing an expert system for medical diagnostics is well established and after a period of experimentation and intense consultation is likely to prove acceptable to the orthopaedic fraternity. Successful medical applications of expert systems already permeate many parts of the profession: well established examples include INTERNIST/CADUCEUS for diagnosing diseases in internal medicine and MYCIN for diagnosing blood infections.

- It is recommended that the preferred long-term deployment of the neural network technique is as a stand-alone or alternative evaluation paradigm. The neural network is a powerful data-mining technique in its own right, capable of incorporating patient, prosthesis, surgical practices, hospital regimes and other key variables and of recognising patterns and relationships in historical data. The integrity and the consistency of these data are a primary consideration but the way different countries design and assemble their medical databases is not yet harmonised [83].

- It is recommended that efforts to gain direct access to the raw data contained in the Swedish National Hip Register be redoubled. These data, despite some limitations, are consistently compiled and facilitate the implementation of a neural network combining both patient and implant variables and, ultimately, hospital and surgeon-related variables also.

- The range of prosthesis design variables on which data are now recorded should be extended in the light of experience gained in constructing and testing neural networks. The number of prosthesis variables which can potentially affect the outcome of a THR has been shown to be considerably greater than is currently recorded. (It is highly probable that a similar comment can be applied to recorded data on other orthopaedic implants, in particular the knee and the elbow, for which the neural networks developed in this thesis have compelling relevance).

- Computerisation has resulted in the proliferation of medical databases nationally and internationally and is giving renewed impetus to the theory and practice of neural networks for the mining of these data. In the context of the pressing medical and social demand for more rapid progress in THR it is important that this vital area of human concern should not fall further behind.

8.4 FURTHER WORK

- All the evaluation techniques critiqued in Chapter 4, some more than others, require development and adaptation to keep pace with rapid advances in software capability and computing power. For example, FEA and fit-and-fill analysis, currently based on 2-D imaging, need to be converted to use 3-D CAT scans (see Section 4.1.1.1); and clinical hip score analysis is particularly disadvantaged by data limitations. Many of the improvements necessary can only be achieved in operating conditions and through a close working relationship between the surgeon and the domain specialists.

- Progress in refining the rule-based expert system must go hand-in-hand with improvements in the evaluation techniques because the rules are only as good as the data on which they are based. It is estimated that reliable data are needed to sustain about 150 rules, as opposed to the 30 rules prototyped in the research model. The expert system can only be refined and validated in realistic laboratory conditions.

- Further work on fuzzy expert systems must await the prior development and broad-based acceptance of a comprehensive rule-based expert system.

- The potential and the need to consolidate the neural network approach introduced and progressed in this thesis is obvious. The first step is to combine patient and prosthesis variables and then to add other variables as the neural network is trained and as the data become available:

this would yield much useful information on the complex variable interactions and dependencies involved in THR. A neural network easily accommodates this increase in complexity.

- The most crippling impediment to progress in THR research and practice is the absence of agreement on how to assemble and regulate the content of prosthesis data-bases. This deficiency is seriously impairing progress and has been repeatedly highlighted in editorials and in articles by a succession of eminent authorities. Disagreement exists at the most fundamental level. What is the best measure of THR success? Should failure by recognised at the onset of moderate pain [11] or when the necessity for revision occurs [83]?

- The standardisation and harmonisation issue must be urgently addressed as considerable resources are now being allocated to the computerisation of hospital records. The present Labour government in Britain has already committed itself to the full scale computerisation of all records in their National Health Service. Hopefully, such an action is imminent in Ireland.

- If progress is not made on these fundamental issues it is likely that THR may be further outstripped by other areas of medical technology. A number of countries, in addition to the UK, are now committed to full computerisation of their health service records. IBM has recently set up a dedicated research group to apply data mining techniques to medical data bases [85]. This decision was motivated by a turning point in the man *versus* machine epic: when an IBM parallel computer defeated the human chess grandmaster Kasparov. The new machine was capable of examining 50 billion chess positions in the three minutes nominally allocated for a single move. When this incredible processing power was harnessed to the recorded knowledge of computer scientists and chess champions, it created a watershed in the history of artificial intelligence.

Appendix A

Prolog Expert System Source Code

This appendix contains the Prolog source code listings.

```
/* Suitability Evaluation

suitability(Prosthesis,Answer) :-
      Answer is the reply to a request by Prosthesis
      for suitability. */

suitability(Prosthesis,Answer) :-
      ok_profile(Prosthesis),
      fit_rating(Prosthesis,FitRating),
      hip_score_analysis(Prosthesis,HipScoreAnalysis),
      fea_analysis(Prosthesis,FEA),
      evaluate(profile(FitRating,HipScoreAnalysis,FEA),
            Answer) , !.
```

```
/* The Fit Rating Module

fit_rating(Prosthesis,Rating) :-

      Rating is a qualitative description assessing

      the fit offered by Prosthesis to cover the request

      for suitability. */

fit_rating(Prosthesis,Rating) :-

      fit_profile(Prosthesis,PrimaryDimensions,SecondaryDimensions,

            TertiaryDimensions),

      fit_evaluation(PrimaryDimensions,SecondaryDimensions,

            TertiaryDimensions,Rating).


fit_profile(Prosthesis,PrimaryDimensions,SecondaryDimensions,

      TertiaryDimensions) :-

      requested_suitability(Prosthesis,Suitability),

      fit_percent(primary_dimensions,Prosthesis,Suitability,

            PrimaryDimensions),

      fit_percent(secondary_dimensions,Prosthesis,Suitability,

            SecondaryDimensions),

      fit_percent(tertiary_dimensions,Prosthesis,Suitability,

            TertiaryDimensions).


fit_percent(Type,Prosthesis,Total,Value) :-

      findall(X,(fit(Fit,Type),

amount(Fit,Prosthesis,X)),Xs),

      sumlist(Xs,Sum),
```

```
      Value is Sum*100/Total.

/* Evaluation Rules */

fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,good) :-
      PrimaryDimensions =< 100.
fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,excellent) :-
      PrimaryDimensions < 70,
      PrimaryDimensions + SecondaryDimensions =< 100.
fit_evaluation(PrimaryDimensions,SecondaryDimensions,
      TertiaryDimensions,excellent) :-
      PrimaryDimensions < 40,
      PrimaryDimensions + SecondaryDimensions < 70,
      PrimaryDimensions + SecondaryDimensions +
            TertiaryDimensions =< 100.

/* Arthroplasty Data - Classification of Fit */

fit(isthmus,primary_dimensions).
fit(femoral_head_offset,primary_dimensions).
fit(canal_width,secondary_dimensions).
fit(canal_width_plus_20,tertiary_dimensions).
fit(canal_width_minus_20,tertiary_dimensions).

/* Hip Score Analysis Rating Module
```

```
hip_score_analysis(Prosthesis,Rating) :-
      Rating is a qualitative description assessing
      the hip score analysis record offered by Prosthesis to
      support the request for suitability. */
hip_score_analysis(Prosthesis,Rating) :-
hip_score_factors(Factors),
score(Factors,Prosthesis,0,Score),
calibrate(Score,Rating).

/* Hip Score Evaluation Rules */
calibrate(Score,bad) :- Score =< -500.
calibrate(Score,medium) :- -500 < Score, Score < 150.
calibrate(Score,good) :- 150 =< Score, Score < 1000.
calibrate(Score,excellent) :- Score >= 1000.

/* Arthroplasty Data - Weighting Factors */
hip_score_factors([(age,10),
(sex,2),
(cause,5),
(activity_level,5),
(weight,2)  ]).

score([(Factor,Weight)|Factors],Prosthesis,Acc,Score) :-
value(Factor,Prosthesis,Value),
Acc1 is Acc + Weight*Value,
```

```
score(Factors,Prosthesis,Acc1,Score).
score([],Prosthesis,Score,Score).

/* Final Evaluation */
evaluate(Profile,Outcome) :-
      Outcome is the reply to the prosthesis's Profile. */
evaluate(Profile,Answer) :-
      rule(Conditions,Answer), verify(Conditions,Profile).

verify([condition(Type,Test,Rating)|Conditions],Profile) :-
scale(Type,Scale),
select_value(Type,Profile,Fact),
compare(Test,Scale,Fact,Rating),
verify(Conditions,Profile).
verify([],Profile).

compare('=',Scale,Rating,Rating).
compare('>',Scale,Rating1,Rating2) :-
precedes(Scale,Rating1,Rating2).
compare('>=',Scale,Rating1,Rating2) :-
precedes(Scale,Rating1,Rating2) ; Rating1 = Rating2.
compare('<',Scale,Rating1,Rating2) :-
precedes(Scale,Rating2,Rating1).
compare('=<',Scale,Rating1,Rating2) :-
precedes(Scale,Rating2,Rating1) ; Rating1 = Rating2.
```

```
precedes([R1|Rs],R1,R2).
precedes([R|Rs],R1,R2) :- R \== R2, precedes(Rs,R1,R2).

select_value(fit,profile(C,F,Y),C).
select_value(success,profile(C,F,Y),F).
select_value(analysis,profile(C,F,Y),Y).

/* Utilities */
sumlist(Is,Sum) :-
            umlist(Is,0,Sum).
sumlist([I|Is],Temp,Sum) :-
      Temp1 is Temp + I,
      sumlist(Is,Temp1,Sum).
sumlist([],Sum,Sum).

/* Arthroplasty Data and Rules */
rule([condition(fit,'>=',excellent),condition(success,'>=',good),
      condition(analysis,'>=',reasonable)],give_suitability).
rule([condition(fit,'=',good),condition(success,'=',good),
      condition(analysis,'>=',reasonable)],consult_expert).
rule([condition(fit,'=<',moderate),condition(success,'=<',medium)],
      refuse_suitability).

scale(fit,[excellent,good,moderate]).
```

```
scale(success,[excellent,good,medium,bad]).

scale(analysis,[excellent,reasonable,poor]).


/* Prosthesis Data */

fea_analysis(charnley_medium,excellent).

requested_suitability(charnley_medium,40).


amount(isthmus,charnley_medium,5).

amount(femoral_head_offset,charnley_medium,4).

amount(canal_width,charnley_medium,2).

amount(canal_width_plus_20,charnley_medium,10).

amount(canal_width_minus_20,charnley_medium,10).


value(average_time_before_first_revision,charnley_medium,20).

value(average_time_before_second_revision,charnley_medium,10).

value(measure_of_patient_satisfaction,charnley_medium,49).

value(measure_of_bone_resorption,charnley_medium,9).


ok_profile(charnley_medium).
```

Appendix B

Theory of Fuzzy Expert Systems

Fuzzy set theory, proposed in 1965 by Lotfi Zadeh at the University of California, Berkeley, was an attempt to generalise classical set theory. Since then, the theory has been extended to other fields that are based on set theory, including logic. In classical (two-valued) logic, such as propositional and predicate logics, sentences take on one of two possible interpretations or meanings, the values true or false. This is in keeping with "crisp" set theory, wherein objects either belong to a set or do not belong to a set. There is no in-between (the excluded middle).

In classical set theory, a *set* is any well-defined collection of objects. The usual notation and definitions of sets are used. Sets with a number of objects (the elements or members of the set) can be listed. For example, the set of small numbers is listed $\{1, 2, 3, 4\}$. Sets with an infinite number of elements can be described without listing ($\Re = \{x | x \text{ is a real number }\}$, $\aleph = \{x | x$

is a positive integer or zero}, $Z = \{x | x \text{ is an integer}\}$. If x is an element of the subset A we write $x \in A$ and if x is not an element of A, we write $x \notin A$. Sets can also be defined by their properties: $\{x | P(x)\}$ are elements x satisfying property P. Equal sets A and B are denoted as $A = B$. A is a *subset* of B if every element of A is also an element of B (written as $A \subseteq B$, proper subsets as $A \subset B$ and $A \not\subset B$ when A is not a subset of B).

Sets can be characterised through an indicator or characteristic function. The characteristic function of f of subset A, is defined as

$$f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \tag{B.1}$$

Likewise, characteristic function values can be defined for derivative sets such as the intersection and union of two sets:

$$f_{A \cap B}(x) = f_A(x) f_B(x), \tag{B.2}$$

$$f_{A \cap B} = f_A + f_B - f_A f_B, \tag{B.3}$$

$$f_{A \oplus B} = f_A + f_B - 2 f_A f_B. \tag{B.4}$$

B.1 FUZZY SET THEORY

Fuzzy set theory is a generalisation of classical set theory. Definitions, theorems, proofs and results of classical set theory, in general, hold for fuzzy set theory. The theory of fuzzy logic is founded on fuzzy set theory in the same way that classical logics are formulated from (two-valued) set theory.

But, fuzzy logic representations try to capture the way humans represent and reason with real world knowledge. In real world situations, a multitude of inexact concepts must be dealt with: generalities (a concept applying to many things), ambiguities, vagueness, chance events, incomplete knowledge and even unbelievable or contradictory information. Representing such concepts with conventional set theory and logic is difficult if not impossible. For example, it becomes very cumbersome in classical logic to try to classify objects described by expressions such as "slightly more beautiful," "not quite as tall as," "much more expensive," "good, but not as useful," etc.

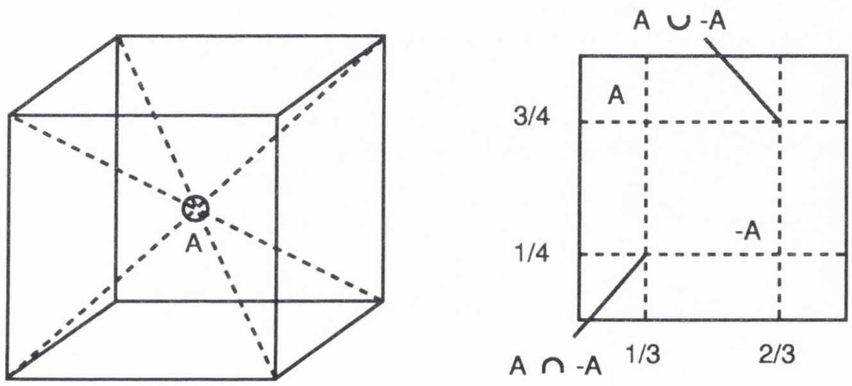
Fuzzy sets deal with subsets of the universe that have no well-defined boundaries. Members of a fuzzy set can have varying degrees of membership ranging between 0 and 1. For somebody who is 200cm in height, membership in the fuzzy subset "tall people" is near 1. Whereas somebody who's 100cm is more correctly nearer 0. But does a 150mm person belong to the set of tall people?

The characteristic function is used to formally define fuzzy sets similar to the way it was used to define crisp sets. Let X denote the universe of objects under consideration. Then the fuzzy subset A in X is a set of ordered pairs

$$A = \{(x, \mu_A(x))\}, x \in X \tag{B.5}$$

where $(0 \leq \mu_A(x) \leq 1)$ is the characteristic or membership function that denotes the degree of membership or inclusion of x in A. A value of $\mu = 0$ means that x is not included in A at all and a value of $\mu = 1$ signifies that x is a "full" member of A, corresponding to crisp membership values. Values of μ between 0 and 1 are the relative degrees of set inclusion ranging between complete inclusion and none.

A geometrical way to view a fuzzy set is as a point in the unit cube I^n. Vertices of the cube are non-fuzzy points. Maximal fuzziness then occurs at the midpoint of the unit cube (0.5, 0.5, ... , 0.5) (Figure B.1a).



(a) Maximum fuzziness. (b) $A \cap -A$.

Figure B.1: "Points as sets" geometrical interpretation of a fuzzy set.

Referring to Figure B.1, we can see how intersection and union of the fuzzy set A and its complement set $-A$ are interpreted. For example in Figure B.1b if the fuzzy set A is defined as $A = (\frac{1}{3}, \frac{3}{4})$, then the complement set is $-A = (\frac{2}{3}, \frac{1}{4})$ and intersection and union are then given by $A \cap -A = (\frac{1}{3}, \frac{1}{4})$ and $A \cup -A = (\frac{2}{3}, \frac{3}{4})$ respectively.

Operations on fuzzy sets such as intersection, union, complementation are somewhat similar to operations on crisp sets ("$-$" is used to denote the complement set). The following relationships are also defined for fuzzy sets:

$$A = B \quad \text{iff} \quad \mu_A(x) = \mu_B(x), \tag{B.6}$$

$$A \subseteq B \quad \text{iff} \quad \mu_A(x) \leq \mu_B(x), \tag{B.7}$$

$$A \cup B : \mu_{A \cup B}(x) \equiv \mu_A(x) \vee \mu_B(x), \tag{B.8}$$

$$A \cap B : \mu_{A \cap B}(x) \equiv \mu_A(x) \wedge \mu_B(x). \tag{B.9}$$

Here $\vee$ is the symbol for maximum, so $\mu_{A \cup B}$ is the smallest fuzzy subset having both A and B as subsets. In a similar way, $\wedge$ is the symbol for minimum, so $\mu_{A \cap B}$ is the largest fuzzy subset that is a subset of both A and B. The following relationship also holds

$$-A : \mu_{-A}(x) \equiv 1 - \mu_A(x). \tag{B.10}$$

In general

$$A \cap (-A) \neq \varnothing \tag{B.11}$$

and

$$A \cup (-A) \neq X \tag{B.12}$$

since for $\mu_A(x) = c$, with $0 < c < 1$,

$$\mu_{A\cup -A}(x) = \max(c, 1-c) \neq 1, \tag{B.13}$$

$$\mu_{A\cap -A}(X) = \min(c, 1-c) \neq 0. \tag{B.14}$$

An example of a fuzzy set is the fuzzy subset A of "small integers." If X is the set of all nonnegative integers, we might define A by

$$\mu_A(x) = \frac{1}{1+(\frac{x}{4})^2} \tag{B.15}$$

where $x = (1, 2, \dots)$. In the above example it was suggested that "we might define ... " since membership function definitions are subjective. They are a matter of personal choice. Of course, one could seek a consensus among a group of people as to how membership functions are defined and use the means of the consensus for the definitions.

As another example, let X be the set of integers in the interval $[0, 120]$ and x interpreted as "age." We might then define the fuzzy subset A as "old" with membership values as depicted in Figure B.2. Note that the ***linguistic*** variable AGE can take words as values (VERY YOUNG, YOUNG, ... , OLD, VERY OLD) and these linguistic variables each have fuzzy membership functions as suggested in Figure B.2.

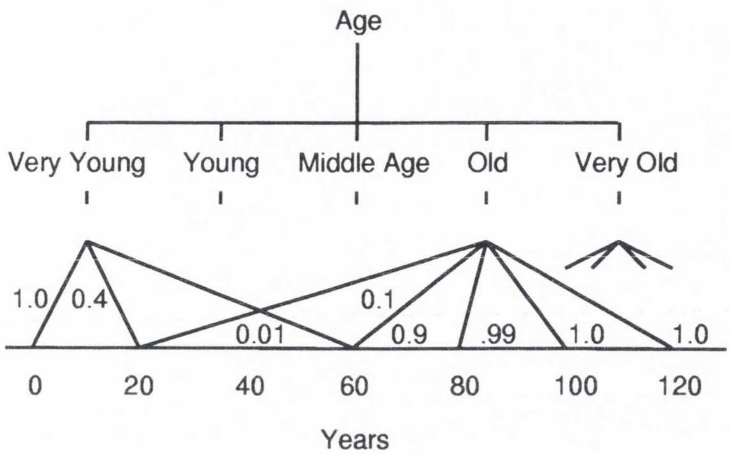


Figure B.2: The fuzzy membership function for OLD.

Some special operations defined for fuzzy sets including the following:

$$\text{Dilation:} \qquad \text{Dil}(A) = [\mu_A(x)]^{\frac{1}{2}}, \tag{B.16}$$

$$\text{Concentration:} \qquad \text{Con}(A) = [\mu_A(x)]^2, \tag{B.17}$$

$$\text{Normalization:} \qquad \text{Norm}(A) = \frac{\mu_A x}{\max_x\{\mu_A(x)\}}. \tag{B.18}$$

Other examples of fuzzy measures include cardinality and entropy. The cardinality M or "size" of a fuzzy set A is defined as

$$M(A) = \sum_{i=1}^{n} \mu_A(x_i). \tag{B.19}$$

For example, the cardinality of the fuzzy set of Figure B.1b is just

$$M(A) = \sum_{i=1}^{2} \mu_A(x_i) = \frac{1}{3} + \frac{3}{4} = \frac{13}{12}.$$

The entropy of a fuzzy set A is defined as

$$E(A) = \frac{d(A_{\text{near}})}{d(A_{\text{far}})} \tag{B.20}$$

where d is (Euclidean) distance, and A_{near} and A_{far} are line segments from the point A to the nearest and farthest vertex in I^n, respectively. Again referring to Figure B.1b the fuzzy entropy E of the fuzzy set A in the figure is just

$$E(A) = \frac{d(A_{\text{near}})}{d(A_{\text{far}})} = \frac{\frac{1}{3}+\frac{1}{4}}{\frac{2}{3}+\frac{3}{4}} = \frac{7}{17} \quad (d \text{ is distance}). \tag{B.21}$$

Just a few examples of the many operations, definitions and theory related to fuzzy sets have been given here.

B.2 FUZZY LOGIC

Predicate logic is a two-valued logic, based on traditional set theory. Predicates define classes of objects, and objects that satisfy a given predicate are members of the respective class. Inferences in predicate logic are performed using inferring rules such as *modus ponens.* If P and Q are predicated and $\rightarrow$ is the implication connective (read as IF ... THEN), then modus ponens can be summarized as

$$\frac{P, P \rightarrow Q}{Q}. \tag{B.22}$$

The formulas above the line are called the *premises*, and the formula below the line is called the *conclusion* of the inference rule. Similar inferring rules have been defined for fuzzy logic based on fuzzy relations among fuzzy subsets. In fuzzy logic, modus ponens is concerned with the degree of truth between the premise and the consequent,

$$\frac{\mu_P(x), \mu_P(x) \to \mu_Q(y)}{\mu_Q(y)}. \tag{B.23}$$

Membership in P implies membership in Q. We know P's degree of membership (truth). Therefore, we can infer Q's membership (truth).

Appendix C

Fuzzy Expert System Example

The rule-based expert systems use "if ... then" rules to represent "chunks" of knowledge. Similar expert systems have been implemented with fuzzy "if ... then" rules. Fuzzy expert systems are called fuzzy expert systems. They are often able to model common-sense reasoning better than conventional rule-based expert systems. The basic fuzzy expert system operates in three stages: converting input variable values to fuzzy set values, rule instantiation and converting from fuzzy set values back to "crisp" output variable values. The process is illustrated in Figure C.1. The method of operation of the fuzzy

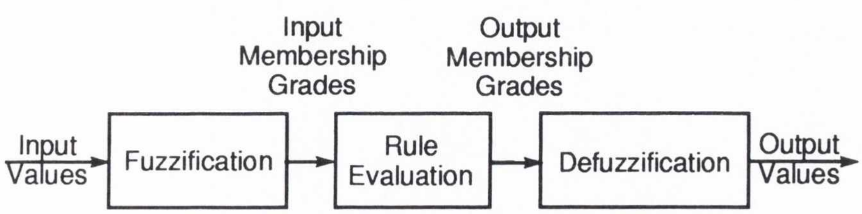


Figure C.1: Fuzzy inference in a fuzzy rule expert system.

expert system is best described by an example. The application chosen is a fuzzy decision support system for the trading of stocks. The member ship functions defined for the *share price* level of Company X are illustrated in Figure C.2 where the triangular shaped fuzzy membership values of negative

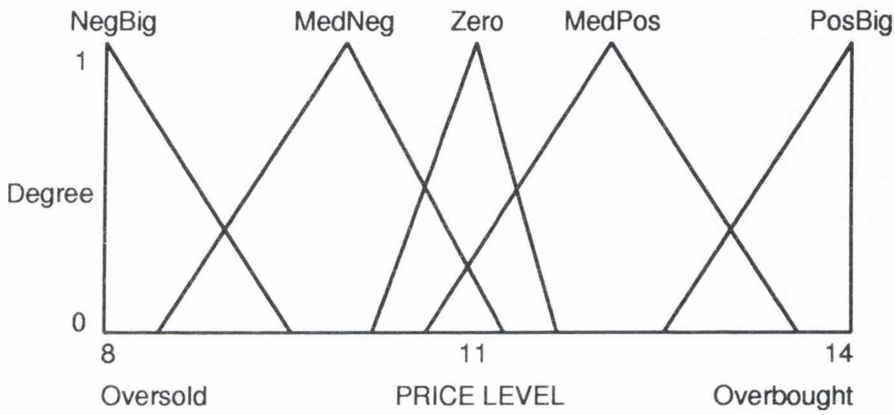


Figure C.2: Membership functions for share price level in Company X.

big (NegBig), medium negative (MedNeg), zero, medium positive (MedPos) and positive big (PosBig) are depicted.

Fuzzification of price level is accomplished by mapping from price value to membership function value. For example, if the price of Company X's shares is \$9.00 the input grades are given by Table C.1. The mapping is

Label	Grade (Value)
PosBig	0
MedPos	0
Zero	0
MedNeg	0.6
BigNeg	0.2

Table C.1: Input grades.

illustrated in Figure C.3.

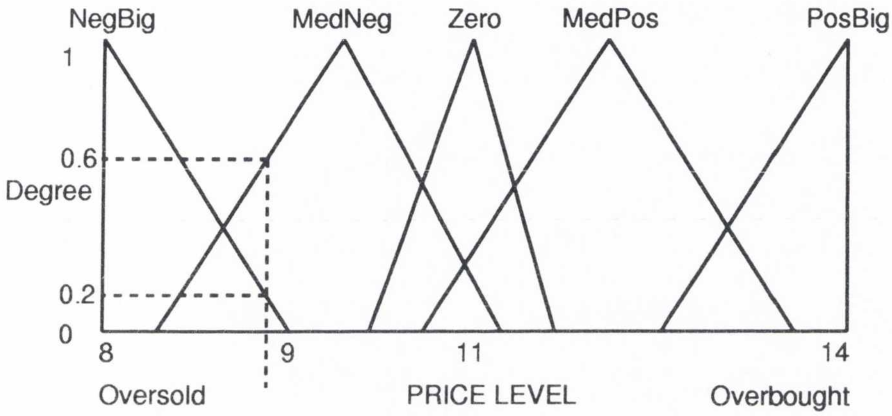


Figure C.3: The fuzzification mapping process.

