Utility Theory and its use in

Managerial Systems (an NHS perspective)

by

Melvyn Langford

Doctoral Thesis

Submitted in fulfilment of the

requirements for the award of

Doctor of Philosophy of Loughborough University

June 2009

Abstract

This thesis originated from a research question created by a focus group of National Health Service (NHS) senior estates managers, who considered that the systems of internal control do not give adequate assurance that NHS healthcare building services engineering day-to-day maintenance activities conform to the national guidance. The initial aim of this research was to test their concerns against empirical evidence from NHS Trusts. This was achieved by identifying the gaps within the participating Trusts' maintenance managerial activities when assessed against national standards. Central to the methodology used to assess the level of dynamic risk being generated was the rejection of the NHS national standard 5x5 risk criticality grid in common use throughout the health service, in favour of a series of specific 'Utility Functions'. This has created greater transparency and robustness of the risk assessment process. To the researcher's knowledge, this is the first time that 'Utility Theory' has been used in such scenarios. The result of this analysis has shown their fear to be correct. And for each of the 31 NHS Trusts taking part, the multi-professional focus groups composed of their own senior managers confirmed that there are areas of non-conformance within their maintenance regime, which were previously unknown. In all cases the organisations considered that their failings were exposing their patients, staff, public and stakeholders to substantial/intolerable risk through a 'systematic' failure of the Trusts' governance systems.

The aim of this research then expanded to design techniques that specifically assesses the resource needs to close these managerial gaps employing industry standard techniques. Then again employing 'Utility Theory' examined various revenue levels of directly employed maintenance artisan resource with respect to risk, via a specifically designed simulation model. This has proved that the historical NHS methodology for assessing workforce planning to be fundamentally flawed, as it seriously underestimates the resource need. This research then developed and designed a generic day-to-day monitoring assurance framework from reference to the research into High Reliability Organisations, Normal Accident Theory and managerial governance needs.

The overriding recommendation from the research was that NHS Trusts must redesign their governance systems to ensure that they are aware of their estates department's non-conformances when assessed against national standards.

List of Contents	Page

Acknowledgements

Preface

A National Health Service (NHS) report (Department of Health 2000) suggested that an estimated 850,000 (range 300,000 to 1.4 million) adverse events might occur each year in the NHS hospital sector, resulting in a £2 billion direct cost through patients needing additional hospitalisation. Some of these adverse events will be inevitable complications of treatment but around half might be avoidable. As part of this thesis research due to the concerns raised within the report, a focus group was convened consisting of ten experienced NHS estate managers from England and Northern Ireland. The focus group discussed the Department of Health document and identified the main managerial systems within an estates department that they considered to be generating the greatest unaddressed perceived risk. The output from these discussions was apprehension relating to the level of conformance to national standards in relation to the engineering work controlled by these departments. This led to the following research question: -

'Do the managerial systems of internal control within the National Health Service (NHS) assure that the day-to-day building service engineering maintenance activities conform to their respective national guidelines?

The results of this research conclusively show that of the 31 NHS healthcare organisations analysed as part of this research, not one conformed to these national standards. In all cases the organisation's senior managers considered that their governance failings were:

Exposing their organisation, patients, staff, public and stakeholders to substantial/intolerable risk through a 'systematic' failure of their governance systems.

During the year 2001 the author began to assemble detailed sample information of the actual engineering operational policies and procedures practised. This data was then compared against the national guidance that had been specifically developed by the NHS for such maintenance activities, and assessed the risk of any non-conformances. The data collection has continued up to the present day. This research created four research objectives to split the governance problem into fine detail. The research

1

instruments developed form robust and transparent research methodologies, that identified latent and dangerous conditions that were previously unknown; created assessments of the resource needs to close these managerial gaps; estimated the revenue consequences of keeping the gaps closed; and then developed a monitoring regime methodology to give assurance that the gaps remain closed.

This thesis is composed of six chapters; a summary of each is presented below:

Chapter One - Introduction

This chapter traces the history of the National Health Service (NHS) from its creation in 1948 to the present day, mapped in ten year timeframes. The fifth decade (1988 – 1997) saw the introduction of the internal market within the Service. This split the NHS into 'purchasers' and 'providers'. Purchasers were health authorities and some family doctors. These organisations commissioned medical services from the providers, where the providers were: the acute hospitals, organisations providing care for the mentally ill, people with learning disabilities and the elderly, along with ambulance services. Providers became NHS Trusts with their own independent organisations and management structures.

Included within this Chapter is a discussion relating to the government's strategy to assess the performance of the NHS against its conformance to national standards. It was from this backdrop that the research topic was developed that questioned the managerial control and assurance systems then in use, to ensure that NHS Trust engineering day-to-day maintenance activities were conforming to national codes of practice.

Chapter Two – Literature Review

This Chapter begins with a short description of reliability and risk, and the balance that must be achieved between these two factors when designing a maintenance strategy. The Department of Health's national guidance relating to the formation of a Trust's maintenance strategy issued in the year 2002 is discussed, and compared to that published in 1973. From