Another input variable is share position, the number of shares held (long) or owed (short) in inventory by Company X. The membership functions for share position are illustrated in Figure C.4.

The next stage of processing is rule evaluation. The fuzzy rules used have two conjuncts (if conditions) in the rule premise corresponding to the two fuzzy input variables and one output variable (action). The rules have the following form:

```
[R1]  If price_level is MedPos and share_position is PosBig
      then position_change is MedNeg;
[R2]  If price_level is NegMed and share_position is Zero
      then position_change is PosBig.
```

A complete decision table for the rules is illustrated in Table C.2 where the row and column entries are the rule conjuncts and the table value is the

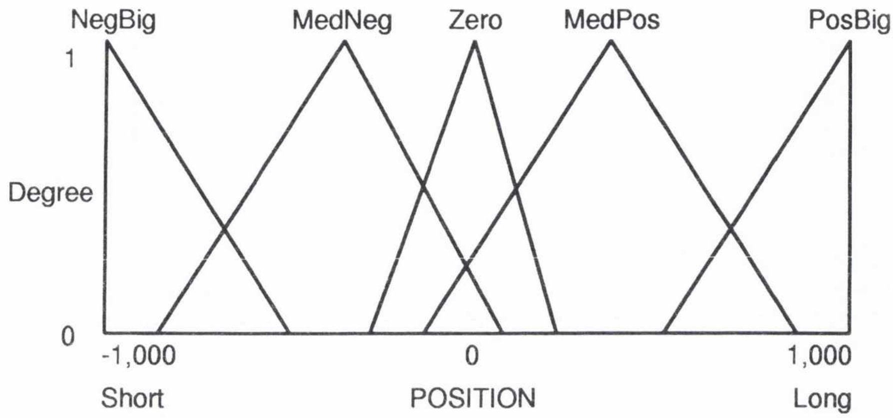


Figure C.4: Membership functions for share position in Company X.

	Price Level				
Position	NB	NM	Z	PM	PB
PB	PB	PB	Z	NM	NB
PM	PB	PB	PM	NM	NM
Z	PB	PB	PM	Z	Z
NM	PM	PM	Z	Z	PM
NB	PB	PB	PM	Z	Z

Table C.2: Fuzzy decision table for share trading decision support system.

corresponding rule action.

The final stage of processing is defuzzification, mapping from the "then" part of the rule membership function to variable values—the recommended action of the system.

A portion of the complete fuzzy expert system is illustrated graphically in Figure C.5. where the two input variables are share price and share position and the output variable is the approve/refuse recommendation. Fuzzification takes place in the first stage on the left where the fuzzy values are combined

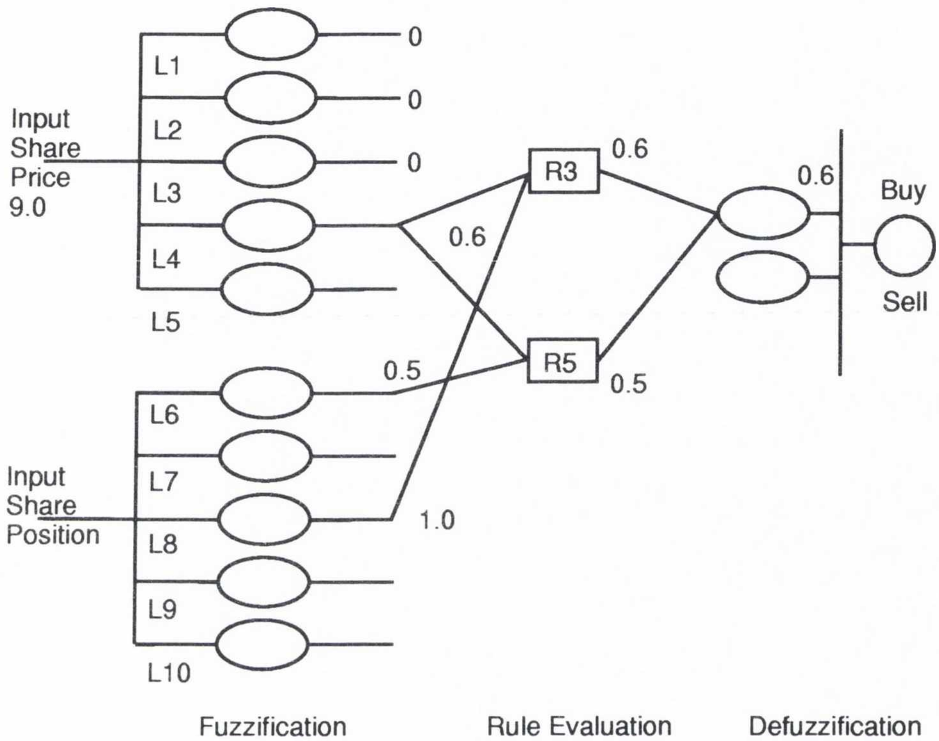


Figure C.5: A portion of a fuzzy expert system.

and passed to the rules R3 and R5 in parallel. The action parts of the rules are then combined and defuzzified to give the recommended action to buy or sell shares in Company X.

Appendix D

Prolog Fuzzy Expert System Source Code

This appendix contains the Prolog Fuzzy Expert System source code listings.

```
/* RISK.PRO:

Developed by Alberto Pacheco, 1996-1997

Modified by John Cogan, 1997-1998

This version supports:

One-goal, one-sample-at-a-time

Linear Fuzzy Membership Representations Sigmoid Curve

      Membership Function (S-Curve Fuzzy Set) (as/ds)

Zadeh Fuzzy Set Operators

Mapping from Membreship Degree to

      I/O Domain (member_to_domain)

Scalable Monotonic Chaining (monotonic_scaling) */
```

```
/* Operator Definitions */
:- op(700, xfx, is).
:- op(720, xfy, and).
:- op(740, xfy, or).
:- op(760, xfx, then).
:- op(780, fx,  if).


/* Main Procedure:
      1) Initialization;
      2) Goal with Output Variable */
main :- init(combination_risk), one_goal(combination_risk).


/* Initialization: Clear global working memory */
init(Var) :-
      retractall(_),
      assert(sum1(Var,0.0)),
      nl,write('Fuzzy Patient/Prosthesis Combination
            Risk Assessment Model'), nl,
      input_value( _, Input_Variable, Value ),
      write(Input_Variable),write(': '),
      write(Value),nl, fail.
init(_):-nl,!.


/* Probes all rules with conclusion 'Var' */
one_goal(Var) :- prove( Var is X ), fail.
```

```
/* Reports the composite solution */
one_goal(Var) :- output_value(Var,X),
      nl, write(Var), write(' = '), write(X), nl.

/* Production Rules (Fuzzy Model's Rules)
A Patient/Prosthesis Combination Risk Assessment Model
Based on Earl Cox's Case Study */
if fit_and_fill_level is loose
      then operation_risk is increased.
if fea_stress_level is adverse
      then operation_risk is increased.
if survival_analysis_level is short
      then operation_risk is  increased.
if hip_score_level is high
      then operation_risk is increased.

/* Fuzzy Sets Definitions
Note: The Maximum Domain Limit acts as a
Weighted Measure for its corresponding set. */
fuzzy_set(fit_and_fill_level, loose, as,
      0.0, 1.5, 3.0, 0.0).
fuzzy_set(fea_stress_level, adverse, at,
      0.0, 24.0, 0.0, 0.0).
fuzzy_set(survival_analysis_level, short, dt,
```

```
      0.0, 100.0, 0.0, 0.0). /* Greatest Weight */
fuzzy_set(hip_score_level, high, dt,
      0.0, 10.0, 0.0, 0.0 ). /* Lowest Weight */

/* Output Fuzzy Sets */
fuzzy_set(combination_risk, increased, as,
      0.0, 1000.0, 0.0, 0.0).
fuzzy_set(combination_risk, high_risk, at,
      0.0, 5000.0, 0.0, 0.0).

/* Input Values */
input_value( 1, fit_and_fill_level, 0.8000000).
input_value( 1, fea_stress_level, 12.000000).
input_value( 1, survival_analysis_level, 40.000000).
input_value( 1, hip_score_level, 3.000000).

/* Prolog Interpreter in Prolog
Based on Dennis Merrit's Article
"Building Custom Rule Engines," PC AI, Mar/Apr 1996. */
prove(ATTR is VALUE and REST) :-
      getav(ATTR, VALUE),
      prove(REST),
      apply_fuzzy_oper(and_z).
prove(ATTR is VALUE or REST) :-
      getav(ATTR, VALUE),
```

```
        prove(REST),
        apply_fuzzy_oper(or_z).
prove(ATTR is VALUE) :-
        getav(ATTR,VALUE).
getav(ATTR,VALUE) :-
        if CONDITIONS then ATTR is VALUE,
        prove(CONDITIONS),
        retract(prem(Mx)),
        monotonic_scaling(ATTR,VALUE,Mx).
getav(ATTR,VALUE) :-
        not(if _ then ATTR is _),
        rule_translation(ATTR,VALUE).

/* Fuzzy Rule Processing */
rule_translation( T, Cj ) :-
        clause( fuzzy_set(T,_,_,_,_,_,_), _ ), !,
        input_value( _, T, X ), !,
        fuzzification( T, Cj, X ).

/* Discrete Inference Rule Processing */
rule_translation( T, Cj ) :-
        input_value( _, T, Cj ), !, is_true.
rule_translation( T, Cj ) :-
        is_false, nl,write('Error in rule_translation():
                Undefined set'),nl,write(T),nl,write(Cj).
```

```
/* Fuzzification */

fuzzification( Name, Set, X_Value ) :-
      fuzzy_set( Name, Set, Type, A, B, C, D ),
      degree_of_membership( Type, A, B, C, D,
            X_Value, Membership ),
      assert( prem(Membership) ), !.

/* dt - Linear Decreasing Fuzzy Set (i,i,i,i,i,i,o)
      A - Minimum Value
      B - Maximum Value */

degree_of_membership( dt, A, _, _, _, X, 1.0 ) :- X =< A, !.
degree_of_membership( dt, _, B, _, _, X, 0.0 ) :- X >= B, !.
degree_of_membership( dt, A, B, _, _, X, M ) :-
      line_eq( dt, A, B, X, M ), !.

/* at - Linear Increasing Fuzzy Set
      A - Minimum Value
      B - Maximum Value */

degree_of_membership( at, A, _, _, _, X, 0.0 ) :- X =< A, !.
degree_of_membership( at, _, B, _, _, X, 1.0 ) :- X >= B, !.
degree_of_membership( at, A, B, _, _, X, M ) :-
      line_eq( at, A, B, X, M ), !.

/* tp - Trapezoidal or Triangular Fuzzy Set
```

```
         -- TRAPEZOIDAL --          -- TRIANGULAR --

        A - Minimum Value          A - Minimum Value

        B - Left Shoulder          B - Center

        C - Right Shoulder         C - Center

        D - Maximum Value          D - Maximum Value */

degree_of_membership( tp, A, _, _, _, X, 0.0 ) :- X =< A, !.

degree_of_membership( tp, A, B, _, _, X, M ) :-

      X > A, X =< B, line_eq( at, A, B, X, M ), !.

degree_of_membership( tp, _, B, C, _, X, 1.0 ) :-

      X > B, X < C, !.

degree_of_membership( tp, _, _, C, D, X, M ) :-

      X > C, X < D, line_eq( dt, C, D, X, M ), !.

degree_of_membership( tp, _, _, _, _, _, 0.0 ).


/* as - Increasing S-Curve Fuzzy Set

      A - Minimum Value

      B - Inflexion Point

      C - Maximum Value */

degree_of_membership( as, A, _, _, _, X, 0.0 ) :- X =< A, !.

degree_of_membership( as, A, B, C, _, X, M ) :-

      X > A, X =< B, M is (2.0 * ((X-A)/(C-A))**2.0), !.

degree_of_membership( as, A, B, C, _, X, M ) :-

      X > B, X < C, M is (1.0 - 2.0*(((C-X)/(C-A))**2.0)), !.

degree_of_membership( as, _, _, _, _, _, 1.0 ).
```

```
/* Finds the Expected Output Value of the Fuzzy Set
Given its Degree of Memembership
member_to_domain( A, B,
      Membership-Degree, X-Domain-Value ) - (i,i,i,o) */
member_to_domain( X1, X2, Memb, X ) :-
      X is ( Memb * (X2 - X1) + X1 ), !.

/* Fuzzy Operators */
apply_fuzzy_oper( and_z ) :-
      retract(prem(M1)),
      write(M1),
      retract(prem(M2)), !,
      write(' and '), write(M2), nl,
      min(M1,M2,M),
      assert(prem(M)), !.
apply_fuzzy_oper( or_z ) :-
      retract(prem(M1)),
      write(M1),
      retract(prem(M2)),!,
      write(' or '), write(M2), nl,
      max(M1,M2,M),
      assert(prem(M)), !.

/* Monotonic Scaling Chaining Model - (i,i,i) */
monotonic_scaling( Var, OutSet, Memb ) :-
```

```
        fuzzy_set( Var, OutSet, _, X1, X2, _, _ ), !,
        member_to_domain( X1, X2, Memb, X ),
        retract( sum1(Var,Q) ),
        R is ( X + Q ),
        assert( sum1(Var,R) ),
        write(Var),write(' :  amount = '),write(X),
        write('  membership = '),write(Memb),nl,!.

/* End of Defuzzification Method */
output_value( Var, X ) :-
        retract( sum1( Var, cummulative_risk ) ),
        nl, write('Cummulative Risk is '),
        write(cummulative_risk),nl,
        fuzzification( Var, high_risk, cummulative_risk ),
        retract( prem( Memb ) ),
        X is (Memb * 1000.0).

/* Linear Interpolation */
line_eq( dt, X1, X2, X, Y ) :-
        Y is (X2 - X) / (X2 - X1).
line_eq( at, X1, X2, X, Y ) :-
        Y is (X - X1) / (X2 - X1).

/* Fuzzy Function Primitives */
min( X, Y, X ) :- X < Y, !.
```

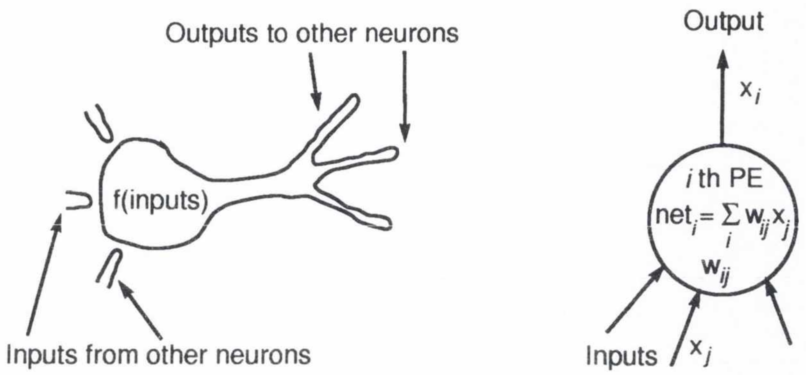
```
min( _, Y, Y ).

max( X, Y, X ) :- X > Y, !.

max( _, Y, Y ).


/* Boolean Values */

is_true :- assert(prem(1.0)), !.

is_false :- assert(prem(0.0)), !.
```

Appendix E

Theory of Feed-Forward Neural Networks

E.1 THE GENERAL NEURAL NETWORK PROCESSING ELEMENT

Recent advances in the construction of computer based artificial neural networks (ANNs) derives from enhanced knowledge of the biological neuron of the human brain [43, 42]. Figure E.1a is a simplified representation of a single neuron cell capable of a crude computation. The brain organises such cells into networks capable of incredible feats of learning. ANNs are an attempt to mimic the brain's learning processes. Figure E.1b is an artificial neuron which is a computer representation of the biological neuron. Artificial neurons, also called units, nodes or processing elements (PEs), form the basic units of artificial neural networks.



(a) Biological neuron.

(b) Artificial neuron.

Figure E.1: Neurons.

Generally speaking, output values from units will be positive numbers. Weights w_{ij} can be either positive or negative (see Figure E.1b). Inputs to a unit are categorised according to their effect. Inputs whose connections have a positive weight contribute a net positive value to the overall excitation of the unit. Those inputs whose connections have negative weights detract from the overall excitation. The former type are referred to as **excitatory connections**, and the latter as **inhibitory connections**.

The overall excitation of a unit is its **net input**. The value of the excitation is usually calculated by summing the products of the input values and the weights assigned to the associated connections. For the ith unit, the net input is:

$$\mathrm{net}_i = \sum_{j=1}^{n} x_j w_{ij}. \tag{E.1}$$

The output value is written as

$$x_i = f_i(net_i). \tag{E.2}$$

The output value of a unit can be expressed in the form of a differential equation. Similar to their biological counterparts, the outputs are dynamic functions of time. The simplest form of the equation of importance for the outputs is

$$\dot{x}_i = -x_i + f(\text{net}_i) \tag{E.3}$$

where $f(\text{net}_i)$ is the function referred to as the **output function**. We apply arbitrary input values to the PE so that $net_i > 0$. If the inputs are applied for a sufficiently long time, the output value will reach an equilibrium value, when $\dot{x}_i = 0$, given by

$$x_i = f_i(net_i) \tag{E.4}$$

which is the same as Equation E.2.

The function $f(\text{net}_i)$ can take many different forms, the simplest being the identity function; that is, $f(\text{net}_i) = \text{net}_i$. Equation E.3 can then be written as

$$\dot{x}_i = -x_i + \text{net}_i. \tag{E.5}$$

Integrating this differential equation in order to study the behaviour of the variable x as a function of time gives

$$x_i(t) = \text{net}_i(1 - e^{-t}). \tag{E.6}$$

From Equation E.6, the value of x asymptotically approaches the value of net_i.

The most useful output function is called a **sigmoid**. One version of a sigmoid function is defined by the equation

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{E.7}$$

One of the important features of the sigmoid function is that it is differentiable everywhere. If the output function of a unit is a sigmoid function, then the following relationship holds

$$f'(net_i) = f(net_i)(1 - f(net_i)).$$

If we define $o_i = f(net_i)$, then we can write

$$f'(net_i) = o_i(1 - o_i). \tag{E.8}$$

This equation will be useful later.

The **threshold function** defined below is another useful function

$$f(x) = \begin{cases} 1 & x \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{E.9}$$

where θ is the threshold value.

E.2 LEARNING IN NEURAL NETWORKS

Learning or training in neural networks involves finding a set of weights such that the network correctly performs the intended data-processing function. For some simple cases weights can be arrived at by a trial-and-error procedure. Consider, for example, constructing a single-unit system that computes the AND function of two binary inputs (see Figure E.2 and Table E.1). The

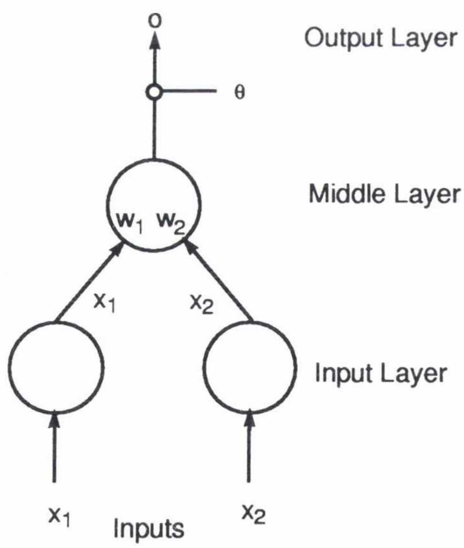


Figure E.2: Simple network with two input units and a single output unit.[1]

unit in Figure E.2 has two weights and a threshold output function, as in

[1] We wish to find weights such that, given any two binary inputs, the network correctly computes the AND function for those two inputs.

Input #1	Input #2	Output
0	0	1
1	0	0
0	1	1
1	1	1

Table E.1: The AND truth table.

Equation E.9. The threshold condition can be rewritten as follows. The network will have an output of one if

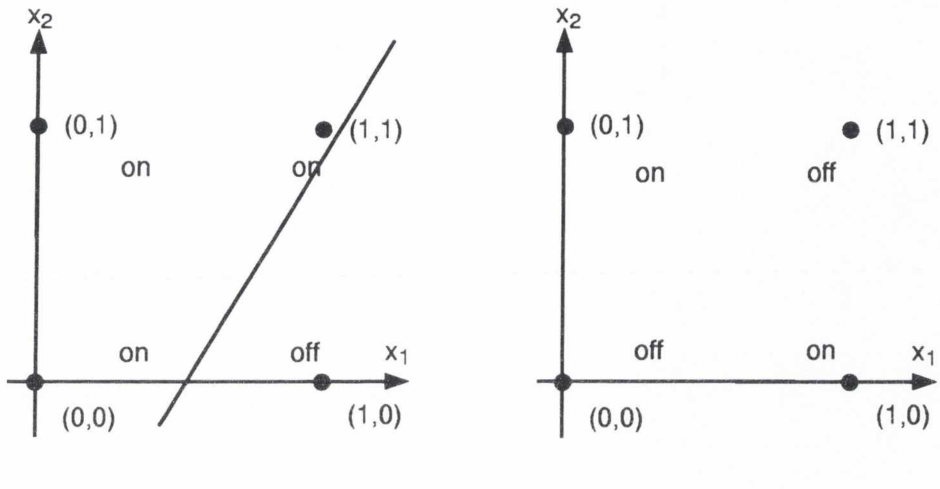
$$\sum_{i=1}^{n} w_i x_i - \theta > 0 \tag{E.10}$$

and otherwise, the output will be zero. n refers to the number of inputs.

If we replace the inequality in Equation E.10 with an equality, the equation becomes the equation of a line in the $x_1 x_2$ plane. If we position that line properly, we can determine the weights that will allow the network to solve the AND problem (see Figure E.3a). There are an infinite number of other lines that also yield weights that solve this problem.

The line in the problem just discussed is an example of a **decision surface**. The line breaks up the space into two regions, one where the points in the region, when used as inputs, would satisfy the threshold condition, and one where the points in the region would not satisfy the threshold condition.

The device in Figure E.2 is often referred to as a **perceptron**. Perceptrons were an early development in neural networks, dating from the late 1950s. They were invented by a psychologist named Frank Rosenblatt, who was primarily concerned with collections of the above devices which he called perceptrons rather than with an individual unit. Rosenblatt favoured ran-



(a) Linearly separable. (b) Linearly inseparable.

Figure E.3: Linearly separable sets and inseparable sets.

dom connectivity among layers of these devices in his models of perception and vision.

E.2.1 Weakness of Perceptron Model

The individual unit, however, has a serious flaw that limits its use. Rosenblatt's former schoolmate, Marvin Minsky, along with Seymour Papert, in their book *Perceptrons* [88] expounded in great detail on the weaknesses of the perceptron model.

In order to explore this weakness we shall consider once again a single unit, as in Figure E.2. This time, however, we wish to solve the XOR problem. Figure E.3b and Table E.2 show the space of input points for this problem.

There is no orientation of the decision surface (line) that will correctly

Input #1	Input #2	Output
0	0	0
1	0	1
0	1	1
1	1	0

Table E.2: The XOR truth table.

separate the points having an output of zero from the points having an output of one. The linear decision surface is a characteristic of the perceptron unit. We say that the perceptron unit can only separate categories, or classes, if they are naturally **linearly separable**. This characteristic is considered by many to be a serious weakness, since many real problems, of which the XOR is a very simple example, do not have classes that are linearly separable.

David Rumelhart, Geoffrey Hinton, and Ronald Williams in 1986 published a seminal work, "Learning Internal Representations by Error Propagation" that enabled research into artificial neural networks to make a significant leap forward. They developed a *multi-layer feed-forward network*, that was not restricted to linearly separable training sets, and they complemented this with a reasonably effective training algorithm (to determine weights for it). They thereby demonstrated that artificial neural networks could provide real solutions to practical problems.

E.3 MULTI-LAYER FEED-FORWARD NETWORK

Although the problem of linear separability appeared formidable, it was relatively easily overcome. A solution to the XOR problem was constructed by using a network of the type shown in Figure E.4. In this case a hidden

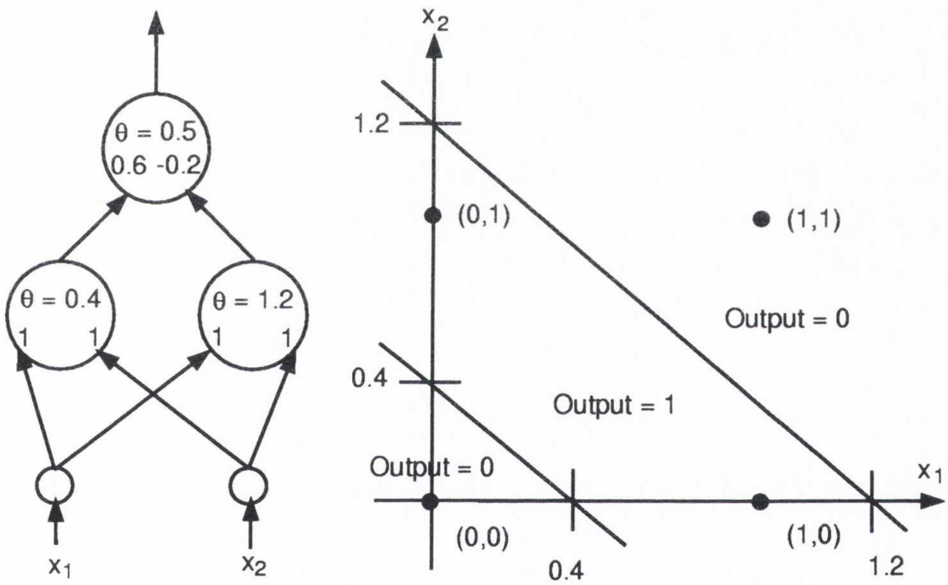


Figure E.4: Network made up of three layers: an input layer, a hidden layer comprising two units, and a single unit output layer.[2]

layer of units was constructed between the input and output layers. It is this hidden layer that facilitates a solution. Each hidden-layer unit produces a decision surface, as shown in the figure. The first hidden-layer unit (the one

[2]All units are threshold units with the value of θ as the threshold in each case. The two hidden-layer units construct the two decision surfaces shown in the graph on the right. The output unit performs the logical function: (hidden unit 1) AND (NOT (hidden unit 2)).

on the left) will produce an output of one if either or both inputs are one. The hidden-layer unit on the right will produce an output of one only if both input are one.

The output unit will produce an output of one only if the output of the first hidden unit is one AND the output of the second hidden unit is zero; in other words, if only one, but not both, of the inputs are one.

This is a primitive form of multi-layered feed-forward network. The term "feed-forward" means that information flows in one direction only. The inputs to neurons in each layer come exclusively from the outputs of neurons in previous layers, and outputs from these neurons pass exclusively to neurons in following layers.

E.3.1 Training by Error Minimisation

So far it has been possible to arrive at the weights for the network by trial-and-error, as in Figure E.4, which enabled the network to behave as an XOR function. For more complex functions, where there may be many inputs, hidden-layer units, and outputs, this is not feasible. A method must be introduced whereby the network can be trained to learn its own weights.

A method with this capability has its roots in a search technique called **hill climbing**. In the Section E.3.1.1 the training of a system comprising a single unit is examined. In Section E.3.2 we extend this method to cover the case of multiple units and multiple layers of interconnected units.

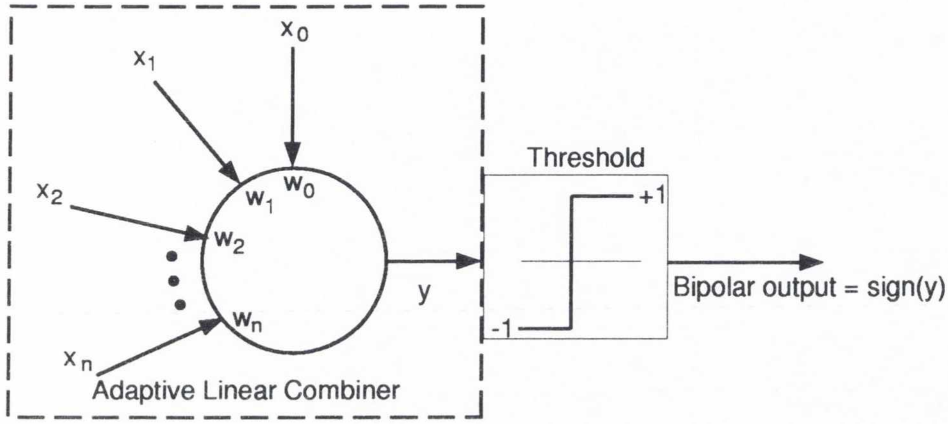


Figure E.5: Adaline.[3]

E.3.1.1 Adaline and the Adaptive Linear Combiner

The Adaline comprises two major parts, as illustrated in Figure E.5: an **adaptive linear combiner** (ALC), a unit almost identical in structure to the general processing element described in Section E.1, and a bipolar output function, which determines its output based on the sign of the net-input value of the ALC. **Adaline** is an acronym for ADAptive Linear Neurone.

Notice the addition of a connection with weight, w_0, which is referred to as the **bias** term. This term is a weight on a connection that has its input value always equal to one. The inclusion of such a term is largely a matter of experience.

The net input to the ALC is calculated as before as the sum of the products of the inputs and the weights. In the case of the ALC, the output function is the identity function, so the output is the same as the net input.

[3]The complete Adaline consists of the adaptive linear combiner, in the dashed box, and a bipolar output function. The adaptive linear combiner resembles the general processing element described in Section E.1.

If the output is y, then

$$y = w_0 + \sum_{i=1}^{n} w_i x_i$$

where the w_i are the weights and x_i are the inputs. If we make the identification $x_0 = 1$, then we can write

$$y = \sum_{i=1}^{n} w_i x_i$$

or in terms of the vector dot product

$$y = \mathbf{w} \cdot \mathbf{x}. \tag{E.11}$$

The final output of the Adaline is

$$o = Sign(y) \tag{E.12}$$

where the value of the Sign function is +1, 0, -1, depending on whether the value of y is positive, zero, or negative.

E.3.1.2 The LMS Learning Rule

Suppose we have an ALC with four inputs and a bias term. Furthermore, suppose that we desire that the output of the ALC be the value 2.0, when the input vector is {1, 0.4, 1.2, 0.5, 1.1} where the first value is the input to the bias term. We can represent the weight vector as $\{w_0, w_1, w_2, w_3, w_4\}$. There is an infinite number of weight vectors that will solve this particular problem.

However, suppose we have a set of input vectors, $\{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_L}\}$, each having its own, perhaps unique, *correct* or *desired* output value, $d_k, k = 1, L$. The problem of finding a single weight vector that can successfully associate each input vector with its desired output value is no longer simple. In this section we develop a method called the **least-mean-square** (LMS) learning rule, or the **delta rule**, which is one method of finding the desired weight vector. We refer to this process of finding the weight vector as *training* the ALC. Moreover, we call the process a **supervised learning** technique, in the sense that there is some external teacher that knows what the correct response should be for each given input vector. The learning rule can be embedded in the device itself, which can then *self-adapt* as inputs and desired outputs are presented to it. Small adjustments are made to the weight values as each input-output combination is processed until the ALC give correct outputs. In a sense, this procedure is a true training procedure, because we do not calculate the value of the weight vector explicitly.

Weight Vector Calculations Before developing the LMS rule, some insight can be gained into the procedure by looking at a method for calculating the weight vector. To begin, the problem can be restated: given examples, (also called **exemplars**), $(\mathbf{x_1}, d_1), (\mathbf{x_2}, d_2), \ldots, (\mathbf{x_L}, d_L)$, of some processing function that associates (or *maps*) input vectors, $\mathbf{x_k}$, with output values, d_k, what is the *best* weight vector, $\mathbf{w_{min}}$, for an ALC that performs this mapping? We shall assume $L > n+1$, where n is the number of inputs and there is one additional weight for the bias term. This assumption means that we cannot find the weight vector by solving a system of simultaneous equations because such a system is over-determined.

The answer to the question posed in the previous paragraph depends on how we define the word *best* within the context of the problem. Once we find this best weight vector, we would like the application of each input vector to result in the precise, corresponding output value. Since it may not be possible to find a set of weights that allows this mapping to be performed without error, we would like at least to minimise the error. Thus, we choose to look for a set of weights that minimises the mean-squared error over the entire set of input vectors. If the actual output value for the kth input vector is y_k, then we define the error as

$$\varepsilon_k = d_k - y_k \tag{E.13}$$

and the mean-squared error is

$$\xi = \langle \varepsilon_k^2 \rangle = \frac{1}{L} \sum_{k=1}^{L} \varepsilon_k^2 \tag{E.14}$$

where the angled brackets indicate the mean, or expectation, value.

Substituting Equations E.11 and E.13 into Equation E.14 shows that the mean-squared error is an explicit function of the weight values

$$\xi = \langle d_k - \mathbf{w} \cdot \mathbf{x_k} \rangle^2. \tag{E.15}$$

Expanding this equation we find

$$\xi = \langle d_k^2 \rangle + \mathbf{w}^t \langle \mathbf{x_k x_k^t} \rangle \mathbf{w} - 2\langle d_k \mathbf{x_k^t} \rangle \mathbf{w}. \tag{E.16}$$

The fact that ξ is a function of the weights means that it is possible to find weights that minimise ξ. The function $\xi(\mathbf{w})$ is plotted for the case of an ALC with only two inputs and no bias term. Using the following definitions

$$d = \langle d_k^2 \rangle,\ \mathbf{R} = \langle \mathbf{x_k x_k^t} \rangle \text{ and } \mathbf{p} = \langle d_k \mathbf{x_k^t} \rangle$$

and without specifying the actual input vectors, we can construct the graph. The surface of the function $\xi(\mathbf{w})$ (see Figure E.6) is a paraboloid. The

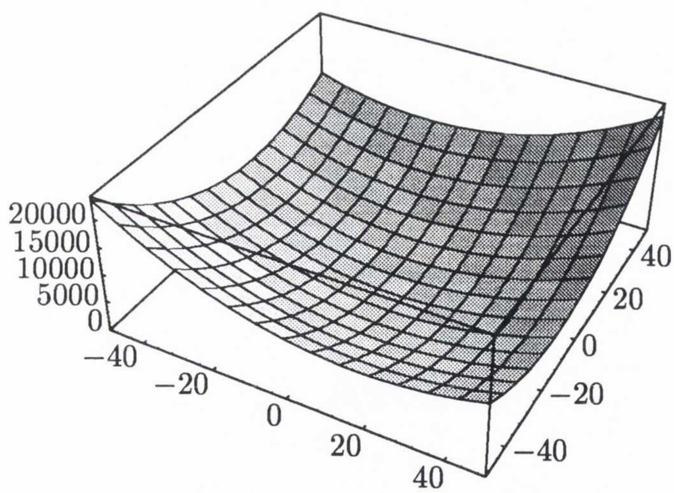


Figure E.6: A graph of the function $\xi(\mathbf{w})$.

function has a single minimum point. The weights corresponding to that minimum point are the best weights for this example. A contour plot of this function is shown in Figure E.7. We can find the minimum point by taking

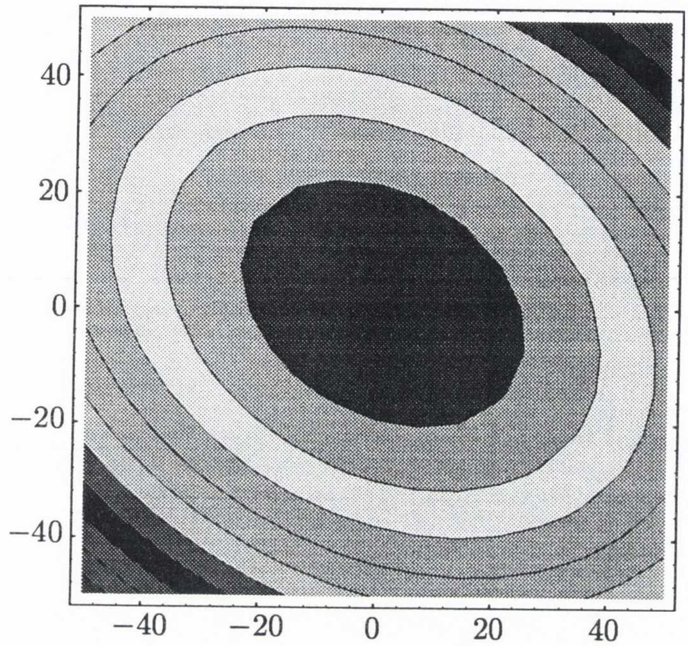


Figure E.7: A contour plot of the function $\xi(\mathbf{w})$.

the derivative of Equation E.16. The result is the weight vector that gives the minimum error

$$\mathbf{w}_{\mathbf{min}} = \mathbf{R}^{-1} \cdot \mathbf{p}. \tag{E.17}$$

Gradient Descent on the Error Surface Given the knowledge of **R**, also called the **input correlation matrix**, and **p**, we saw how it was possible to calculate the weight vector directly. In many problems of practical interest, we do not know the values of **R** and **p**. In these cases we must find an alternate method for discovering the minimum point on the **error surface**.

To initiate training, we assign arbitrary values to the weights, which establishes the error, ξ, at a certain value. As we apply each training pattern to the network, we can adjust the weight vector slightly in the direction of the greatest downward slope.

To perform this gradient descent, the equation of the surface must be known in which case the weight vector can be calculated directly. This discussion of the principles of gradient descent is by way of introduction to the next section that investigates how to approximate the process in the absence of complete knowledge of the error surface.

The Delta Rule Suppose we cannot specify the **R** matrix or **p** vector in advance, or suppose that the number of input vectors is so large as to make the calculations excessively time consuming. There may also be a case in which the distribution function of the input vectors changes as a function of time.

The gradient descent method can still be used by employing a local approximation to the error surface which is valid for a particular input vector.

First, apply a particular input pattern, say the kth, and note the output, y_k. Then determine the error ε_k. Instead of applying other patterns and accumulating the squared error, this error value is used directly. As an approximation to the mean-squared error in Equation E.14, the local value of the squared error can be used for a particular pattern. That is

$$\xi = \langle \varepsilon_k^2 \rangle \approx \varepsilon_k^2 = \xi_k. \tag{E.18}$$

Since ξ_k is a function of the weights we can compute the gradient

$$\xi_k \approx \left(d_k - \sum_{i+1}^{n} w_i(x_i)_k)\right)^2$$

$$\frac{\partial \xi_k}{\partial w_i} = -2\left(d_k - \sum_{i+1}^{n} w_i(x_i)_k)\right)(x_i)_k = -2\varepsilon_k(x_i)_k.$$

We then adjust the weight value, in the case w_i, by a small amount in the direction opposite to the gradient. In other words, we update the weight value according to the following prescription

$$w_i(t+1) = w_i(t) + \eta\varepsilon(x_i)_k \tag{E.19}$$

or in vector form

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta\varepsilon(\mathbf{x})_k \tag{E.20}$$

where η is called the **learning rate parameter** and usually has a value much less than one.

Equations E.19 and E.20 are expressions of a learning law called the LMS rule, or delta rule. By repeated application of this rule, using all of the input vectors, the point on the error surface moves down the slope toward the minimum point, though it does not necessarily follow the exact gradient of the surface. As the weight vector moves toward the minimum point, the error values will decrease. Iteration continues until the errors have been reduced to an acceptable value, the definition of *acceptable* being determined by the requirements of the application.

E.3.2 Training by Back-Propagation of Error

In the Section E.3.1 describing the learning paradigm, the delta rule was used to calculate an approximation to the optimum weight vector that would allow an ALC to correctly map input vectors to output values in accordance with certain examples used during the training process. Extending the rule to multiple-layer networks requires that a non-linear output function be added to the units, and this fact complicates the situation. Moreover, since we have no foreknowledge of the *correct* output values for units on any layer other than the output layer, we have to resort to other methods to determine the weight updates.

This section looks at a method for calculating weight updates that is known as **back-propagation of errors** (BPN). BPN is quite expensive computationally, especially during the training process. Many people have attempted, therefore, to modify the basic back-propagation algorithm to speed up training. A few of these methods are examined.

E.3.2.1 The Generalised Delta Rule

In this section we extend the delta rule to multi-layered networks. Before performing the derivation of the **generalised delta rule** (GDR) the architecture of multi-layered neural networks is examined with particular reference to the features of BPN.

BPN Architecture The standard BPN architecture appears in Figure E.8. The **bias units** shown in that figure are optional. Bias units always have an output of one and they are connected to all units on their respective layer. The weights on the connections from bias units are called **bias terms** or **bias weights**. Units on all layers calculate their net-input values

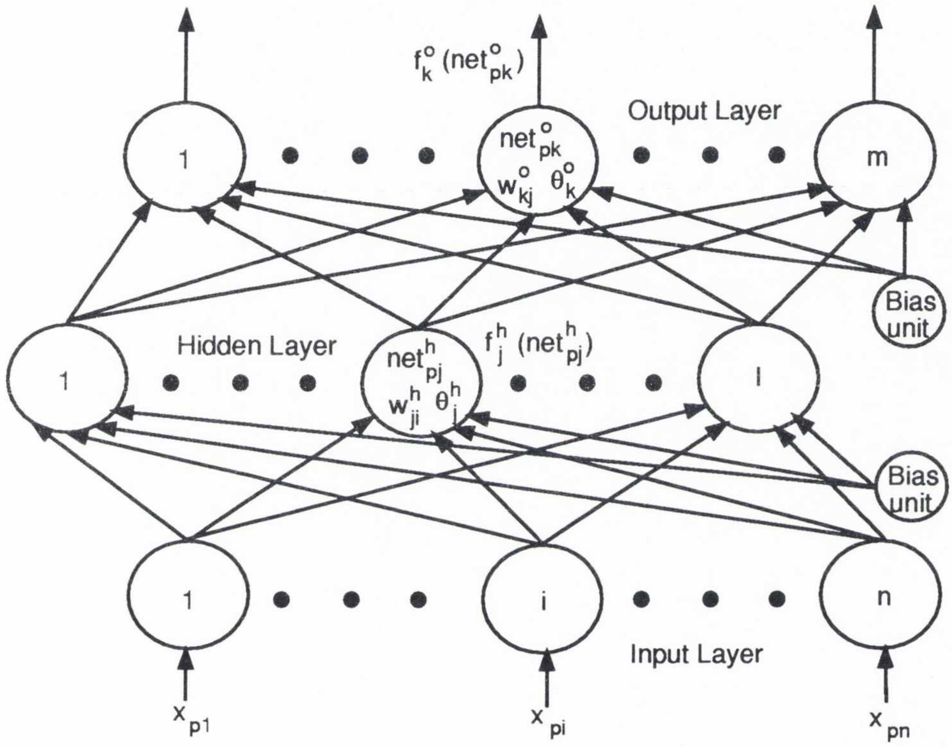


Figure E.8: Typical structure for a BPN.[4]

in accordance with the standard sum-of-productions calculation described in Section E.1. For the hidden-layer units

$$net_{pj}^h = \sum_{i=1}^{N} w_{ji}^h x_{pi} + \theta_j^h \tag{E.21}$$

[4]Although there is only one hidden layer in this figure, you can have more than one. The superscripts on the various quantities identify the layer. The p subscript refers to the pth input pattern.

and for the output-layer units

$$net_{pk}^{o} = \sum_{j=1}^{L} w_{kj}^{o} i_{pj} + \theta_{k}^{o} \tag{E.22}$$

where i_{pj} is the input from the jth hidden-layer unit to the output layer units for the pth input pattern, and the θs are the bias values. N and L refer to the number of units on the input and hidden layers respectively.

Unlike ALC, the output function of these units is not necessarily the simple identity function, although it can be in the case of the output units. As a rule, the output function will be the sigmoid function (see Equation E.7). Then the outputs of the units are

$$i_{pj} = f_{j}^{h}(net_{pj}^{h}) = \frac{1}{1 + e^{-net_{pj}^{h}}} \tag{E.23}$$

for units on the hidden layer, and

$$o_{pk} = f_{k}^{o}(net_{pk}^{o}) = \frac{1}{1 + e^{-net_{pk}^{o}}} \tag{E.24}$$

for units on the output layer.

We can use the identity function on the output-layer unit, in which case we have

$$o_{pk} = net_{pk}^{o}.$$

If we were to use the identity function on the hidden-layer units, then the network would not be able to perform many of the complex input-output mappings that would otherwise be possible.

When we propagate data through the network from inputs to outputs, we can streamline the calculation by putting all of the weight values for a single layer into a **weight matrix**. Each row of the matrix represents the weights on a single unit of the layer. There would then be L rows, where L is the number of units on the layer. If there are N inputs, there would be N or $N+1$ columns, the latter figure including a place for the bias weight.

Derivation of the GDR As was done with the ALC, the problem is stated in more formal terms. Suppose we have a set of P vector-pairs (exemplars), $(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), \ldots, (\mathbf{x_P}, \mathbf{y_P})$, that are examples of functional mapping

$$\mathbf{y} = \Phi(\mathbf{x}), \ \mathbf{x} \in R_N, \ \mathbf{y} \in R_M \tag{E.25}$$

where $\mathbf{x}$ and $\mathbf{y}$ are N- and M-dimensional real vectors respectively. We wish to train a neural network (i.e., find a set of weights) to learn an approximation to that functional mapping. To develop the training algorithm we use the same approach that was used for the ALC in Section E.3.1, that is gradient descent down an error surface.

The error that is minimised by the training algorithm is

$$E_p = \frac{1}{2}\sum_{k=1}^{M} \delta_{pk}^2 \tag{E.26}$$

where

$$\delta_{pk} = (y_{pk} - o_{pk}). \tag{E.27}$$

The subscript, p, refers to the pth exemplar, o_{pk} is the output of the kth output-layer unit for the pth exemplar, and there are M output-layer units.

Equation E.26 represents a local approximation to the global error surface

$$E = \sum_{p=1}^{P} E_p.$$

Using the local approximation simplifies the calculation here, as it did in Section E.3.1.

Substituting Equation E.27 into Equation E.26, and using Equation E.24, we find

$$E_p = \frac{1}{2} \sum_{k=1}^{M} (y_{pk} - f_k^o(net_{pk}^o))^2.$$

The gradient of E_p with respect to the output-layer weights, is

$$\frac{\partial E_p}{\partial w_{kj}^o} = -(y_{pk} - o_{pk}) \frac{\partial f_k^o(net_{pk}^o)}{\partial (net_{pk}^o)} \frac{\partial (net_{pk}^o)}{\partial w_{kj}^o}.$$

For now we shall write the partial derivative of the output function

$$\frac{\partial f_k^o(net_{pk}^o)}{\partial (net_{pk}^o)} = f_k^{o\prime}(net_{pk}^o).$$

Using Equation E.24 we can show that

$$\frac{\partial (net_{pk}^o)}{\partial w_{kj}^o} = i_{pj}.$$

Finally, we can write the gradient of the error surface as

$$\frac{\partial E_p}{\partial w_{kj}^o} = -(y_{pk} - o_{pk}) f_k^{o\prime}(net_{pk}^o) i_{pj}. \tag{E.28}$$

By a similar, and only slightly more complicated analysis, we can find the gradient of the error surface with respect to the hidden-layer weights

$$\frac{\partial E_p}{\partial w_{ji}^h} = -f_j^{h\prime}(net_{pj}^h) x_{pi} \sum_{k=1}^{M} (y_{pk} - o_{pk}) f_k^{o\prime}(net_{pk}^o) w_{kj}^o. \tag{E.29}$$

The derivatives of the output functions are "primed" functions instead of explicitly calculated, because the value of that derivation depends on the form of the output function. The two primary cases of interest are the sigmoid and the identity function. In these two cases, the derivatives of the functions for output-layer units are

$$f_k^{o\prime}(net_{pk}^o) = o_{pk}(1 - o_{pk}) \tag{E.30}$$

for the sigmoid, and

$$f_k^{o\prime}(net_{pk}^o) = 1 \tag{E.31}$$

for the identity function.

As each training pattern is presented to the network, the information is propagated forward to determine the actual network outputs. Then the error terms on the output layer and the gradient of the error surface are calculated with respect to each of the output-layer weights. Next, the gradient of the error surface is calculated with respect to each of the weights on the hidden layer. A study of Equation E.29 shows that, for any given unit on the hidden layer, the gradient of the error surface depends on *all* of the errors on the output layer. This dependency is reasonable, since any change on a hidden-layer weight will have an effect on all of the output values of the output layer. Here is where the concept of *back-propagation* enters formally: having calculated errors on the output layer first, these are brought *back* to the hidden layer to calculate the surface gradients there. Having calculated the gradients, each weight value is adjusted by a small amount in the direction of the negative of the gradient. The proportionality constant is called the **learning-rate parameter**, just as it was for the ALC in Section E.3.1.

The next input pattern is then presented and the weight-update process repeated. The process continues until all output-layer errors have been reduced to an acceptable value.

The notation can be simplified through the use of some auxiliary variables. Defining the **output-layer delta** as

$$\delta^o_{pk} = (y_{pk} - o_{pk}) f^{o\prime}_k(net^o_{pk}) = \delta_{pk} f^{o\prime}_k(net^o_{pk}) \tag{E.32}$$

and the **hidden-layer delta** as

$$\delta^h_{pj} = f^{h\prime}_j(net^h_{pj}) x_{pi} \sum_{k=1}^{M} \delta^o_{pk} w^o_{kj}. \tag{E.33}$$

Using these definitions the weight-update equations on both layers take on a similar from

$$w_{kj}^{o}(t+1) = w_{kj}^{o}(t) + \eta \delta_{pk}^{o} i_{pj} \tag{E.34}$$

on the output layer, and

$$w_{ji}^{h}(t+1) = w_{ji}^{h}(t) + \eta \delta_{pj}^{h} i_{pi} \tag{E.35}$$

on the hidden layer. η is the learning-rate parameter, and it was assumed that it is the same on all units on all layers. This assumption is typically a good one, and is employed in this thesis.

E.3.2.2 BPN Variations

The BPN algorithm requires a large amount of computation for each iteration. Two ways to speed convergence are mentioned below.

Momentum The BPN algorithm can be modified by the addition of a term called **momentum** to the weight-update equations. The term will have a significant effect on the learning speed, in terms of the number of iterations required.

The idea behind momentum in a neural network is straightforward: once you start adjusting weights in a certain direction, keep them moving generally in that direction. In more practical terms, after you adjust the weights during one training iteration, save the value of that adjustment; when calculating the adjustment for the next iteration, add a fraction of the previous change to the new one. In terms of an equation (in this case, for the hidden-layer weights)

$$w_{ji}(t+1) = w_{ji} + \eta \delta_{pj} x_{pi} + \alpha \Delta w_{ji}(t) \tag{E.36}$$

where α is called the momentum term, typically a positive number less than one, and

$$\Delta w_{ji}(t) = w_{ji}(t) - w_{ji}(t-1).$$

An additional modification is to set a maximum acceptable error for any one pattern to some number, say 0.1. A conditional statement is added to the program so that, if the error for an input pattern is less than this acceptable value, no weight updates occur during that iteration. That way, the network is not over learning one pattern at the expense of the others.

Competitive Weight Updates The BPN can be modified to include a competitive algorithm to update the weight values. From Equations E.34 and E.35, weight changes are proportional to the delta terms; thus, we might reason that the unit with the largest value of delta should adjust its weighs by the largest amount. All other units on the layer should adjust their weights in the direction opposite to that *winning* unit. In other words, after we calculate the delta values on a layer, we search for the unit with the largest delta (we need to look for the largest magnitude). That unit is declared the winner of the competition, and the delta value for all units on the layer becomes a function of that unit's delta.

For the hidden layer, the delta value is given by Equation E.33

$$\delta_{pj}^{h} = f_j^{h'}(net_{pj}^{h})x_{pi}\sum_{k=1}^{M}\delta_{pk}^{o}w_{kj}^{o}.$$

Let

$$\varepsilon_{pj}^{h} = \begin{cases} \max_m(\delta_{pm}^{h}) & j = \text{winning unit} \\ -\frac{1}{4}\max_m(\delta_{pm}^{h}) & \text{otherwise.} \end{cases} \tag{E.37}$$

Then Equation E.35 becomes

$$w_{ji}^{h}(t+1) = w_{ji}^{h}(t) + \eta\varepsilon_{pj}^{h}x_{pi} \tag{E.38}$$

with a similar equation for the output layer.

E.4 IMPLICATIONS FOR PROBLEM SOLVING

The previous sections have outlined the fundamental theoretical tenants supporting feed-forward neural networks. This section considers their implications for problem solving.

Neural networks in general, and multi-layer feed-forward networks in particular suffer from some small but annoying shortcomings. It is nearly impossible to specify an effective neural network design or architecture based only on the description of a problem. *Experimentation* is also required. After a network is trained, it may be difficult to understand how it works. Worse still, the supposition that it will work correctly when presented with any possible test must usually be taken on trust. Techniques for strict mathematical verification of a neural network's performance are still in their infancy. It is well established, however, that these networks *do* perform well in practice. It is exceedingly rare for a network to be well trained, verified with a reasonable test set, and then to fail in practice. Performance quality is, however, difficult to *prove* at the present time. Because of this shortcoming there is a continual search for a powerful theorem to establish theoretically why neural networks work.

The current state of knowledge about feed-forward neural networks enables us to formulate a number of propositions.

- Firstly, neural networks approximate functions and the converse of this is that if a problem can be expressed as a function it can be modelled as a neural network. A function here is taken to be a mapping from the real-valued vector domain $\Re_n$ to the real numbers $\Re$.

  The results are stated in terms of *what function a network can learn.* Each problem is posed as a definition of a function that a network is asked to learn bearing in mind that a multi-layer feed-forward network implements a function itself. Inputs are applied to it, and deterministic outputs are produced. So we are asking one function, our network, to approximate another function, the problem. The network is said to be able to solve the problem if it is able to learn to approximate the function with ever increasing accuracy.

- Secondly, the structure of the neural network is contingent upon the type of function being approximated. Functions, which can be learned, are classified as follows:

1. Functions which consists of a finite collection of points can be learned by a three-layer network (one hidden layer).
2. Functions which are continuous and defined on a compact domain can be learned by a three-layer network (one hidden layer). Roughly speaking, "compact domain" means that the inputs have definite bounds, rather that having no limits on what they can be.
3. Functions that do not meet the above criteria can also be learned by a three-layer network (one hidden layer). In particular, discontinuities can be theoretically tolerated under all conditions likely to be met in real life. Also, functions whose inputs are normally distributed random variables can be learned by a three-layer network under some conditions.
4. Under very general conditions, all other functions that can be learned by a neural network, can be learned by a four-layer (two hidden) network.

- Thirdly, the majority of practical functions can be approximated by a one hidden layered neural network.

  This means that, theoretically at least, we are always reasonably safe using a single hidden layer. Furthermore, we should never (at least theoretically) need more than two hidden layers. A network having two hidden layers is a universal approximator.

In practice, the need for a second hidden layer comes about in essentially only one way. That is when we need to learn a function that is mostly continuous, but has a few discontinuities. We occasionally are confronted with a function defined on a compact domain that is generally continuous, but has one or more sudden jumps where continuity is lost. These piecewise-continuous functions cannot in general be learned easily by a network having only one hidden layer (though it is theoretically possible). Two hidden layers are usually required.

The most common reason why a function cannot be learned by *any* multi-layer feed-forward network is when it seriously violates the assumption of a compact domain. We cannot expect a function whose behaviour remains unpredictable as its inputs tend toward infinity to be approximated well by ordinary neural networks. Even so, use of trigonometric activation functions can create neural networks that behave like Fourier approximators, thus circumventing even that limitation.

In conclusion, a multi-layer feed-forward network *can* learn virtually all functions. If there are problems, they are *not* due to the model itself. They are due to insufficient training, or insufficient numbers of hidden neurons, or an attempt to learn a supposed function that is not deterministic.

The learning and generalisation capabilities of multiple-layer feed-forward networks are astounding. In practice surprisingly few hidden neurons are normally required. With proper design of the network and training set, the training time is usually moderate. Multiple-layer feed-forward networks should be considered for nearly every neural network task.

Appendix F

Feed-Forward Neural Network Examples

There are very many examples of neural network techniques being successfully used in medicine and in other disciplines. The cases described here illustrate their versatility and capability.

The wide popularity of feed-forward networks for many diverse applications stems from their ability to map or represent many different functions. If a good set of training data is available for a particular application, it is likely that a single or at most a two hidden-layer feed-forward network can learn to master the desired tasks. Although their performance is acknowledged to be as good as other networks, and other non-network techniques, in the solution of diverse problems, neural networks require experience and practice to train. It also remains to be confirmed if their performance holds when scaled up to large networks.

Before describing ANN solutions for each of four broad categories of applications, namely classification and diagnosis, control and optimisation, prediction and forecasting, and pattern recognition, a brief overview of neural network applications generally is offered in the next section.

F.1 NEURAL NETWORK CAPABILITIES

Imaginative researchers are devising new applications for artificial neural networks daily. Some of the more familiar applications include:

Classification Neural networks can be used to determine crop types from satellite photographs, to distinguish a submarine from a boulder given its sonar return, and to identify specific diseases of the heart from electrocardiograms. Any task that can be done by traditional discriminant analysis can be done at least as well (and almost always much better) by a neural network.

Noise Reduction An artificial neural network can be trained to recognise a number of patterns. These patterns may be parts of time-series, images, etc. If a version of one of these patterns, corrupted by noise, is presented to a properly trained network, the network can reproduce the original pattern on which it was trained. This technique has been used with outstanding success in some image restoration problems.

Prediction A very common problem is that of predicting the value of a variable given historic values of that variable (and perhaps other variables). Economic and meteorological models spring to mind. Neural networks have frequently been shown to outperform traditional techniques like frequency domain analysis.

Artificial neural networks are most likely to be superior to other methods under the following conditions:

1. The input data on which conclusions are to be based are "fuzzy." Examples are human opinions and ill-defined categories subject to large error. This type of problem requires the robust behaviour of neural networks.

2. The patterns important to the required decision are subtle or deeply hidden. One of the principal advantages of a neural network is its ability to discover patterns in data which are obscure or imperceptible to human researchers and which are unresponsive to standard statistical methods. One of the first major commercial uses of neural networks was to predict the credit worthiness of loan applicants, based on their spending and payment history. The correct decision depended on much more than simple factors like salary and debt level. Neural networks were found to provide decisions superior to those made by trained humans in this type of situation.

3. The data exhibits significant unpredictable non-linearity. Traditional time-series models for predicting future values, such as Kalman filters, are based on strictly defined models. If the data do not fit the models, results may be useless. Neural networks, on the other hand, are eminently adaptable.

4. The data is chaotic (in the mathematical sense). Chaos can be found in telephone line noise, stock market prices and in a host of other physical processes. Such behaviour is devastating to most other techniques, but neural networks are generally robust with inputs of this type.

The excellent performance of artificial neural networks is not surprising when one considers the solid theoretical foundations on which many of them rest. The standard workhorse, the three-layer (one hidden layer) feed-forward network, has powerful function-approximation capabilities. In particular, *any* continuous function defined over a compact subset of $\Re^n$ can be approximated to arbitrary accuracy given sufficient hidden neurons. The importance of this result cannot be overstated. When combined with the robustness of the three-layer feed-forward network as regards input errors, it is a powerful tool. Rigorous mathematical discussion of these properties is given in [58, 7].

In summary, many artificial neural networks possess both substantial theoretical foundations and practical utility. Any problem that can be solved with traditional modelling or statistical methods can most likely be solved more effectively with a neural network.

F.2 CLASSIFYING CELLS FOR CANCER DIAGNOSIS

This applications falls into the "classification and diagnosis" category of problems which are ubiquitous and occur in many problem domains. An electronic or mechanical system that develops operational problems will have developed one or more faults that needs diagnosing for identification and repair. Manufactured products must be inspected for quality and either accepted or rejected, a form of classification task, and so on. The example given here is the classification of cells in the diagnosis of bladder cancer.

The structure and other characteristics of cells observed from urine samples of patients can provide an accurate indication of bladder cancer. A simple two-category classification scheme of "Well" and "Not Well" is sufficient for the cell diagnosis task when well chosen, discriminant features are used. Several approaches to the classification problem have been proposed, including the use of Selective Mapping Tree Classifiers, feed-forward networks and others [89]. Of the methods reported, only the ANN approach has achieved levels of classification accuracy acceptable for clinical use. For example, the tree classifier system accuracy was of the order of 23% with other non-ANN methods reporting even poorer performance levels. The accuracy achieved using a simple two-layer feed-forward network is of the order of 96%, a significant improvement over the other approaches.

In this example, 43 microscopic images containing 597 objects were used to train and test the network. The images were first examined visually and cells were classified by experts. Of the total objects, 77 were classified as Well and 520 as Not Well. For the experiments, about one half of the objects were selected randomly for training and the balance for testing. Several descriptive features were selected for cell descriptors. These included: (1) the cell area, which ranged between 100 and 400 pixels for Well cells (larger for Not Well cells), (2) cell circularity, a measure of how well the cell approximated a circle, defined as the ratio of the cell's area to the area of a rectangular box containing the cell ($\pi/4$ for a true circle), (3) area of the cell nucleus, (4) circularity of the nucleus and (5) the ratio of the entire cell area to the nucleus. These features were used as input to the neural network which had five input nodes, two hidden layers, each with ten nodes, and a single output node. A cell was interpreted as Well if the output node value was greater than or equal to 0.5 and Not Well otherwise.

Each microscopic image consisted of 256×240 pixels with 256 grey levels. Pre-processing of the images was performed as follows: the grey-level images were first partitioned into 16 segments of 32×120 pixels to account for different lighting backgrounds and shading conditions. The segments were then digitised and converted to binary form after determining threshold levels from a histogram computation and analysis of the segments. The threshold intensity level was chosen to separate the cytologic objects from the background and thereby permit object segmentation. Segmentation was performed using a blob colouring algorithm which assigns homogenous neighbouring pixels 0/1 values [4]. The five-object descriptors defined above were then computed and fed to the neural network as an input pattern vector.

After training the network, tests were conducted on the data set and an accuracy of 93.4% was reported in accepting Well cells and 97.0% in detecting Not Well cells. This gave an average error rate of 3.5%, an acceptable level for clinical use. The time spent for the diagnoses was also quite acceptable. Pre-processing time for each image required 2.6 seconds and classification 0.4 seconds for a total diagnostic time of 3.0 seconds. This compares favourably with other methods that took in the order of 32 seconds per image. Overall, the neural network approach outperformed the other approaches by a significant margin.

F.3 AUTONOMOUSLY DRIVEN LAND VEHICLE

The second example falls into the "control and optimisation" category of problems which are among the more difficult applications for ANNs to master. The mapping functions that must be learned are generally very complex in nature and problem constraints that must be satisfied are often conflicting. Even so, the application of feed-forward networks to such problems has been moderately successful. The example described here relates to the control of an autonomously driven (driver-less) vehicle.

Carnegie-Mellon University converted a commercial van into a laboratory vehicle (Navlab I) in 1986 to act as a test bed for autonomous driving experiments [130, 68]. One of the control systems, an ALVINN (Autonomous Land Vehicle in a Neural Network) is neural net based. The van is equipped with several video cameras, a scanning laser range finder, a global positioning system, an inertial navigation system and sonar sensors. It also carries several computers onboard.

The ALVINN automatic road following control system uses a fully connected three-layer BP network with colour vision inputs from a video camera. The input image is reduced resolution with input retina of 30×32 (also 45×48) pixels. The hidden layer has nine units and the output layer has 45 units. The network learns different sets of weights to follow different types of roads. The basic network architecture is illustrated in Figure F.1.

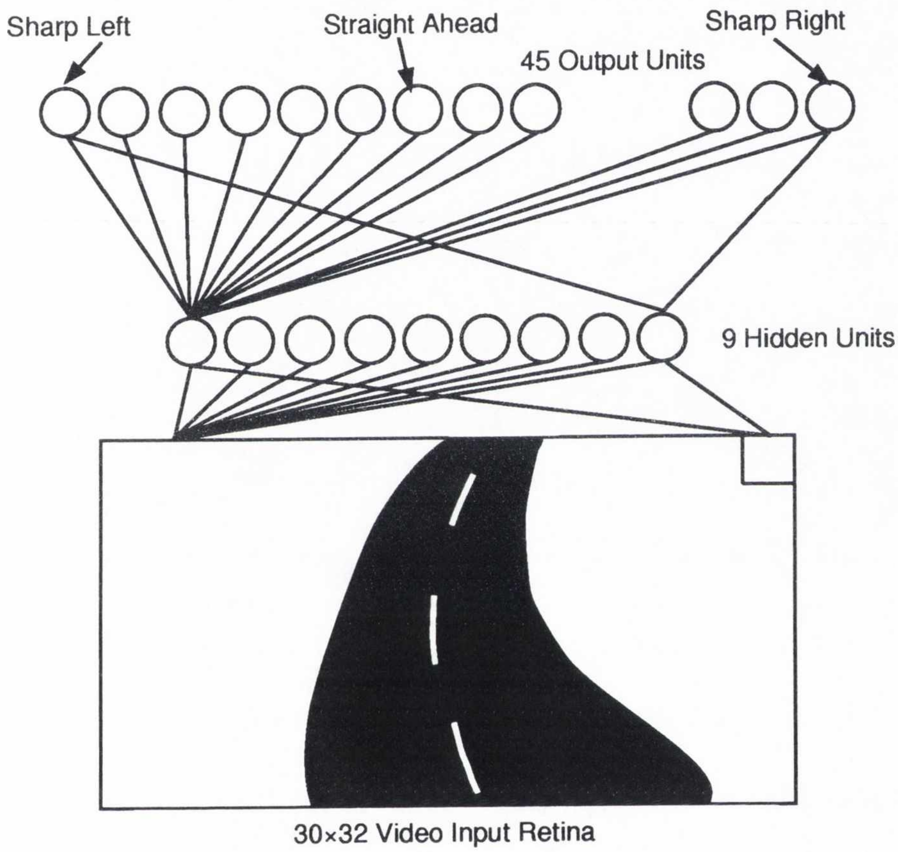


Figure F.1: Multi-layer network control unit for NavLab.

During operation, the image is pre-processed to enhance road contours and the result is fed directly to the network input. The network computes the steering angle directly with no reasoning about road location. The output nodes provide commands to steer the vehicle with varying degrees of angular turning; (45 different angle positions) sharp-left, varying degrees of left, straight ahead, varying degrees of right and far-right. During operation, the output is updated 15 times per second to provide real-time control of the vehicle while travelling at speeds of up to 55mph.

To train the network, a unique "on-the-fly" procedure is used. Images are inputted and processed while someone drives the vehicle down a section of the selected road or highway. Thus, the training set consists of "snapshots" of the highway images (input vector) and the angular position of the steering mechanism (target vector). Each image is reused in several positions during training. The image is transformed through lateral translations to provide several positions, shifted to simulate erroneous correspondence between road and steering angle. The appropriate commands to correct the error are also determined. These modified images and steering angles are all provided as part of the training set. This type of training gives a "view" of the road consistent with ideal human driving. The system also uses three-dimensional images from laser range finders to detect obstacles along the roadway (trees, cars, mailboxes, rocks).

The ALVINN project has met with good success where the vehicle has travelled at speeds up to 55mph for distances of 90 miles or more. It has also been tested successfully on various road types (dirt, paved single and double-lane) and under various weather conditions.

ALVINN is not the only successful ANN control system for autonomous vehicle driving. The Advance Research Projects Agency (ARPA) have also built autonomous vehicles (ALV) as well as European and Japanese organisations. The potential for (commercial and military) driver-less vehicle applications is very great and warrants considerable interest and resources.

F.4 PREDICTING CREDIT WORTHINESS FOR LOAN APPLICATIONS

Prediction is a task every organisation must learn to do. A consumer products company will want to know the expected growth in sales for a new product that it plans to introduce. Meteorologists need to predict the weather. Banks want to predict the credit worthiness of companies as a basis for granting loans. Airport management groups want to predict the growth in passenger arrivals at busy airports, and electric power companies want to know customer demand for power in the future, and so on. ANNs have been shown to be successful as predictive tools in a variety of ways—predicting that some event will or will not occur, predicting the time at which an event will occur, or predicting the level of some outcome. To predict with an acceptable level of accuracy, an ANN must be trained with a sizeable number of examples of past patterns, together with known future outcome values. The training set will generally come from historical data that has been collected over a given time period. The ANN must then learn to generalise and extrapolate from new patterns to predict future outcomes. The example of this category of problem described here is the assessment of the credit worthiness of loan applicants.

Chase Manhatten Bank implemented one of the most successful applications of predictive models using neural networks in 1990. The system is a hybrid, statistical-based network that assesses the credit worthiness of public corporations seeking business loans. Chase loans out some US$ 300 million (€ 257.18 million) annually to qualifying companies. Since this can be a significant source of profit (or loss), the ability to accurately forecast the credit worthiness of potential customers is essential.

The overall system known as Creditview performs three-year forecasts that assign a risk classification of *good*, *criticised*, or *charged-off*. The system also provides a detailed listing of items that significantly contribute to the forecast. A conventional expert system interprets the items and produces various comparison reports for senior loan officers. A front-end system to the neural network, known as ADAM, receives historical input data from a financial database, together with good and bad data on the customer seeking the loan. These data serve as part of the training set. ADAM also receives input on industry norms, financial data for the normalisation of specific industry categories. ADAM generates candidate variables that may indicate the future financial condition of a company. This is compiled as profile data in the form of logical feature vectors that form the basis for a neural network model of the company in question. The neural network, known as the public company model (PCLM) Forecaster then produces a rating for the company. The combined system is illustrated in Figure F.2.

The PCLM accepts six years of past financial data for the company being rated. It uses the expression produced by ADAM to predict the financial health of the company three years into the future. The predictions determine the likelihood of a company being rated as good, criticised or charge-off. In addition, the PCLM identifies the strengths and weaknesses in the financial structure of the company. Extensive reports giving comparative risk esti-

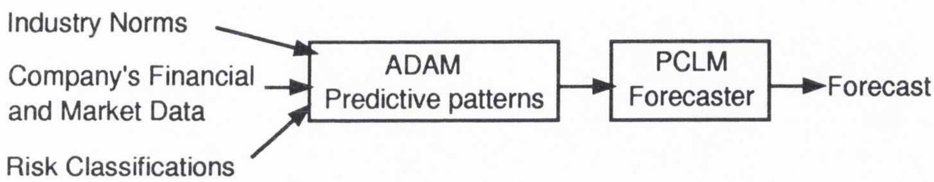


Figure F.2: Chase Manhatten Bank's hybrid neural network credit rater.

mates and a text explanation of the analysis are provided to the user. Chase [90, 91] has tested the system extensively and uncovered a number of troublesome loans. The system was put in operation around 1990 and is being extended and enhanced to include private corporation evaluations as well.

F.5 DETECTION OF EPILEPTIC ATTACKS

The "pattern recognition" category of problem includes such applications as speech and character recognition for various languages, visual image recognition and classification, and various types of signal and chart analysis including electrocardiograms, electroencefelographs, electrocardiographs, and various graph analysis for process alarm monitoring. Pattern recognition applications are generally difficult to master. They are closely related to cognitive tasks which humans perform almost effortlessly. In this section an application describing the early detection of epileptic attacks is described.

Early detection of an impending attack in epileptic patients makes it possible to start early reactions and treatment. The preictal[1] period is not easily detected however, since the temporal data vary greatly from patient to patient and the course of the attack is also a source of variability. Because of the strong patient dependency, conventional statistical and AI methods have been very ineffective in early detection. With patient specific data available on prior attacks, the use of ANN technology is possible since training data sets can be constructed.

Patient data are collected by implanting EEG electrodes subdurally in a patient and recording the analogue brain wave signals over a period of time prior to an attack. A number of 100–200Hz signals from different locations are recorded graphically for analysis by an electrophysiologist and also digitised and stored for subsequent use in training feed-forward networks. The physiologist can analyse and edit the data using a graphical screen editor, eliminating artefacts and other irrelevant data. Further pre-processing of the data has been found to be effective in reducing the size of the data set. Sliding windows over the time series data are used to selectively sample sections of pattern data points. This data is used in the computation of averages, variances, histograms and other summary statistics. The data from different channels are then divided among several feed-forward networks for independent training and the outputs from these networks are then used for input to a final prediction network. By combining the outputs of several independent networks into a vector for input to a final network, researchers have been able to improve the response-quality of the early detection system [48]. The detection system is illustrated in Figure F.3.

[1] Occurring before a convulsion or stroke.

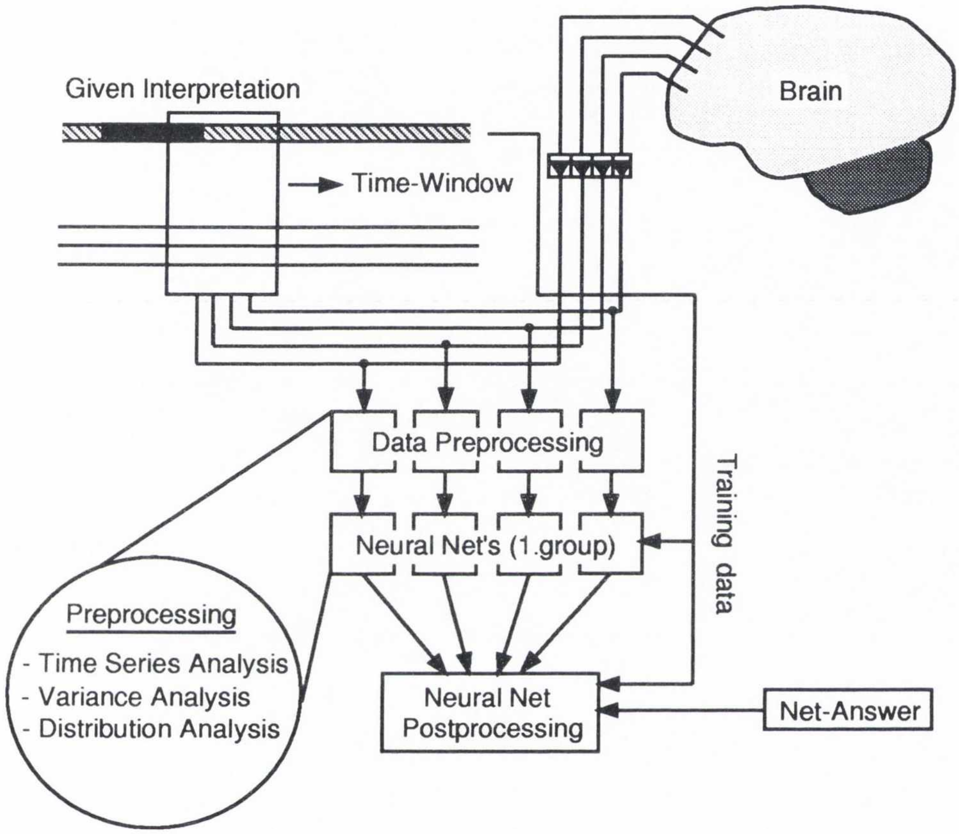


Figure F.3: System for early detection of epileptic attacks.

The success of the system was found to be heavily dependent on the quality of the human data interpretations, and particularly on the pre-processing methods applied to the source data. The use of ANNs for early detection has led to more efficient analysis and has resulted in more precise and more stable detection. This work has helped to pave the way for the development of a real-time automatic detection system that is now deemed feasible.

Appendix G

Mathematica^{TM} Source Code Listings

This appendix contains the *Mathematica*^{TM} source code listings for patient-related and implant related feed-forward neural networks. All the data used to train and validate the neural network are included here for completeness. This data was randomly divided into test and validation sets for the experiment.

G.1 PATIENT-RELATED FEED-FORWARD NEURAL NETWORK

```
<<DiscreteMath`Backpropagation`
ioPairsTC= ioPairs=
```

```
{{{0.900,0.100,0.100,0.900,0.100,0.100},{0.900}},
{{0.900,0.100,0.100,0.900,0.100,0.153},{0.900}},
{{0.900,0.100,0.100,0.900,0.100,0.260},{0.888}},
{{0.900,0.100,0.100,0.900,0.100,0.367},{0.872}},
{{0.900,0.100,0.100,0.900,0.100,0.473},{0.840}},
{{0.900,0.100,0.100,0.900,0.100,0.580},{0.800}},
{{0.900,0.100,0.100,0.900,0.100,0.687},{0.848}},
{{0.900,0.100,0.100,0.900,0.100,0.793},{0.736}},
{{0.900,0.100,0.100,0.900,0.100,0.900},{0.708}},
{{0.900,0.100,0.100,0.900,0.367,0.100},{0.900}},
{{0.900,0.100,0.100,0.900,0.367,0.153},{0.900}},
{{0.900,0.100,0.100,0.900,0.367,0.260},{0.884}},
{{0.900,0.100,0.100,0.900,0.367,0.367},{0.860}},
{{0.900,0.100,0.100,0.900,0.367,0.473},{0.844}},
{{0.900,0.100,0.100,0.900,0.367,0.580},{0.796}},
{{0.900,0.100,0.100,0.900,0.367,0.687},{0.740}},
{{0.900,0.100,0.100,0.900,0.367,0.793},{0.688}},
{{0.900,0.100,0.100,0.900,0.367,0.900},{0.644}},
{{0.900,0.100,0.100,0.900,0.633,0.100},{0.900}},
{{0.900,0.100,0.100,0.900,0.633,0.153},{0.900}},
{{0.900,0.100,0.100,0.900,0.633,0.260},{0.892}},
{{0.900,0.100,0.100,0.900,0.633,0.367},{0.876}},
{{0.900,0.100,0.100,0.900,0.633,0.473},{0.848}},
{{0.900,0.100,0.100,0.900,0.633,0.580},{0.804}},
{{0.900,0.100,0.100,0.900,0.633,0.687},{0.780}},
```

```
{{0.900,0.100,0.100,0.900,0.633,0.793},{0.752}},
{{0.900,0.100,0.100,0.900,0.633,0.900},{0.732}},
{{0.900,0.100,0.100,0.900,0.900,0.100},{0.900}},
{{0.900,0.100,0.100,0.900,0.900,0.153},{0.900}},
{{0.900,0.100,0.100,0.900,0.900,0.260},{0.896}},
{{0.900,0.100,0.100,0.900,0.900,0.367},{0.884}},
{{0.900,0.100,0.100,0.900,0.900,0.473},{0.876}},
{{0.900,0.100,0.100,0.900,0.900,0.580},{0.860}},
{{0.900,0.100,0.100,0.900,0.900,0.687},{0.848}},
{{0.900,0.100,0.100,0.900,0.900,0.793},{0.844}},
{{0.900,0.100,0.100,0.900,0.900,0.900},{0.844}},
{{0.900,0.100,0.100,0.100,0.100,0.100},{0.900}},
{{0.900,0.100,0.100,0.100,0.100,0.153},{0.900}},
{{0.900,0.100,0.100,0.100,0.100,0.260},{0.892}},
{{0.900,0.100,0.100,0.100,0.100,0.367},{0.880}},
{{0.900,0.100,0.100,0.100,0.100,0.473},{0.860}},
{{0.900,0.100,0.100,0.100,0.100,0.580},{0.832}},
{{0.900,0.100,0.100,0.100,0.100,0.687},{0.796}},
{{0.900,0.100,0.100,0.100,0.100,0.793},{0.772}},
{{0.900,0.100,0.100,0.100,0.100,0.900},{0.752}},
{{0.900,0.100,0.100,0.100,0.367,0.100},{0.900}},
{{0.900,0.100,0.100,0.100,0.367,0.153},{0.900}},
{{0.900,0.100,0.100,0.100,0.367,0.260},{0.896}},
{{0.900,0.100,0.100,0.100,0.367,0.367},{0.884}},
{{0.900,0.100,0.100,0.100,0.367,0.473},{0.860}},
```

```
{{0.900,0.100,0.100,0.100,0.367,0.580},{0.840}},
{{0.900,0.100,0.100,0.100,0.367,0.687},{0.812}},
{{0.900,0.100,0.100,0.100,0.367,0.793},{0.780}},
{{0.900,0.100,0.100,0.100,0.367,0.900},{0.768}},
{{0.900,0.100,0.100,0.100,0.633,0.100},{0.900}},
{{0.900,0.100,0.100,0.100,0.633,0.153},{0.900}},
{{0.900,0.100,0.100,0.100,0.633,0.260},{0.896}},
{{0.900,0.100,0.100,0.100,0.633,0.367},{0.884}},
{{0.900,0.100,0.100,0.100,0.633,0.473},{0.860}},
{{0.900,0.100,0.100,0.100,0.633,0.580},{0.836}},
{{0.900,0.100,0.100,0.100,0.633,0.687},{0.808}},
{{0.900,0.100,0.100,0.100,0.633,0.793},{0.784}},
{{0.900,0.100,0.100,0.100,0.633,0.900},{0.772}},
{{0.900,0.100,0.100,0.100,0.900,0.100},{0.900}},
{{0.900,0.100,0.100,0.100,0.900,0.153},{0.900}},
{{0.900,0.100,0.100,0.100,0.900,0.260},{0.898}},
{{0.900,0.100,0.100,0.100,0.900,0.367},{0.890}},
{{0.900,0.100,0.100,0.100,0.900,0.473},{0.882}},
{{0.900,0.100,0.100,0.100,0.900,0.580},{0.874}},
{{0.900,0.100,0.100,0.100,0.900,0.687},{0.872}},
{{0.900,0.100,0.100,0.100,0.900,0.793},{0.870}},
{{0.900,0.100,0.100,0.100,0.900,0.900},{0.868}},
{{0.100,0.900,0.100,0.900,0.100,0.100},{0.900}},
{{0.100,0.900,0.100,0.900,0.100,0.153},{0.900}},
{{0.100,0.900,0.100,0.900,0.100,0.260},{0.884}},
```

```
{{0.100,0.900,0.100,0.900,0.100,0.367},{0.856}},
{{0.100,0.900,0.100,0.900,0.100,0.473},{0.822}},
{{0.100,0.900,0.100,0.900,0.100,0.580},{0.772}},
{{0.100,0.900,0.100,0.900,0.100,0.687},{0.740}},
{{0.100,0.900,0.100,0.900,0.100,0.793},{0.700}},
{{0.100,0.900,0.100,0.900,0.100,0.900},{0.644}},
{{0.100,0.900,0.100,0.900,0.367,0.100},{0.900}},
{{0.100,0.900,0.100,0.900,0.367,0.153},{0.900}},
{{0.100,0.900,0.100,0.900,0.367,0.260},{0.888}},
{{0.100,0.900,0.100,0.900,0.367,0.367},{0.876}},
{{0.100,0.900,0.100,0.900,0.367,0.473},{0.844}},
{{0.100,0.900,0.100,0.900,0.367,0.580},{0.808}},
{{0.100,0.900,0.100,0.900,0.367,0.687},{0.778}},
{{0.100,0.900,0.100,0.900,0.367,0.793},{0.708}},
{{0.100,0.900,0.100,0.900,0.633,0.100},{0.900}},
{{0.100,0.900,0.100,0.900,0.633,0.153},{0.900}},
{{0.100,0.900,0.100,0.900,0.633,0.260},{0.892}},
{{0.100,0.900,0.100,0.900,0.633,0.367},{0.876}},
{{0.100,0.900,0.100,0.900,0.633,0.473},{0.868}},
{{0.100,0.900,0.100,0.900,0.633,0.580},{0.856}},
{{0.100,0.900,0.100,0.900,0.633,0.687},{0.796}},
{{0.100,0.900,0.100,0.900,0.633,0.793},{0.784}},
{{0.100,0.900,0.100,0.900,0.900,0.100},{0.900}},
{{0.100,0.900,0.100,0.900,0.900,0.153},{0.900}},
{{0.100,0.900,0.100,0.900,0.900,0.260},{0.900}},
```

```
{{0.100,0.900,0.100,0.900,0.900,0.367},{0.900}},
{{0.100,0.900,0.100,0.100,0.100,0.100},{0.900}},
{{0.100,0.900,0.100,0.100,0.100,0.153},{0.900}},
{{0.100,0.900,0.100,0.100,0.100,0.260},{0.892}},
{{0.100,0.900,0.100,0.100,0.100,0.367},{0.880}},
{{0.100,0.900,0.100,0.100,0.100,0.473},{0.844}},
{{0.100,0.900,0.100,0.100,0.100,0.580},{0.800}},
{{0.100,0.900,0.100,0.100,0.100,0.687},{0.772}},
{{0.100,0.900,0.100,0.100,0.100,0.793},{0.724}},
{{0.100,0.900,0.100,0.100,0.100,0.900},{0.702}},
{{0.100,0.900,0.100,0.100,0.367,0.100},{0.900}},
{{0.100,0.900,0.100,0.100,0.367,0.153},{0.900}},
{{0.100,0.900,0.100,0.100,0.367,0.260},{0.892}},
{{0.100,0.900,0.100,0.100,0.367,0.367},{0.884}},
{{0.100,0.900,0.100,0.100,0.367,0.473},{0.868}},
{{0.100,0.900,0.100,0.100,0.367,0.580},{0.836}},
{{0.100,0.900,0.100,0.100,0.367,0.687},{0.820}},
{{0.100,0.900,0.100,0.100,0.367,0.793},{0.806}},
{{0.100,0.900,0.100,0.100,0.367,0.900},{0.804}},
{{0.100,0.900,0.100,0.100,0.633,0.100},{0.900}},
{{0.100,0.900,0.100,0.100,0.633,0.153},{0.900}},
{{0.100,0.900,0.100,0.100,0.633,0.260},{0.898}},
{{0.100,0.900,0.100,0.100,0.633,0.367},{0.892}},
{{0.100,0.900,0.100,0.100,0.633,0.473},{0.876}},
{{0.100,0.900,0.100,0.100,0.633,0.580},{0.870}},
```

```
{{0.100,0.900,0.100,0.100,0.633,0.687},{0.864}},
{{0.100,0.900,0.100,0.100,0.633,0.793},{0.862}},
{{0.100,0.900,0.100,0.100,0.633,0.900},{0.862}},
{{0.100,0.900,0.100,0.100,0.900,0.100},{0.900}},
{{0.100,0.900,0.100,0.100,0.900,0.153},{0.899}},
{{0.100,0.900,0.100,0.100,0.900,0.260},{0.898}},
{{0.100,0.900,0.100,0.100,0.900,0.367},{0.894}},
{{0.100,0.900,0.100,0.100,0.900,0.473},{0.868}},
{{0.100,0.900,0.100,0.100,0.900,0.580},{0.844}},
{{0.100,0.100,0.900,0.900,0.100,0.100},{0.900}},
{{0.100,0.100,0.900,0.900,0.100,0.153},{0.896}},
{{0.100,0.100,0.900,0.900,0.100,0.260},{0.884}},
{{0.100,0.100,0.900,0.900,0.100,0.367},{0.852}},
{{0.100,0.100,0.900,0.900,0.100,0.473},{0.784}},
{{0.100,0.100,0.900,0.900,0.100,0.580},{0.712}},
{{0.100,0.100,0.900,0.900,0.100,0.687},{0.636}},
{{0.100,0.100,0.900,0.900,0.367,0.100},{0.900}},
{{0.100,0.100,0.900,0.900,0.367,0.153},{0.900}},
{{0.100,0.100,0.900,0.900,0.367,0.260},{0.892}},
{{0.100,0.100,0.900,0.900,0.367,0.367},{0.868}},
{{0.100,0.100,0.900,0.900,0.367,0.473},{0.844}},
{{0.100,0.100,0.900,0.900,0.367,0.580},{0.812}},
{{0.100,0.100,0.900,0.900,0.367,0.687},{0.788}},
{{0.100,0.100,0.900,0.900,0.367,0.793},{0.756}},
{{0.100,0.100,0.900,0.900,0.367,0.900},{0.700}},
```

```
{{0.100,0.100,0.900,0.900,0.633,0.100},{0.900}},
{{0.100,0.100,0.900,0.900,0.633,0.153},{0.900}},
{{0.100,0.100,0.900,0.900,0.633,0.260},{0.892}},
{{0.100,0.100,0.900,0.900,0.633,0.367},{0.884}},
{{0.100,0.100,0.900,0.900,0.633,0.473},{0.876}},
{{0.100,0.100,0.900,0.900,0.633,0.580},{0.860}},
{{0.100,0.100,0.900,0.900,0.633,0.687},{0.836}},
{{0.100,0.100,0.900,0.900,0.633,0.793},{0.820}},
{{0.100,0.100,0.900,0.900,0.633,0.900},{0.812}},
{{0.100,0.100,0.900,0.900,0.900,0.100},{0.900}},
{{0.100,0.100,0.900,0.900,0.900,0.153},{0.900}},
{{0.100,0.100,0.900,0.900,0.900,0.260},{0.896}},
{{0.100,0.100,0.900,0.900,0.900,0.367},{0.884}},
{{0.100,0.100,0.900,0.900,0.900,0.473},{0.882}},
{{0.100,0.100,0.900,0.900,0.900,0.580},{0.868}},
{{0.100,0.100,0.900,0.900,0.900,0.687},{0.864}},
{{0.100,0.100,0.900,0.100,0.100,0.100},{0.900}},
{{0.100,0.100,0.900,0.100,0.100,0.153},{0.898}},
{{0.100,0.100,0.900,0.100,0.100,0.260},{0.884}},
{{0.100,0.100,0.900,0.100,0.100,0.367},{0.864}},
{{0.100,0.100,0.900,0.100,0.100,0.473},{0.804}},
{{0.100,0.100,0.900,0.100,0.100,0.580},{0.772}},
{{0.100,0.100,0.900,0.100,0.100,0.687},{0.756}},
{{0.100,0.100,0.900,0.100,0.100,0.793},{0.752}},
{{0.100,0.100,0.900,0.100,0.367,0.100},{0.900}},
```

```
{{0.100,0.100,0.900,0.100,0.367,0.153},{0.900}},
{{0.100,0.100,0.900,0.100,0.367,0.260},{0.892}},
{{0.100,0.100,0.900,0.100,0.367,0.367},{0.884}},
{{0.100,0.100,0.900,0.100,0.367,0.473},{0.860}},
{{0.100,0.100,0.900,0.100,0.367,0.580},{0.844}},
{{0.100,0.100,0.900,0.100,0.367,0.687},{0.816}},
{{0.100,0.100,0.900,0.100,0.367,0.793},{0.784}},
{{0.100,0.100,0.900,0.100,0.367,0.900},{0.760}},
{{0.100,0.100,0.900,0.100,0.633,0.100},{0.900}},
{{0.100,0.100,0.900,0.100,0.633,0.153},{0.900}},
{{0.100,0.100,0.900,0.100,0.633,0.260},{0.898}},
{{0.100,0.100,0.900,0.100,0.633,0.367},{0.884}},
{{0.100,0.100,0.900,0.100,0.633,0.473},{0.868}},
{{0.100,0.100,0.900,0.100,0.633,0.580},{0.852}},
{{0.100,0.100,0.900,0.100,0.633,0.687},{0.836}},
{{0.100,0.100,0.900,0.100,0.633,0.793},{0.812}},
{{0.100,0.100,0.900,0.100,0.633,0.900},{0.812}},
{{0.100,0.100,0.900,0.100,0.900,0.100},{0.900}},
{{0.100,0.100,0.900,0.100,0.900,0.153},{0.900}},
{{0.100,0.100,0.900,0.100,0.900,0.260},{0.896}},
{{0.100,0.100,0.900,0.100,0.900,0.367},{0.884}},
{{0.100,0.100,0.900,0.100,0.900,0.473},{0.882}},
{{0.100,0.100,0.900,0.100,0.900,0.580},{0.878}},
{{0.100,0.100,0.900,0.100,0.900,0.687},{0.876}},
{{0.100,0.100,0.900,0.100,0.900,0.793},{0.872}},
```

```
{{0.100,0.100,0.900,0.100,0.900,0.900},{0.852}}};

inNumber=6

hidNumber=3

outNumber=1

outs={0,0,0}

outs[[1]]

outs[[2]]

outs[[3]]

outs=bpnStandard[6,3,1,ioPairsTC,0.1,500];

graph1 = ListPlot[outs[[3]],PlotJoined->True,

      PlotRange->{0.1,0},

PlotJoined->True,Axes->True];

graph2 = ListPlot[{{0,0.01},{500,0.01}},

      PlotStyle->{RGBColor[1,0,0]},

PlotJoined->True,Axes->True]

graph3  = Show[{graph1,graph2}]

Display["c:\\graph.ps",graph3]
```

G.2 IMPLANT-RELATED FEED-FORWARD NEURAL NETWORK

```
<<DiscreteMath`Backpropagation`

ioPairsTC= ioPairs=
```

```
{{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.153},{0.898}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.207},{0.892}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.260},{0.884}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.313},{0.88}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.367},{0.876}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.420},{0.866}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.473},{0.856}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.527},{0.85}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.580},{0.84}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.633},{0.836}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.687},{0.83}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.740},{0.82}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.793},{0.816}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.847},{0.812}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.100,0.900},{0.8}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.100},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.153},{0.8984}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.207},{0.8972}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.260},{0.8928}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.313},{0.89}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.367},{0.884}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.420},{0.88}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.473},{0.876}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.527},{0.868}},
```

```
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.580},{0.852}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.633},{0.844}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.687},{0.832}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.740},{0.816}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.793},{0.806}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.847},{0.802}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.100,0.900},{0.796}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.100},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.153},{0.8992}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.207},{0.896}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.260},{0.894}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.313},{0.8904}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.367},{0.884}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.420},{0.88}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.473},{0.8744}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.527},{0.8664}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.580},{0.86}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.633},{0.844}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.687},{0.84}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.740},{0.8352}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.793},{0.828}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.847},{0.812}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.100,0.900},{0.812}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.153},{0.892}},
```

```
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.207},{0.876}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.260},{0.838}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.313},{0.792}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.367},{0.752}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.100,0.420},{0.7}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.153},{0.898}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.207},{0.892}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.260},{0.884}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.313},{0.872}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.367},{0.86}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.420},{0.84}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.473},{0.82}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.527},{0.804}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.580},{0.7968}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.633},{0.7968}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.687},{0.7968}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.740},{0.7944}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.793},{0.78}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.847},{0.7704}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.900},{0.7704}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.153},{0.8984}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.207},{0.8944}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.260},{0.892}},
```

```
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.313},{0.8888}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.367},{0.8816}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.420},{0.8776}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.473},{0.8696}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.527},{0.8584}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.580},{0.84}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.633},{0.836}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.687},{0.828}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.740},{0.82}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.793},{0.8136}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.847},{0.8096}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.900},{0.7992}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.153},{0.8992}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.207},{0.8936}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.260},{0.888}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.313},{0.88}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.367},{0.872}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.420},{0.856}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.473},{0.848}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.527},{0.832}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.580},{0.8192}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.633},{0.804}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.687},{0.792}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.740},{0.7672}},
```

```
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.153},{0.8992}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.207},{0.8984}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.260},{0.896}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.313},{0.8904}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.367},{0.884}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.420},{0.88}},
{{0.900,0.100,0.100,0.100,0.100,0.100,0.900,0.473},{0.8752}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.153},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.207},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.260},{0.8976}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.313},{0.892}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.367},{0.884}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.420},{0.8736}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.473},{0.868}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.527},{0.852}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.100},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.153},{0.9}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.207},{0.8976}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.260},{0.8952}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.313},{0.892}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.367},{0.8896}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.420},{0.888}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.473},{0.8816}},
```

```
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.527},{0.88}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.580},{0.88}},
{{0.900,0.900,0.100,0.100,0.100,0.100,0.900,0.633},{0.88}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.153},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.207},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.260},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.313},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.367},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.420},{0.8936}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.473},{0.8936}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.527},{0.8936}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.580},{0.8936}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.153},{0.8984}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.207},{0.896}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.260},{0.892}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.313},{0.8896}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.367},{0.8888}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.420},{0.8776}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.473},{0.86}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.527},{0.8512}},
{{0.900,0.100,0.100,0.900,0.100,0.900,0.900,0.580},{0.8136}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.100},{0.9}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.153},{0.8968}},
```

```
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.207},{0.892}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.260},{0.88}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.313},{0.8784}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.367},{0.872}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.420},{0.8616}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.473},{0.8504}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.527},{0.8104}},
{{0.100,0.900,0.100,0.900,0.900,0.900,0.900,0.580},{0.804}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.100},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.153},{0.9}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.207},{0.8984}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.260},{0.8968}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.313},{0.8952}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.367},{0.8928}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.420},{0.8912}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.473},{0.888}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.527},{0.884}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.580},{0.8904}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.633},{0.876}},
{{0.900,0.100,0.100,0.100,0.100,0.900,0.900,0.687},{0.876}}};
inNumber=8
hidNumber=2
outNumber=1
outs={0,0,0}
outs[[1]]
```

```
outs[[2]]

outs[[3]]

outs=bpnStandard[8,2,1,ioPairsTC,0.1,2000];

graph1 = ListPlot[outs[[3]],PlotJoined->True,
      PlotRange->{0.1,0},Frame->False,
      PlotJoined->True,Axes->True];

graph2 = ListPlot[{{0,0.01},{2000,0.01}},
      PlotStyle->{RGBColor[1,0,0]},Frame->False,
      PlotJoined->True,Axes->True]

graph3 = Show[{graph1,graph2}]

Display["c:\\graph.ps",graph3]
```

G.3 BACKPROPAGATION ALGORITHMS

```
BeginPackage["Backpropagation`"]


sigmoid::usage = "sigmoid[x,opts___Rule]"

bpnTest::usage = "bpnTest[hiddenWts,outputWts,
      ioPairVectors,opts___Rule]"

bpnStandard::usage = "bpnStandard[inNumber, hidNumber,
      outNumber,ioPairs, eta, numIters]"

bpnMomentum::usage = "bpnMomentum[inNumber,hidNumber,
      outNumber,ioPairs,eta, alpha,numIters]"
```

```
Options[sigmoid] = xShift->0,yShift->0,temperature->1;

Options[bpnTest] = printAll->False,bias->False;


Begin["`Private`"] (* begin the private context *)


sigmoid[x_,opts___Rule] := Module[xshft,yshft,temp,
      xshft = xShift /. opts /. Options[sigmoid];
      yshft = yShift /. opts /. Options[sigmoid];
      temp = temperature /. opts /. Options[sigmoid];
      yshft+1/(1+E^(-(x-xshft)/temp)) //N
      ]


bpnTest[hiddenWts_,outputWts_,ioPairVectors_,opts___Rule] :=
Module[inputs,hidden,outputs,desired,errors,i,len,
      prntAll,errorTotal,errorSum,biasVal,
      prntAll= printAll /. opts /. Options[bpnTest];
      biasVal = bias /. opts /. Options[bpnTest];
      inputs=Map[First,ioPairVectors];
      len=Length[inputs];
      If[biasVal,inputs=Map[Append[#,1.0]&,inputs] ];
      desired=Map[Last,ioPairVectors];
      hidden=sigmoid[inputs.Transpose[hiddenWts]];
      If[biasVal,hidden = Map[Append[#,1.0]&,hidden] ];
      outputs=sigmoid[hidden.Transpose[outputWts]];
```

```
      errors= desired-outputs;
      If[prntAll,Print["ioPairs:"];Print[ ];
      Print[ioPairVectors];
      Print[ ];Print["inputs:"];Print[ ];Print[inputs];
      Print[ ];Print["hidden-layer outputs:"];
      Print[hidden];Print[ ];
      Print["output-layer outputs:"];Print[ ];
      Print[outputs];Print[ ];Print["errors:"];
      Print[errors];Print[ ];  ]; (* end of If *)
      For[i=1,i<=len,i++,Print[" Output ",i," = ",
            outputs[[i]]," desired = ",desired[[i]],
            " Error = ",errors[[i]]];Print[];  ];
(* end of For *)
errorSum = Apply[Plus,errors^2,2];
      (* second level *)
errorTotal = Apply[Plus,errorSum];
Print["Mean Squared Error = ",errorTotal/len];
] (* end of Module *)

bpnStandard[inNumber_, hidNumber_, outNumber_,
ioPairs_, eta_, numIters_] :=
Module[errors,hidWts,outWts,ioP,inputs,outDesired,hidOuts,
outputs, outErrors,outDelta,hidDelta,
hidWts = Table[Table[Random[Real,-0.1,0.1],
inNumber],hidNumber];
```

```
outWts = Table[Table[Random[Real,-0.1,0.1],
hidNumber],outNumber];
errors = Table[
(* select ioPair *)
ioP=ioPairs[[Random[Integer,1,Length[ioPairs]]]];
inputs=ioP[[1]];
outDesired=ioP[[2]];
(* forward pass *)
hidOuts = sigmoid[hidWts.inputs];
outputs = sigmoid[outWts.hidOuts];
(* determine errors and deltas *)
outErrors = outDesired-outputs;
outDelta= outErrors (outputs (1-outputs));
hidDelta=(hidOuts (1-hidOuts)) Transpose[outWts].
outDelta;
(* update weights *)
outWts += eta Outer[Times,outDelta,hidOuts];
hidWts += eta Outer[Times,hidDelta,inputs];
(* add squared error to Table *)
outErrors.outErrors,numIters];  (* end of Table *)
Return[hidWts,outWts,errors];
]; (* end of Module *)

bpnMomentum[inNumber_,hidNumber_,outNumber_,ioPairs_,eta_,
alpha_,numIters_] :=
```

```
Module[hidWts,outWts,ioP,inputs,hidOuts,outputs,outDesired,
hidLastDelta,outLastDelta,outDelta,hidDelta,outErrors,
hidWts = Table[Table[Random[Real,-0.5,0.5],
inNumber],hidNumber];
outWts = Table[Table[Random[Real,-0.5,0.5],
hidNumber],outNumber];
hidLastDelta = Table[Table[0,inNumber],hidNumber];
outLastDelta = Table[Table[0,hidNumber],outNumber];
errorList = Table[
(* begin forward pass *)
ioP=ioPairs[[Random[Integer,1,Length[ioPairs]]]];
inputs=ioP[[1]];
outDesired=ioP[[2]];
hidOuts = sigmoid[hidWts.inputs];
(* hidden-layer outputs *)
outputs = sigmoid[outWts.hidOuts];
(* output-layer outputs *)
(* calculate errors *)
outErrors = outDesired-outputs;
outDelta= outErrors (outputs (1-outputs));
hidDelta=(hidOuts (1-hidOuts)) Transpose[outWts].
outDelta; (* update weights *)
outLastDelta= eta Outer[Times,outDelta,hidOuts]
+alpha outLastDelta;
outWts += outLastDelta;
```

```
hidLastDelta = eta Outer[Times,hidDelta,inputs]+
alpha hidLastDelta;
hidWts += hidLastDelta;
outErrors.outErrors, numIters] ;
(* this puts the error on the list *)
(* this many times, Table ends here *)
Print["New hidden-layer weight matrix: "];
Print[]; Print[hidWts];Print[];
Print["New output-layer weight matrix: "];
Print[]; Print[outWts];Print[];
bpnTest[hidWts,outWts,ioPairs,bias->
False,printAll->False];
errorPlot = ListPlot[errorList, PlotJoined->True];
Return[hidWts,outWts,errorList,errorPlot];
] (* end of Module *)

End[] (* end the private context *)

EndPackage[] (* end the package context *)
```

Appendix H

Survivability Graphs

This appendix contains the patient and implant-related survivability data extracted from graphs published in a series of research papers on the Swedish Hip Register [1, 82, 83]. These graphs show the effect of selected variables on survivability. Data on patient and implant factors are available separately, but not in combination.

H.1 24 PATIENT-RELATED SURVIVABILITY GRAPHS

The results from these graphs show significant differences of revision rate for aseptic loosening due to patient age, gender and diagnosis. The survival analysis illustrated here is based on revision as endpoint for failure.

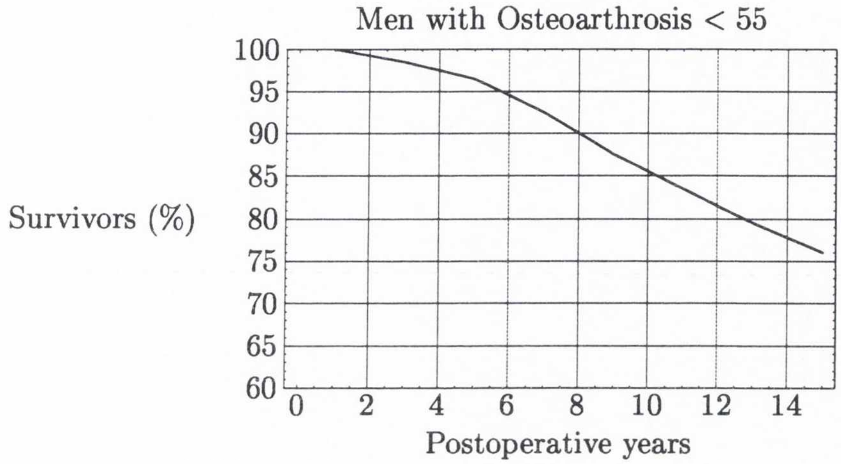


Figure H.1: Survival rate for men with the primary diagnosis osteoarthrosis in the age group less than fifty five.

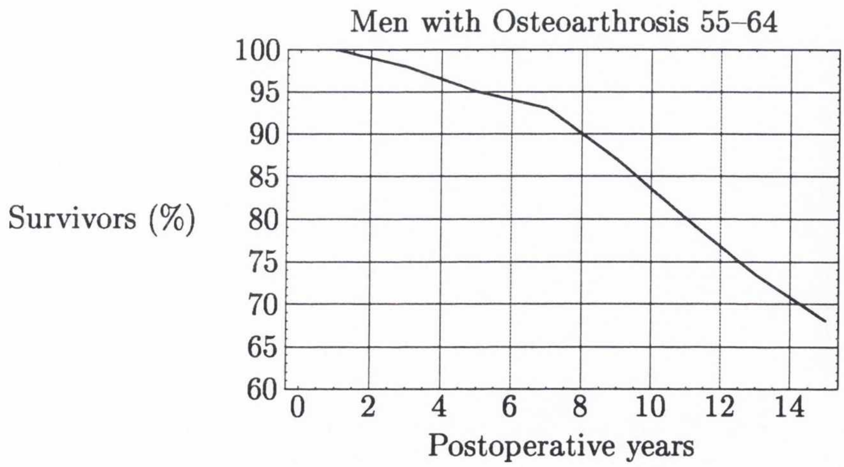


Figure H.2: Survival rate for men with the primary diagnosis osteoarthrosis in the age group fifty five to sixty four.

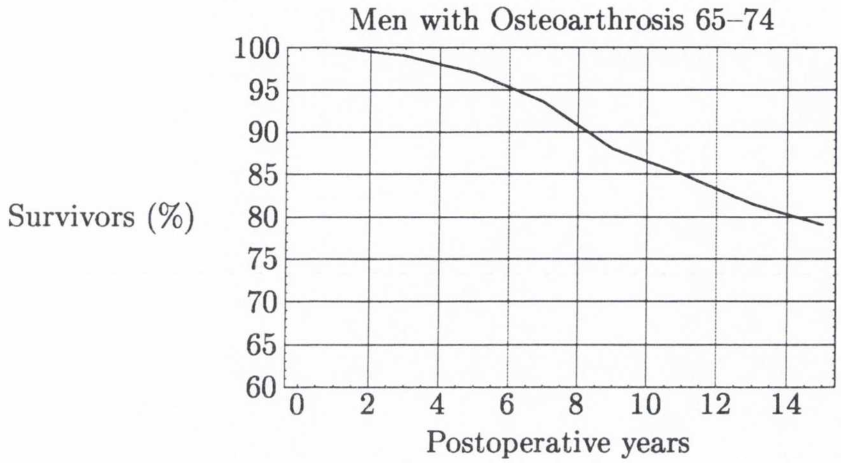


Figure H.3: Survival rate for men with the primary diagnosis osteoarthrosis in the age group sixty five to seventy four.

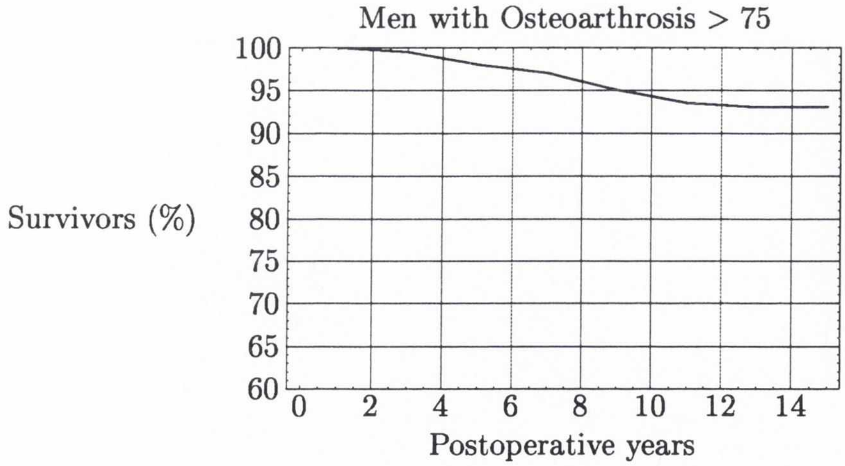


Figure H.4: Survival rate for men with the primary diagnosis osteoarthrosis in the age group greater than seventy five.

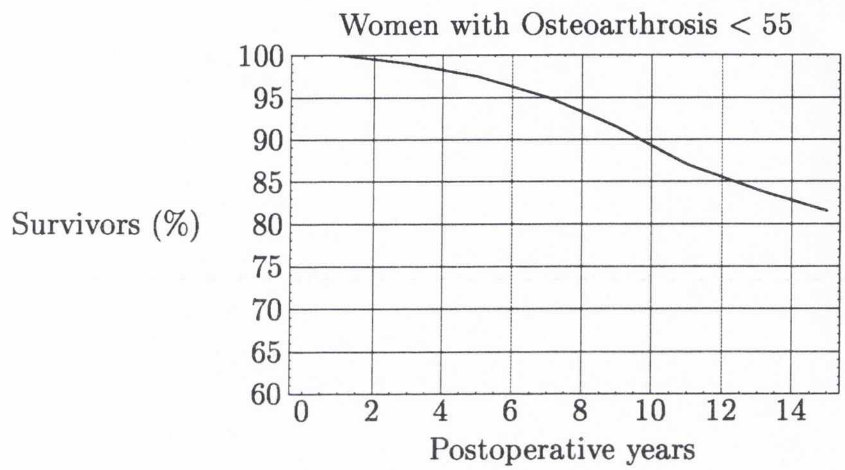


Figure H.5: Survival rate for women with the primary diagnosis osteoarthrosis in the age group less than fifty five.

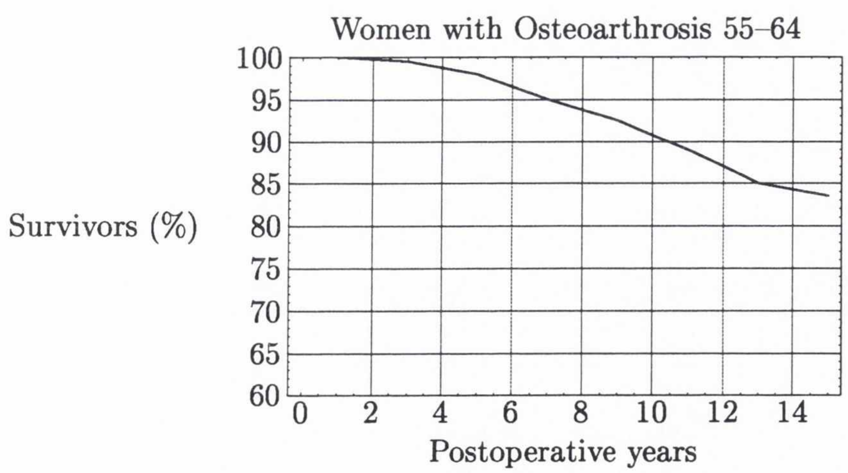


Figure H.6: Survival rate for women with the primary diagnosis osteoarthrosis in the age group fifty five to sixty four.

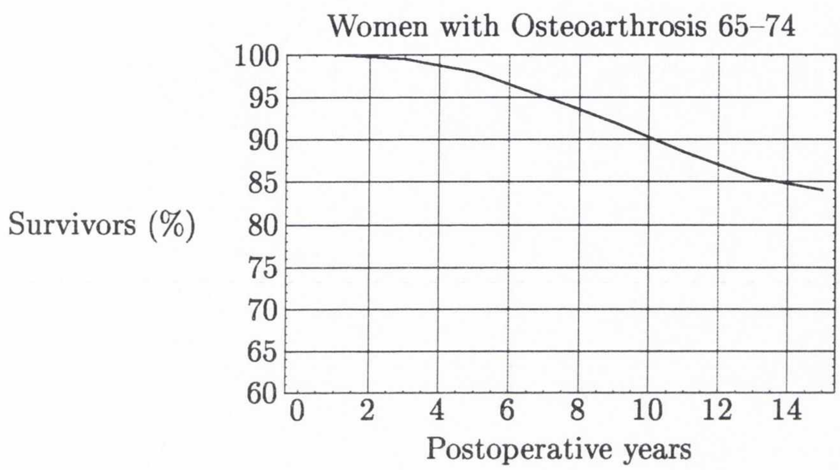


Figure H.7: Survival rate for women with the primary diagnosis osteoarthrosis in the age group sixty five to seventy four.

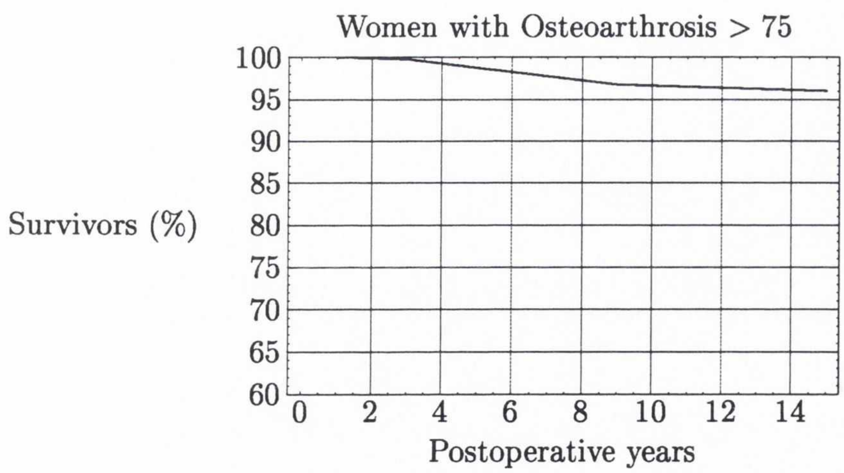


Figure H.8: Survival rate for women with the primary diagnosis osteoarthrosis in the age group greater than seventy five.

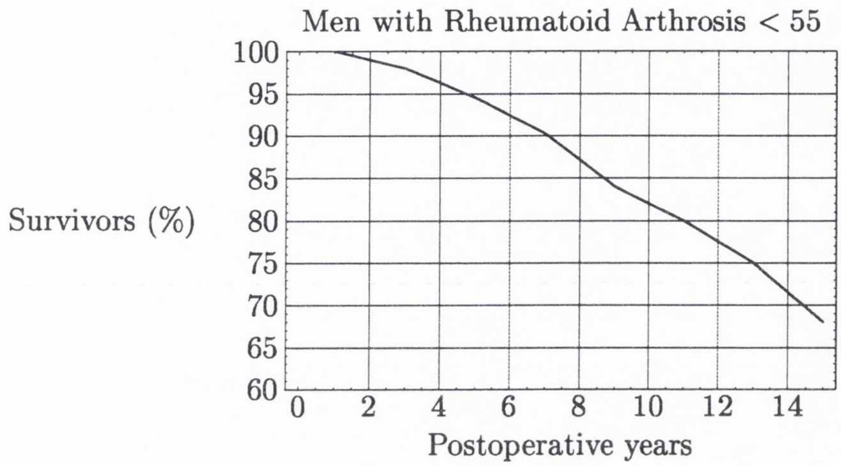


Figure H.9: Survival rate for men with the primary diagnosis rheumatoid arthrosis in the age group less than fifty five.

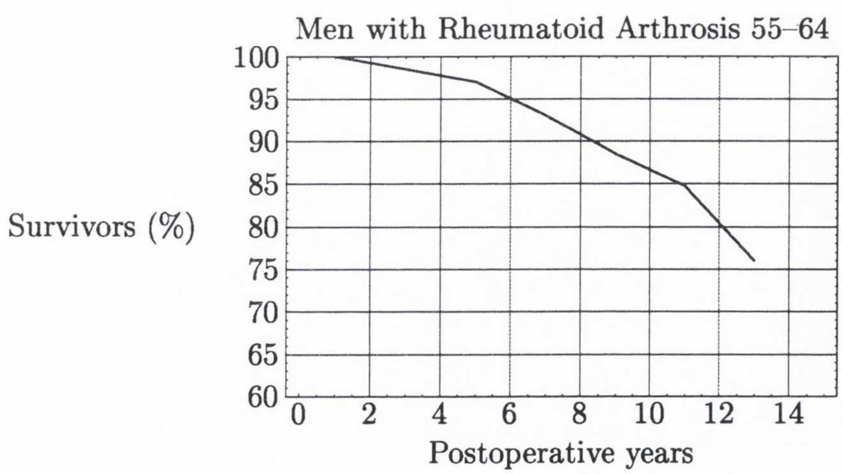


Figure H.10: Survival rate for men with the primary diagnosis rheumatoid arthrosis in the age group fifty five to sixty four.

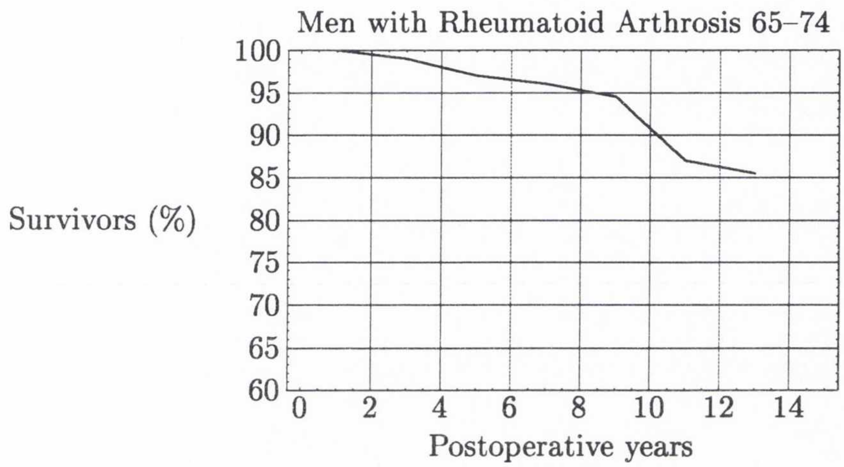


Figure H.11: Survival rate for men with the primary diagnosis rheumatoid arthrosis in the age group sixty five to seventy four.

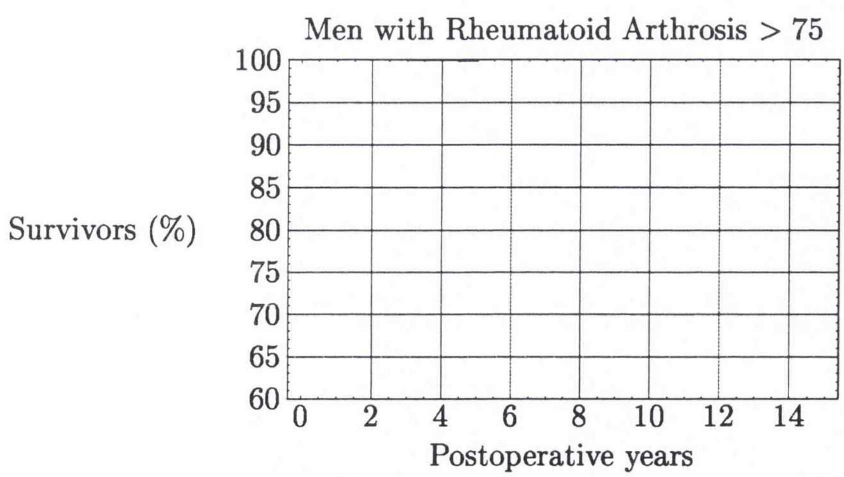


Figure H.12: Survival rate for men with the primary diagnosis rheumatoid arthrosis in the age group greater than seventy five.

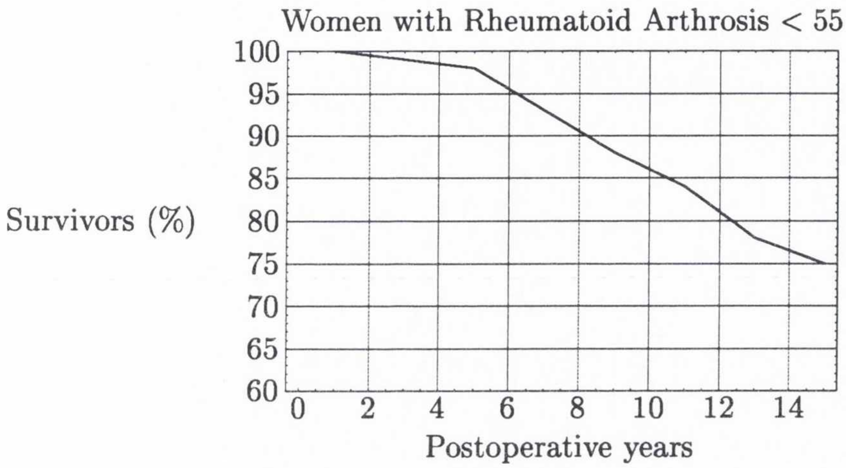


Figure H.13: Survival rate for women with the primary diagnosis rheumatoid arthrosis in the age group less than fifty five.

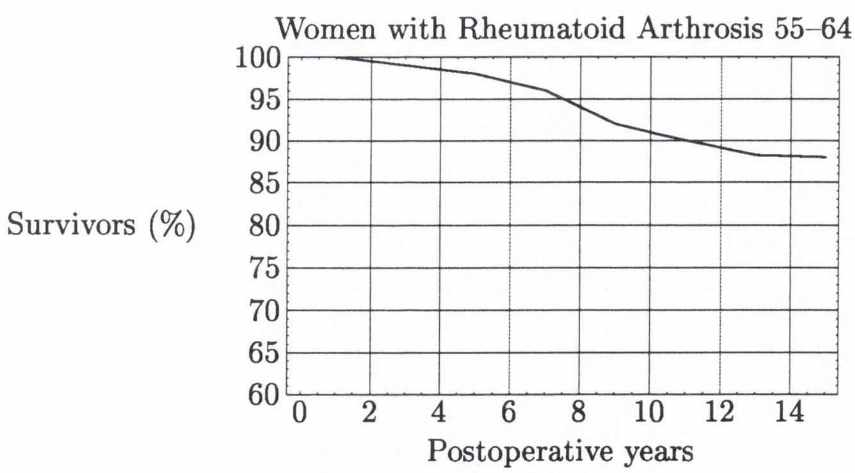


Figure H.14: Survival rate for women with the primary diagnosis rheumatoid arthrosis in the age group fifty five to sixty four.

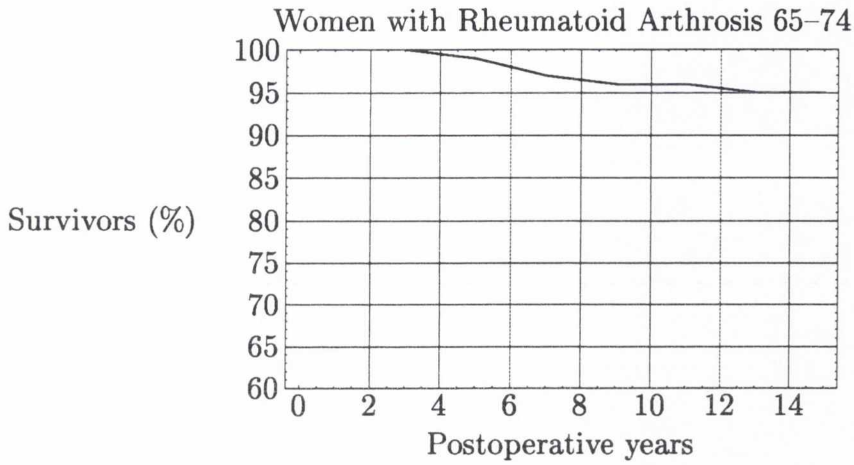


Figure H.15: Survival rate for women with the primary diagnosis rheumatoid arthrosis in the age group sixty five to seventy four.

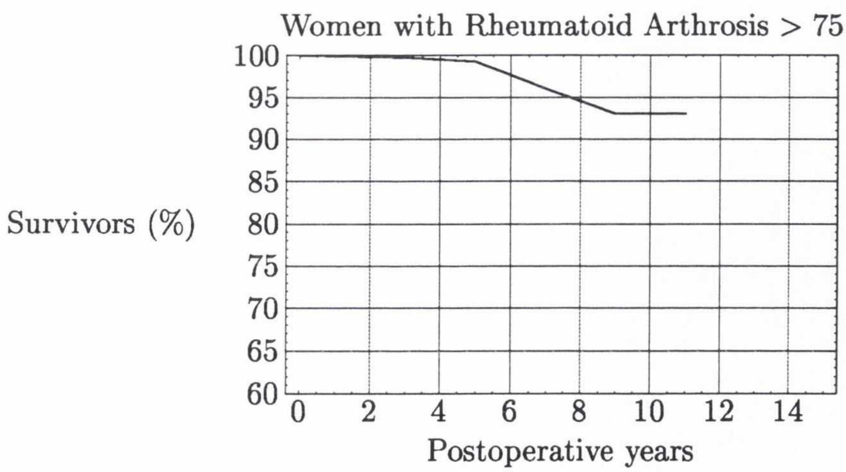


Figure H.16: Survival rate for women with the primary diagnosis rheumatoid arthrosis in the age group greater than seventy five.

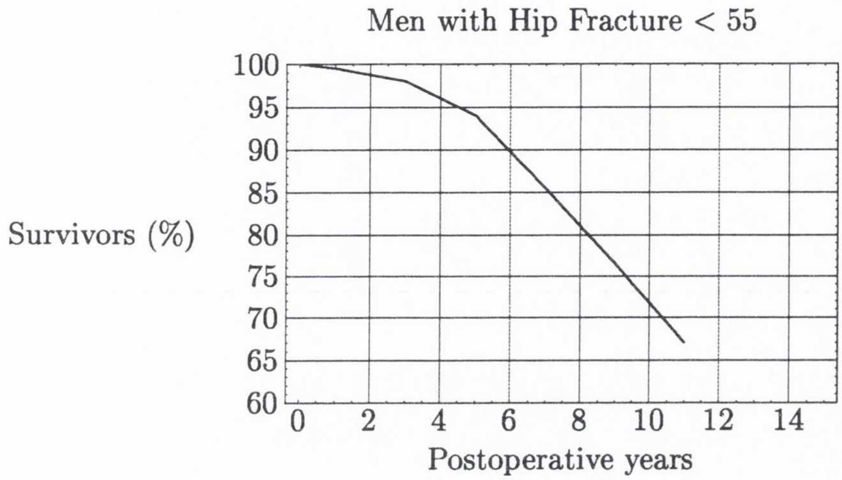


Figure H.17: Survival rate for men with the primary diagnosis hip fracture in the age group less than fifty five.

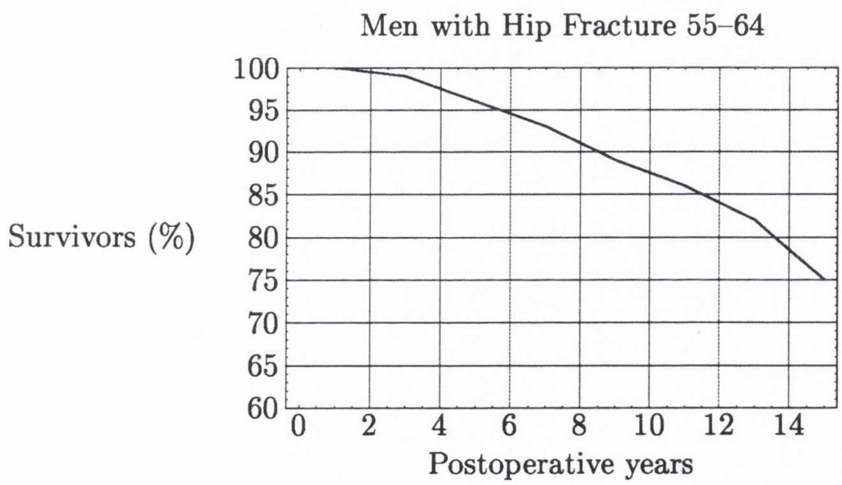


Figure H.18: Survival rate for men with the primary diagnosis hip fracture in the age group fifty five to sixty four.

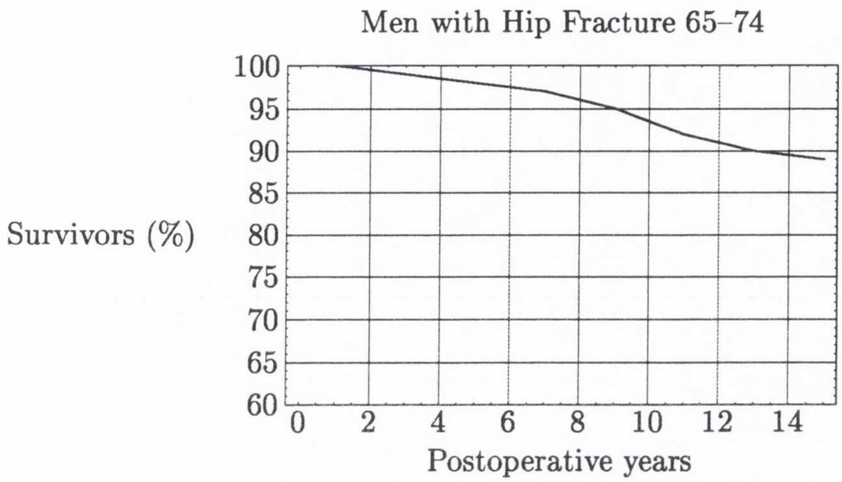


Figure H.19: Survival rate for men with the primary diagnosis hip fracture in the age group sixty five to seventy four.

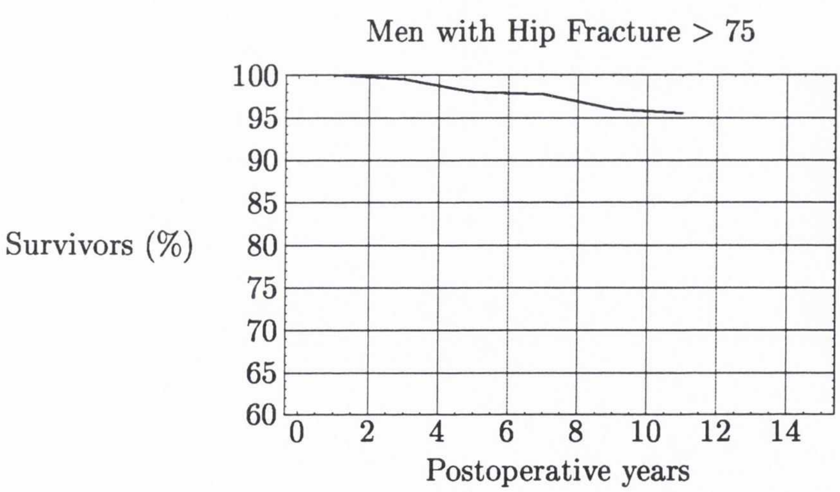


Figure H.20: Survival rate for men with the primary diagnosis hip fracture in the age group greater than seventy five.

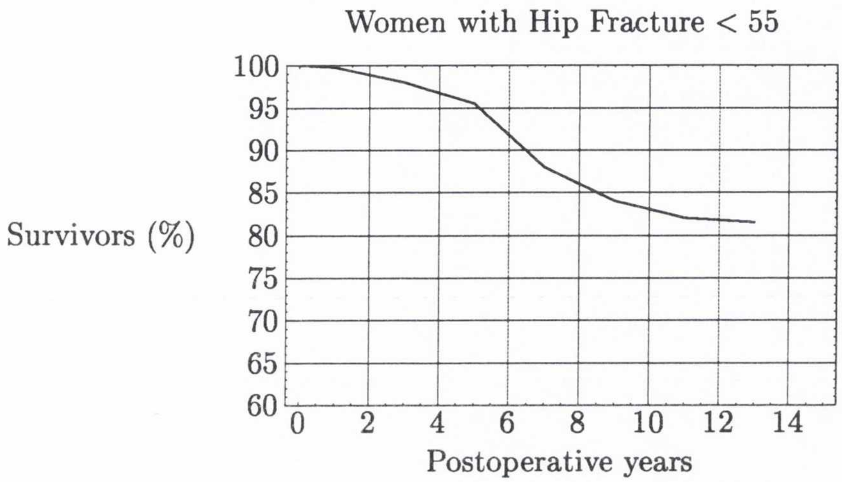


Figure H.21: Survival rate for women with the primary diagnosis hip fracture in the age group less than fifty five.

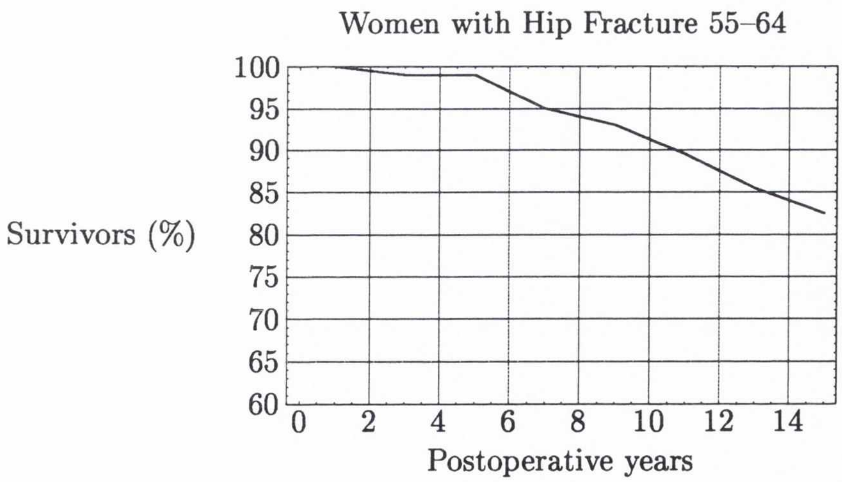


Figure H.22: Survival rate for women with the primary diagnosis hip fracture in the age group fifty five to sixty four.

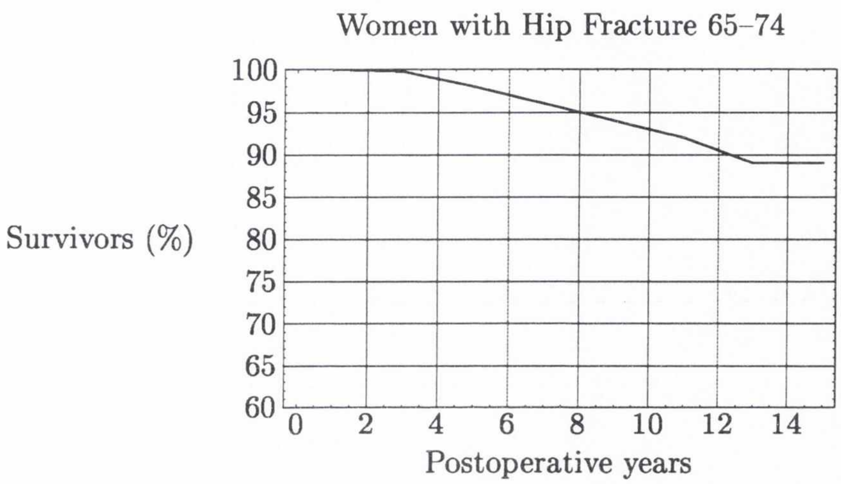


Figure H.23: Survival rate for women with the primary diagnosis hip fracture in the age group sixty five to seventy four.

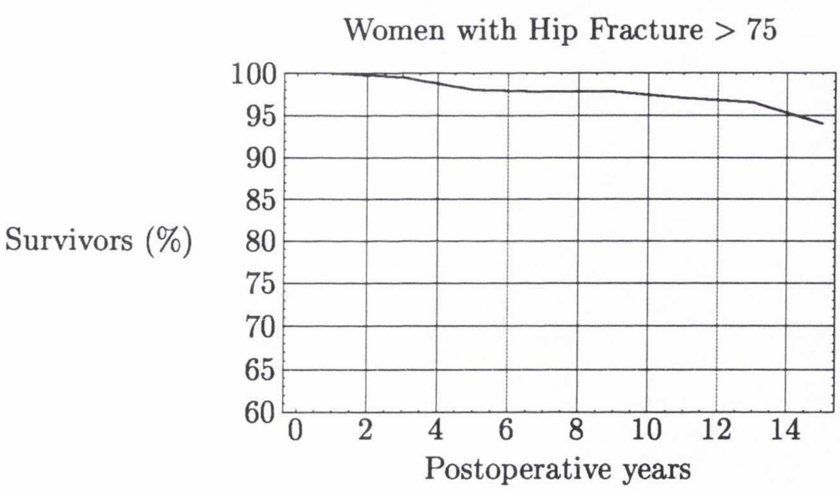


Figure H.24: Survival rate for women with the primary diagnosis hip fracture in the age group greater than seventy five.

H.2 10 IMPLANT-RELATED SURVIVABILITY GRAPHS

The implant survivorship analysis describes the probability of failure due to revision for aseptic loosening. None of the curves are depicted when less than 50 hips remain at risk.

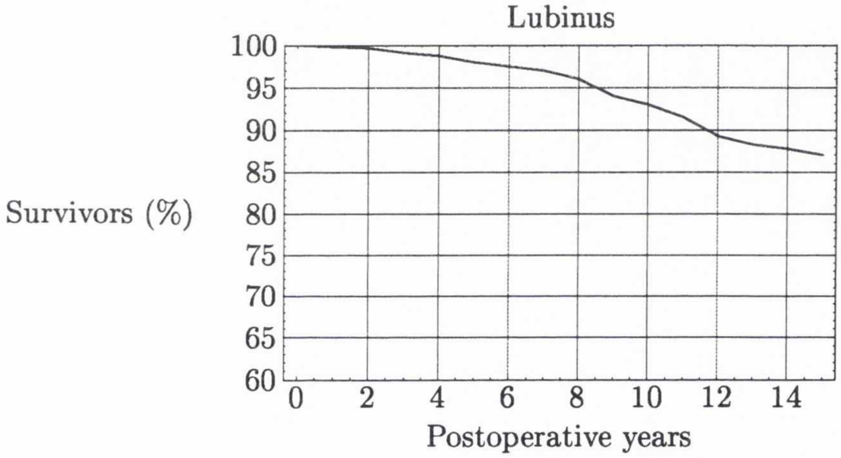


Figure H.25: Survival rate for aseptic loosening using the Lubinus prosthesis.

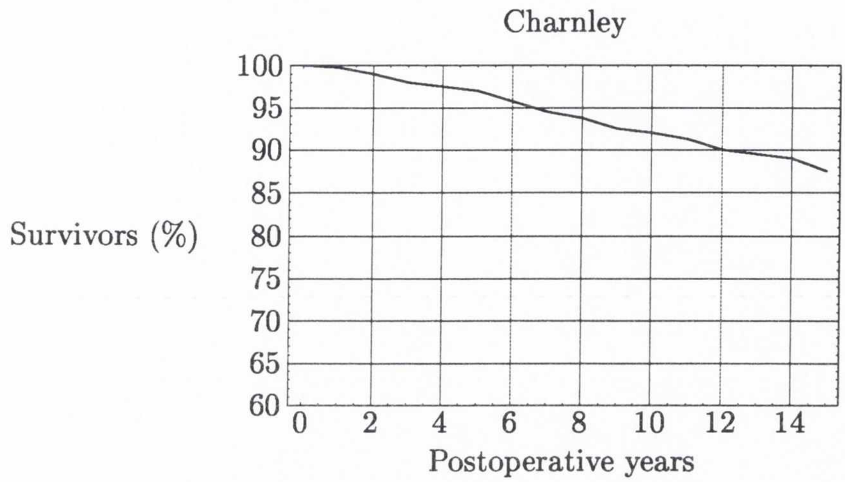


Figure H.26: Survival rate for aseptic loosening using the Charnley prosthesis.

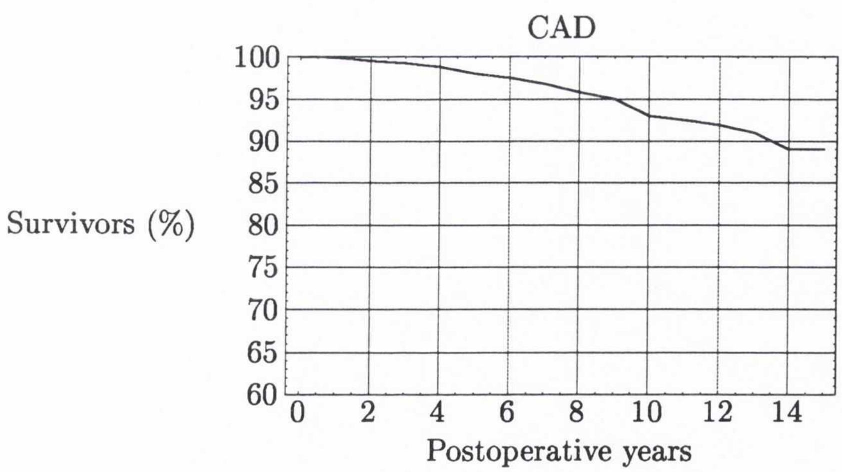


Figure H.27: Survival rate for aseptic loosening using the CAD prosthesis.

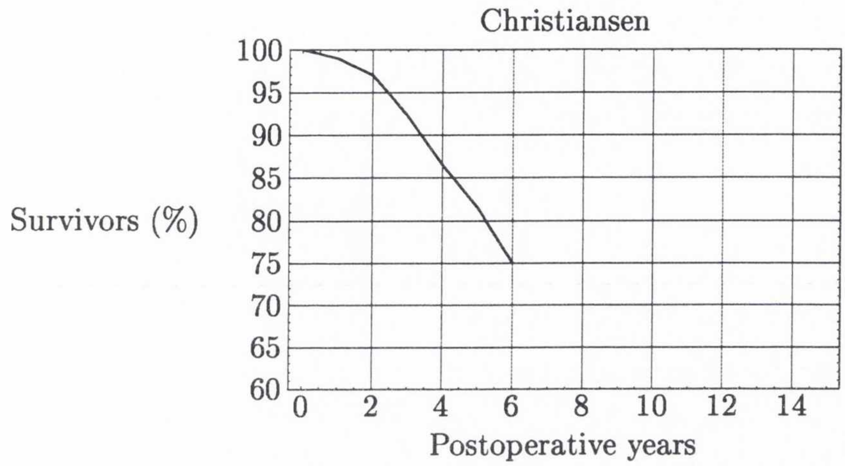


Figure H.28: Survival rate for aseptic loosening using the Christiansen prosthesis.

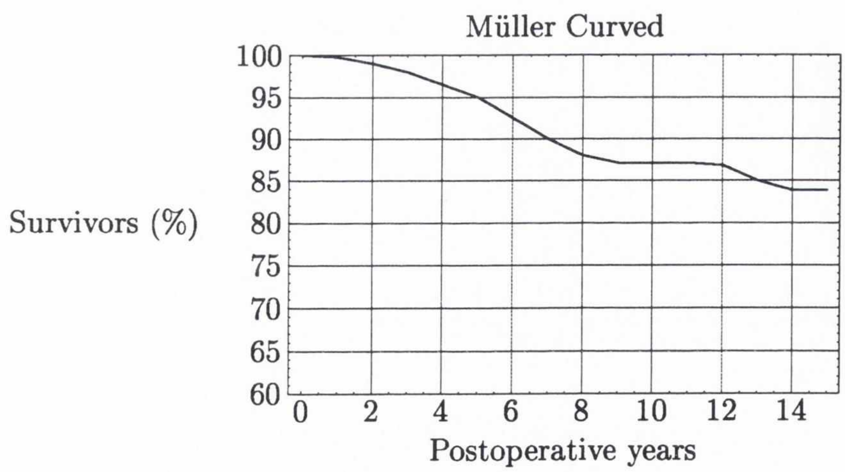


Figure H.29: Survival rate for aseptic loosening using the Muller Curved prosthesis.

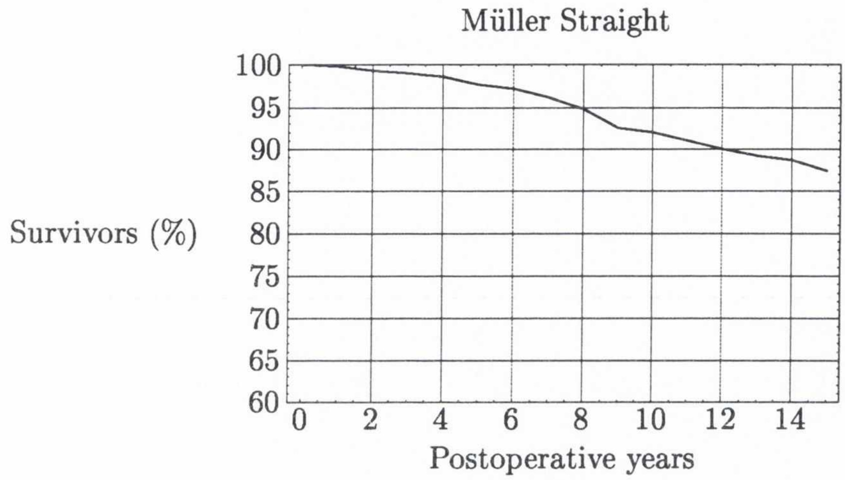


Figure H.30: Survival rate for aseptic loosening using the Muller Straight prosthesis.

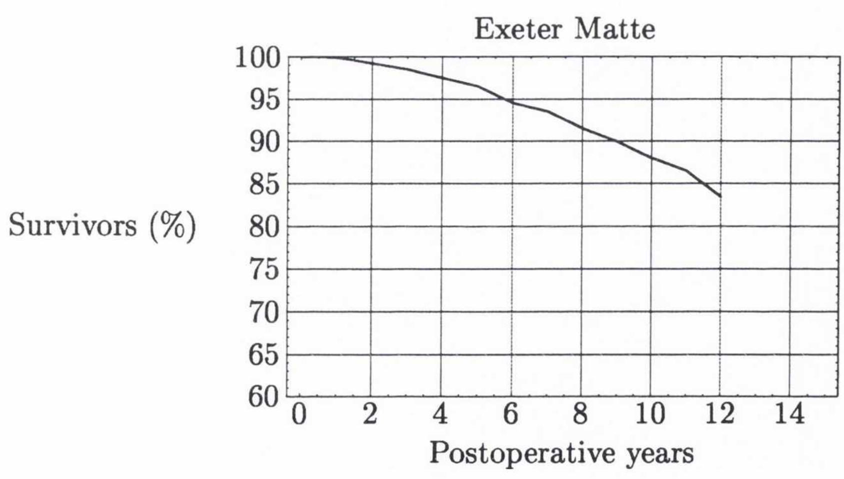


Figure H.31: Survival rate for aseptic loosening using the Exeter Matte prosthesis.

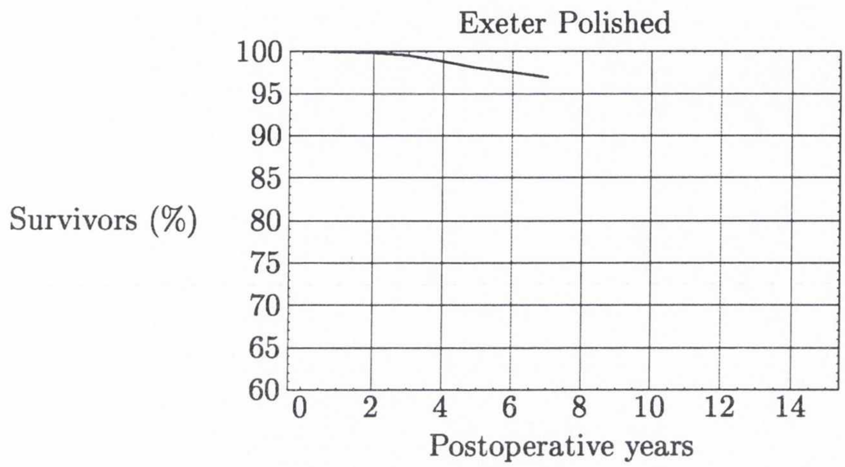


Figure H.32: Survival rate for aseptic loosening using the Exter Polished prosthesis.

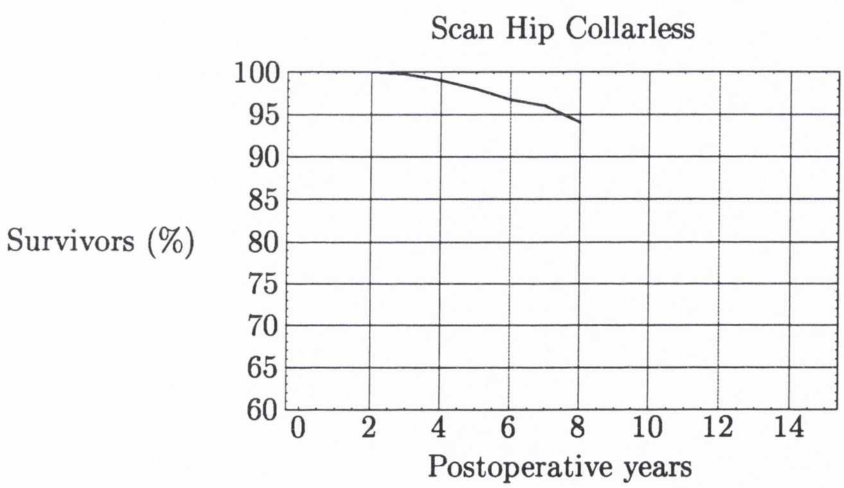


Figure H.33: Survival rate for aseptic loosening using the Scan Hip Collarless prosthesis.

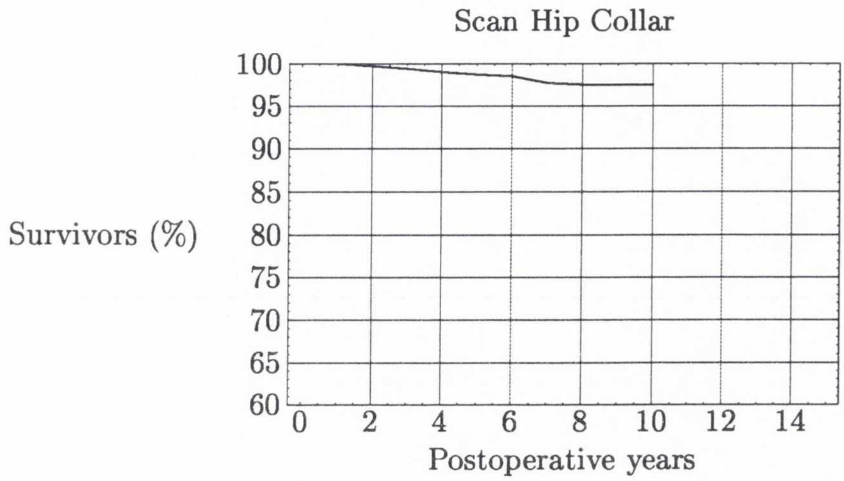


Figure H.34: Scan Hip Collar

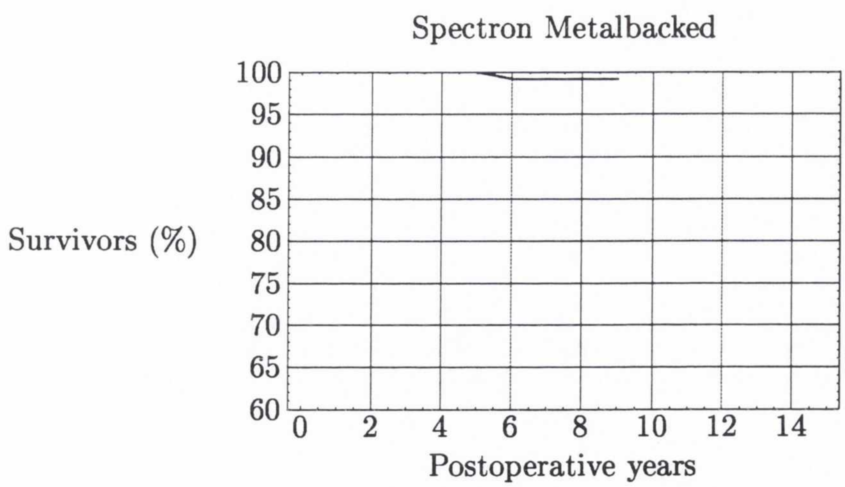


Figure H.35: Survival rate for aseptic loosening using the Spectron Metalbacked prosthesis.

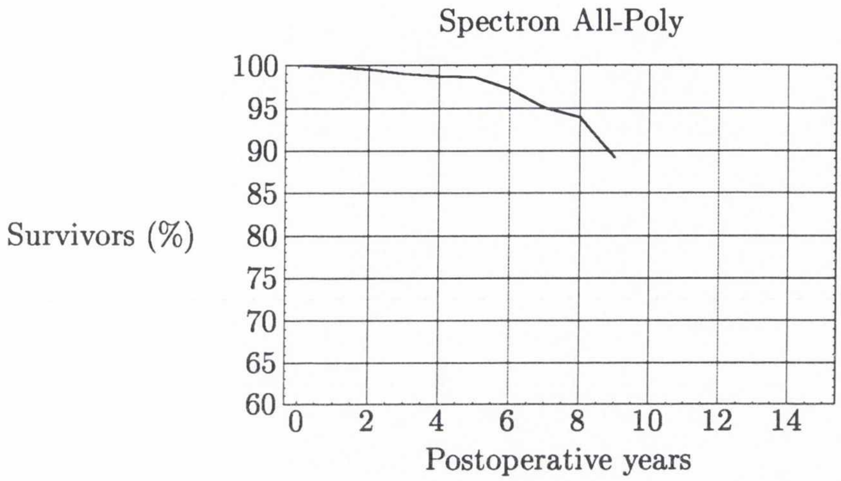


Figure H.36: Survival rate for aseptic loosening using the Spectron All-Poly prosthesis.

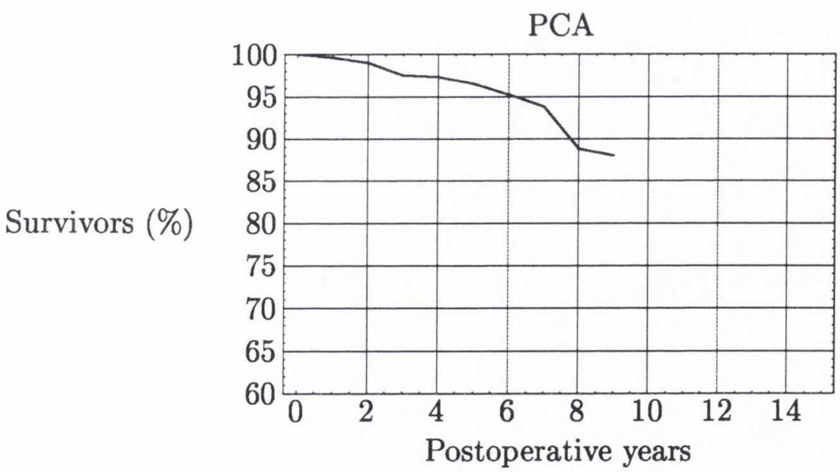


Figure H.37: PCA

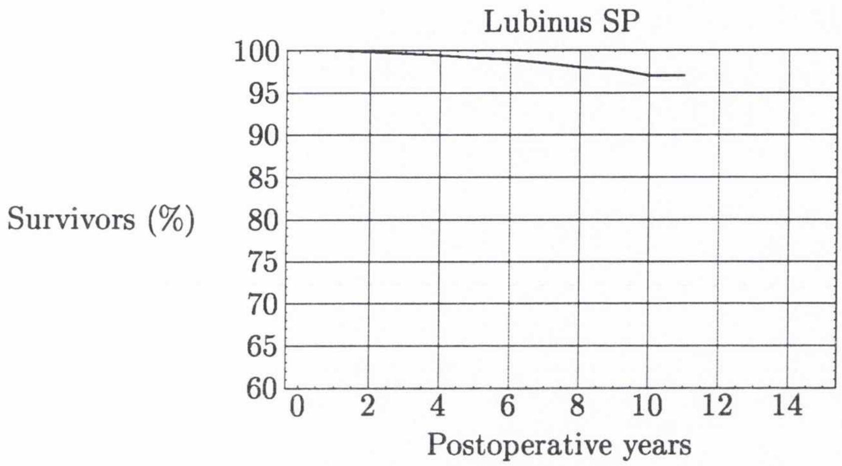


Figure H.38: Survival rate for aseptic loosening using the Lubinus SP prosthesis.

Bibliography

[1] L. Ahnfelt, P. Herberts, H. Malchau, and G. Andersson. Prognosis of total hip replacement: A Swedish multicenter study of 4,664 revisions. *Acta Orthopaedica Scandinavica Supplementum*, 61(238):1–26, 1990.

[2] T.P. Andriacchi and R.P. Mikosz. *Basic orthopaedic biomechanics*, chapter Musculoskeletal dynamics, locomotion and clinical application, pages 15–92. Raven Press, New York, 1991.

[3] Harlan Armstutz, Sam Nasser, and Michael Kabo. Preliminary results of an off-the-shelf press-fit stem, the anthropometric total hip femoral component using exact-fit principles. *Clinical Orthopaedics and Related Research*, (249):73–78, December 1989.

[4] D.H. Ballard and C.M. Brown. *Computer Vision.* Prentice-Hall, Englewood Cliffs, NJ., 1982.

[5] William L. Bargar. Shape the implant to the patient, a rationale for the use of custom-fit cement-less total hip implants. *Clinical Orthopaedics and Related Research*, (249):73–78, December 1989.

[6] Edward Bender. *Mathematical Methods in Artificial Intelligence.* IEEE Press, 1996. ISBN 0–8186–7200–5.

[7] Edward Blum and Leong Li. Approximation theory and feed-forward networks. *Neural Networks*, 4:511–515, 1991.

[8] J. Bobyn, E. Mortimer, A. Glassman, C. Engh, J. Miller, and E. Brooks. Producing and avoiding stress shielding. *Clinical Orthopaedics and Related Research*, (274):79–96, January 1992.

[9] W. Brien, E. Salvati, F. Betts, P. Bullough, T. Wright, C. Rimnac, R. Buly, , and K. Garvin. Metal levels in cemented total hip arthroplasty. *Clinical Orthopaedics and Related Research*, (276):66–74, March 1992.

[10] Christopher Bulstrode. Total hip replacement: The way forward. *Ann R Coll Surg Engl*, 78:129–132, 1996.

[11] Alan Burstein. Podium presentation. In *Annual Meeting of the Hip Society*, 1992.

[12] J. Callaghan, S. Dysart, and C. Savory. The uncemented porous-coated anatomic total hip prosthesis. *The Journal of Bone and Joint Surgery*, 70-A:337–346, 1988.

[13] John J. Callaghan. Total hip arthroplasty, clinical perspective. *Clinical Orthopaedics and Related Research*, (276):33–40, March 1992.

[14] A. Campell, C. Rorabeck, R. Bourne, D. Chess, and L. Nott. Thigh pain after cement-less hip arthroplasty. *British Journal of Bone and Joint Surgery*, 74-B(1):63–66, January 1992.

[15] W. Capello. Fit the patient to the prosthesis. *Clinical Orthopaedics and Related Research*, (249):56–59, December 1989.

[16] H. Chandler, T. Reineck, R. Wixson, and J. McCarthy. Total hip replacement in patients younger than thirty years old, a five-year follow-up study. *The Journal of Bone and Joint Surgery*, (63-A):1426–1434, December 1981.

[17] J. Charnley. Long term results of low friction arthroplasty of the hip performed as a primary intervention. *British Journal of Bone and Joint Surgery*, 54-B:61, 1972.

[18] John Charnley. *Lancet*, 1961.

[19] Clocksin and Mellish. *Programming in Prolog*. Springer-Verlag, 1981.

[20] J.A. Cogan and P.J. Prendergast. Preliminary investigation of knowledge engineering to optimise hip prosthesis design. In *IVth International Symposium on Computer Simulation in Biomechanics*, Paris, 1993. Ecole Nationale Supérieure d'Arts et Métiers.

[21] D. Collins. Cemented total hip replacement in patients who are less than fifty years old. *The Journal of Bone and Joint Surgery*, 66-A:353–359, January 1984.

[22] Delma E. Cowley. Prostheses for primary total hip replacement. *International Journal of Technology Assessment in Health Care*, 11(4):770–778, 1995.

[23] Earl Cox. *The Fuzzy Systems Handbook*. AP Professional, 1994. ISBN 0–12–194270–8.

[24] Merle R. D'Aubigne and M. Postel. Functional results of hip arthroplasty with acrylic prosthesis. *The Journal of Bone and Joint Surgery*, 36-A:451–475, 1954.

[25] J. DeLee and J. Charnley. Radiological demarcation of cemented sockets in total hip replacements. *Clinical Orthopaedics and Related Research*, (121):20–32, 1976.

[26] James K. DeOrio and Kurt E. Blasser. *Joint Replacement Arthroplasty*, chapter Indications and patient selection, pages 547–559. Churchill Livingstone, 1990.

[27] H.S. Dobbs. Survival of total hip replacement. *British Journal of Bone and Joint Surgery*, 62-B:168–173, 1980.

[28] Bruce R. Donald. A search algorithm for motion planning with six degrees of freedom. *Artificial Intelligence*, (31):295–353, 1987.

[29] R.L. Dooley, C. Engel, and M.E. Muller. Automated scanning and digitizing of roentgenographs for documentation and research. *Clinical Orthopaedics and Related Research*, 274:113–9, 1992.

[30] F. Dorey and H.C. Amstutz. The validity of survivorship analysis in total joint arthroplasty. *The Journal of Bone and Joint Surgery*, 71-A:544–548, 1989.

[31] Drucker. *Post Capitalist Society.* Harper-Collins, 1993.

[32] A. Edidin. *Modelling of Bone Material Properties From CT Scans-Considerations in Bio-mechanical Structural Models.* Phd thesis, Cornell University, Ithaca, New York, May 1991.

[33] D. Engelbrecht, F. Weber, M. Sweet, and I. Jakim. Long term results of revision total hip arthroplasty. *British Journal of Bone and Joint Surgery*, 72-B(1):41–45, January 1990.

[34] C. Engh, A. Glassman, W. Griffin, and J. Mayer. Results of cementless revision for failed cemented total hip arthroplasty. *Clinical Orthopaedics and Related Research*, (235):91–110, October 1988.

[35] C.A. Engh, T.F. McGovern, J.D. Bobyn, and W.H. Harris. A quantitative evaluation of periprosthetic bone-remodelling after cement-less total hip arthroplasty. *The Journal of Bone and Joint Surgery*, A-74:1009–1020, 1992.

[36] J. Essinger. A contour-less hip designing process. In *Proceeding Fifth Annual International Symposium on Custom Prostheses*, October 1–3 1992.

[37] European Commission. *Building the Information Society for Us All*, 1996. Interim Report of European Commission.

[38] W. Fakhr, M. Kamel, and M.I. Elmasry. Probability of error, maximum mutual information, and size minimization of neural networks. In *International Joint Conference on Neural Networks*, Baltimore, MD., 1992.

[39] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. *Advances in Knowledge Discovery and Data Mining*, chapter From Data Mining to Knowledge Discovery: An Overview, pages 1–34. MIT Press, Cambridge, Mass., 1996. ISBN 0–262–56097–6.

[40] James A. Freeman. *Simulating Neural Networks with Mathematica*. Addison-Wesley, 1994. ISBN 0–201–56629–X.

[41] James A. Freeman and David M. Skapura. *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley, Reading MA, 1989.

[42] T. Gruen, G. McNeice, and H. Amstutz. Modes of failure of cemented stem-type femoral components. *Clinical Orthopaedics and Related Research*, (141):17–27, June 1979.

[43] R. Gustilo and H. Pasternak. Revision total hip arthroplasty with titanium in-growth prosthesis and bone grafting for failed cemented femoral component loosening. *Clinical Orthopaedics and Related Research*, (235):111–119, October 1988.

[44] R. Haddad, S. Cook, and M. Brinker. A comparison of three varieties of non-cemented porous-coated hip replacement. *British Journal of Bone and Joint Surgery*, 72-B(1):2–8, January 1990.

[45] R.J. Haddad, T.C. Skalley, S.D. Cook, M.R. Brinker, J. Cheramie, R. Meyer, and J. Missry. Clinical and roentgenographic evaluation of non-cemented porous-coated anatomic medulary locking (AML) and porous-coated anatomic (PCA) total hip arthroplasties. *Clinical Orthopaedics and Related Research*, 258:176–182, 1990.

[46] Hamilton and Hufnagel. *Applications of Neural Networks*, chapter Early Detection of Epileptic Attacks, pages 173–178. VCH Verlagsgesellschaft, Weinheim, 1992.

[47] W. Harris, R. Krushell, and J. Galante. Results of cement-less revisions of total hip arthroplasties using the Harris-Galante prosthesis. *Clinical Orthopaedics and Related Research*, (235):120–126, October 1988.

[48] W.H. Harris. Traumatic arthritis of the hip after dislocation in acetabular fractures: Treatment by mould arthroplasty. *The Journal of Bone and Joint Surgery*, 51-A:737–755, 1969.

[49] Anna Hart. *Knowledge Acquisition for Expert Systems.* Kogan Page, 1986. ISBN 0–85091–091–X.

[50] D.L. Hauser. *Optimisation of the fit of femoral prostheses for total hip arthroplasty.* Phd thesis, Cornell University, 1993.

[51] D. Hayes, N. Sharkey, S. Yerby, and H. Paul. Implant stability and design: A correlation analysis using interfacial micro-motions and fit, fill, and gap data. In *Transactions of the 38th Annual Meeting of the Orthopaedic Research Society*, page 383, February 17–20 1992. Washington, D.C.

[52] A. Hedley, T. Gruen, and D. Ruoff. Revision of failed total hip arthroplasties with uncemented porous-coated anatomic components. *Clinical Orthopaedics and Related Research*, (235):75–90, October 1988.

[53] P. Herberts, L. Ahnfelt, H. Malchau, C. Strömberg, and G.B. Andersson. Multi-centre clinical trials and their value in assessing total joint arthroplasty. *Clinical Orthopaedics*, 249:48–55, 1989.

[54] P. Holmberg, J. Bechtold, B. Sun, R. Kyle, and R. Gustilo. Strain analysis of a femur with long stem press fit prosthesis. In *Transactions of the 32nd Annual Meeting of the Orthopaedic Research Society*, New Orleans, Louisiana, February 17–20 1986.

[55] G. Horne. Fit-and-fill: Fashionable fact or fantasy? *The Journal of Bone and Joint Surgery*, 74(2):4–5, January 1992.

[56] Kurt Hornik. Approximation capabilities of multi-layer feed-forward networks. *Neural Networks*, 4(2):251–257, 1991.

[57] R. Huiskes. The various stress patterns of press-fit, ingrown, and cemented femoral stems. *Clinical Orthopaedics and Related Research*, (261):27–38, December 1990.

[58] R. Huiskes. *Basic orthopaedic biomechanics*, chapter Biomechanics of artificial joint fixation, pages 375–442. Raven Press, New York, 1991.

[59] R. Huiskes and E. Chao. A survey of finite element analysis in orthopaedic biomechanics: The first decade. *Journal of Biomechanics*, 16(6):395–409, 1983.

[60] Rik Huiskes. Failed innovation in total hip replacement: Diagnosis and proposals for a cure. *Acta Orthopaedica Scandinavica*, 64(6):699–716, 1993.

[61] D.S. Hungerford and L.C. Jones. The rationale for cement-less revision of cemented arthroplasty failures. *Clinical Orthopaedics and Related Research*, (235):12–24, October 1988.

[62] Peter Jackson. *Introduction to Expert Systems.* Addison-Wesley, second edition, 1990. ISBN 0–201–17578–9.

[63] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgement under Uncertainty: Heuristics and Biases.* Cambridge University Press, 1982.

[64] D. Kahneman and A. Tversky. Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3:430–54, 1972.

[65] E. Kamaric, P. Noble, and H. Tullos. Computer optimisation of the shape of cement-less femoral prostheses. In *Proceedings Fifth Annual International Symposium on Custom Prostheses*, Windsor, England, October 1–3 1992.

[66] T. Kanade, M.L. Reed, and L.E. Weiss. New technologies and applications in robotics. *Communications of the ACM*, 37(3):58–67, 1994.

[67] B. Kavanagh, M. Dewitz, D. Ilstrup, R. Stauffer, and M. Coventry. Charnley total hip arthroplasty with cement. *The Journal of Bone and Joint Surgery*, 71-A:1496–1503, December 1989.

[68] B. Kavanaugh and R. Fitzgerald. Clinical and roentenographic assessment of total hip arthroplasty. *Clinical Orthopaedics and Related Research*, (193), March 1985.

[69] B. Kavanaugh, D. Ilstrup, and R. Fitzgerald. Revision total hip arthroplasty. *The Journal of Bone and Joint Surgery*, 67-A:517–526, November 1985.

[70] Tony Keaveny. *A Finite Element Analysis of Load Transfer and Relative Motion For Contemporary Cement-Less Hip Implants in the Short and Long-Terms.* Phd thesis, Cornel University, Ithaca, New York, January 1992.

[71] C.J. Kershaw, R.M. Atkins, C.A.F. Dodd, and C.J.K. Bulstrode. Revision total hip arthroplasty for aseptic failure—a review of 276 cases. *The Journal of Bone and Joint Surgery*, 73-B:564–568, 1991.

[72] Young-Hoo Kim and Vana E. Kim. Results of the Harris-Galante cement-less hip prosthesis. *British Journal of Bone and Joint Surgery*, 74-B(1):83–87, January 1992.

[73] B.J. Kiratli, J.P. Heiner, A.A. McBeath, and M.A. Wilson. Determination of bone mineral density by dual X-ray absorptiometry in patients with uncemented total hip arthroplasty. *J Orthop Res*, 10(6):836–844, 1992.

[74] G. Lord, J. Marotte, J. Guillamon, and J. Blanchard. Cement-less revisions of failed aseptic cemented and cement-less total hip arthroplas-

ties. *Clinical Orthopaedics and Related Research*, (235):67–74, October 1988.

[75] J. Loudon and M. Older. Subsidence of the femoral component related to long-term outcome of hip replacement. *The Journal of Bone and Joint Surgery*, 71(4):624–628, August 1989.

[76] Peter Lucas and Linda van der Gaag. *Principles of Expert Systems.* Addison-Wesley, 1991. ISBN 0–201–41640–9.

[77] Lundvall and Bengt-Ake. *National Systems of Innovation: Towards a Theory of Innovation and Learning.* Pinter, London, 1992.

[78] O.M. Mahoney and J.H. Dimon. Unsatisfactory results with a ceramic total hip prosthesis. *The Journal of Bone and Joint Surgery*, 73-A:191–193, 1990.

[79] Henrik Malchau. On the importance of stepwise introduction of new hip implant technology: Assessment of total hip replacement using clinical evaluation, radiostereometry, digitised radiography and a national hip registry. From the Department of Orthopaedics, Institute of Surgical Sciences, Göteborg University, Göteborg, Sweden, 1995.

[80] Henrik Malchau, Peter Herberts, and Lennart Ahnfelt. Prognosis of total hip replacement in Sweden: Follow-up of 92,675 operations performed 1978–1990. *Acta Orthopedica Scandinavica*, 64(5):497–506, 1993. ISBN 0001–6470.

[81] Henrik Malchau, Peter Herberts, and Lennart Ahnfelt. Prognosis of total hip replacement. Technical report, The Swedish National Hip Arthroplasty Register, 1996.

[82] Timothy Masters. *Practical Neural Network Recipes in C++*. Academic Press, 1993. ISBN 0–12–479040–2.

[83] Martin McKenna. A technology audit of IBM. Internal study, 1998.

[84] D. Michie, D.J. Spiegelhalter, and C.C. Taylor. *Machine learning, neural and statistical classification*, chapter Introduction, pages 1–5. Ellis Horwood series in artificial intelligence. Ellis Horwood, 1994. ISBN 0–13–106360–X.

[85] R.A Miller, H.E. Pople, and J.D. Myers. Internist-i, an experimental computer-based diagnostic consultatn for general internal medicine. *New England Journal of Medicine*, (307):468–76, 1982.

[86] Marvin Minsky and Papert Seymour. *Perceptrons*. MIT Press, Cambridge, Mass., 1969.

[87] C. Moallemi. Classifying cells for cancer diagnosis using neural networks. *IEEE Expert*, pages 8–12, 1991.

[88] R.A. Morose. A financial neural-network application. *AI Expert*, pages 50–53, May 1990.

[89] R.A. Morose. *Neural Networks in Finance and Investing*, chapter A Financial Neural-Network Application, pages 75–83. Probus Publishing, Chicago, 1993.

[90] R.W. Morris. Editorial: Evidence-based choice of hip prostheses. *British Journal of Bone and Joint Surgery*, 78-B(5):691–693, September 1996.

[91] E. Morscher, editor. *The Cement-Less Fixation of Hip Endoprostheses.* Springer-Verlag, New York, 1984.

[92] R.D. Mulroy and W.H. Harris. The effect of improved cementing techniques on component loosening in total hip replacement. *British Journal of Bone and Joint Surgery*, (B-72):757–760, 1990.

[93] L. Munuera and E. Garcia-Cimbrelo. The femoral component in low-friction arthroplasty after ten years. *Clinical Orthopaedics and Related Research*, (279):163–175, June 1992.

[94] D. Murray. *Outcome Measures in Orthopaedics*, chapter 10, pages 198–227. Butterworth-Heinemann, 1993. ISBN 0–7506–0520–0.

[95] D.W. Murray, A.J. Carr, and C.J. Bulstrode. Which primary total hip replacement? *British Journal of Bone and Joint Surgery*, 77-B(4):520–527, July 1995.

[96] P. Noble, J. Alexander, L. Lindahl, D. Yew, W. Granberry, and H. Tullos. The anatomic basis of femoral component design. *Clinical Orthopaedics and Related Research*, (235):148–165, 1988.

[97] D. Nunn, M. Freeman, K. Tanner, and W. Bonfield. Torsional stability of the femoral component of hip arthroplasty. *The Journal of Bone and Joint Surgery*, 71(3):452–455, May 1989.

[98] OECD, Paris. *The Knowledge-based Economy*, 1996. OECD refers to the Organisation for Economic Co-operation and Development.

[99] Dan W. Patterson. *Artificial neural networks: Theory and applications*. Simon & Schuster, 1996. ISBN 0–13–295353–6.

[100] M. Patterson. Ring uncemented hip replacements, the results of revision. *The Journal of Bone and Joint Surgery*, 69-B(3):374–380, May 1987.

[101] H. Paul, W. Bargar, B. Mittlestadt, B. Musits, R. Taylor, P. Kazanzides, J. Zuhars, B. Williamson, and W. Hanson. Development of surgical robot for cement-less total hip arthroplasty. *Clinical Orthopaedics and Related Research*, (285):57–66, December 1992.

[102] H. Paul, B. Mittelstadt, B. Musits, W. Bargar, P. Kazanaizes, W. Williamson, J. Zuhars, R. Taylor, and W. Hanson. Accuracy of implant interface preparation: Hand-held broach vs. robot machine tool. In *Transactions of the 38th Annual Meeting of the Orthopaedic Research Society*, page 374, Washington, D.C., February 17–20 1992.

[103] R. Peto, M.C. Pike, P. Armitage, N.E. Breslow, D.R. Cox, S.V. Howard, N. Mantel, K. McPherson, J. Peto, and P.G. Smith. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. *Br J Cancer*, 35:1–39, 1977.

[104] T. Phillips and S. Messieh. Cement-less hip replacement for arthritis. *The Journal of Bone and Joint Surgery*, 70(5):750–755, November 1988.

[105] T. Phillips, S. Messieh, and P. McDonald. Femoral stem fixation in hip replacement. *The Journal of Bone and Joint Surgery*, 72(3):431–434, May 1990.

[106] R. Poss, D. Robertson, P. Walker, D. Reilly, F. Ewald, W. Thomas, and C. Sledge. *Non-Cemented Total Hip Arthroplasty: The Bone Interface*, chapter Anatomic stem design for press-fit and cemented application. Churchill Livingstone, New York, 1988.

[107] C. Ranawat, A. Burstein, D. Bartel, and M. Silva. 8311 total hip—a new design to maximise metaphyseal contact for load sharing. In *Proceedings Third Annual International Symposium on Custom Prostheses*, Nice, France, October 3–5 1990.

[108] Edward B. Roberts. Managing technological innovation in the medical devices industry. *Technology Management*, July–August 1989.

[109] D. Robertson, P. Walker, S. Hirano, X. Zhou, J. Granholm, and M. Poss. Improving the fit of press-fit hip stems. *Clinical Orthopaedics and Related Research*, (228):134–140, 1988.

[110] D. Robertson, P. Weiss, E. Fishman, D. Magid, and P. Walker. Evaluation of CT techniques for reducing artefacts in the presence of metallic orthopaedic implants. *Journal of Computer Assisted Tomography*, 12(2):236–241, April 1988.

[111] D.D. Robertson, D. Magid, R. Poss, E. Fishman, A. Brooker, and C. Sledge. Enhanced computed tomographic techniques for the evalu-

ation of total hip arthroplasty. *The Journal of Arthroplasty*, 4(3):271–276, September 1989.

[112] J.A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12:23–41, 1965.

[113] P. Rubin, P. Leyvraz, J. Aubaniac, J. Argenson, P. Esteve, and B. DeRoguin. The morphology of the proximal femur. *The Journal of Bone and Joint Surgery*, 74(1):28–32, January 1992.

[114] Jan-Willem Schimmel and Rik Huiskes. Primary fit of the Lord cementless total hip, a geometric study in cadavers. *Acta Orthopedica Scandinavica*, 59(6):638–642, 1988.

[115] T. Schmalzried, L. Kwong, M. Jasty, R. Sedlacek, T. Haire, D. O'Connor, C. Bradgon, M. Kabo, A. Malcolm, M. Path, and W. Harris. The mechanisms of loosening of cemented acetabular components in total hip arthroplasty. *Clinical Orthopaedics and Related Research*, (274):60–78, January 1992.

[116] W. Scott, L. Riley, and H. Dorfman. Focal lytic lesions associated with femoral stem loosening in total hip prosthesis. *American Journal of Radiology*, (144):977–982, May 1985.

[117] G. Selvik. A roentgenstereophotogrammetric method for the study of the kinematics of the skeletal system. *Acta Orthopaedica Scandinavica Supplementum*, 232, 1989.

[118] F. Seral, J. Villar, A. Esteller, F. Vivar, I. Abad, M. Grande, E. Jorad, and E. Espinar. Five-year follow-up evaluation of the non-cemented press-fit titanium hip-joint endoprosthesis. *Clinical Orthopaedics and Related Research*, (283):49–56, October 1992.

[119] E.H. Shortliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York, 1976.

[120] Alberto Pacheco. Fuzzy logic IF-THEN rules in Amzi! Prolog. World Wide Web, 1997. http://campus.chi.itesm.mx/~apacheco/fuzzy/fuzzy-2.htm.

[121] K. Soballe, N. Olsen, R. Ejsted, F. Christensen, and T. Luxhoj. Revision of the uncemented hip prosthesis. *Acta Orthopaedica Scandinavica Supplementum*, 58:639–633, 1987.

[122] R. Stauffer. Ten-year follow-up study of total hip replacement, with particular reference to roentgenographic loosening of the components. *The Journal of Bone and Joint Surgery*, 64-A(7):983–990, September 1982.

[123] Richard N. Stauffer and Edmund Y.S. Chao. *Joint Replacement Arthroplasty*, chapter Femur, pages 573–585. Churchill Livingstone, New York, 1990.

[124] Leon Sterling and Ehud Shapiro. *The art of Prolog: Advanced programming techniques*. MIT Press, Cambridge, Mass., second edition, 1994. ISBN 0–262–19338–8.

[133] J.T. Wang, N.E. Bennett, and S.R. Goldring. Molecular mechanisms involved in activation of cells by particulate orthopedic implant biomaterials. In *Transactions of the 38th Annual Meeting of the Orthopaedic Research Society*, page 43, Washington, D.C., February 17–20 1992.

[134] M. Williams, S. Frankel, K. Nanchahal, J. Coast, and J. Donovan. Total hip replacement DHA project: Research programme epidemiologically based needs assessment. Health Care Evaluation Unit, 1993.

[135] Patrick Henry Winston. *Artificial Intelligence*. Addison-Wesley, second edition, 1984. ISBN 0–201–08259–4.

[136] Nancy L. Young, David Cheah, James P. Waddell, and James G. Wright. Patient characteristics that affect the outcome of total hip arthroplasty: A review. *Can J Surg*, 41(3):188–195, June 1998.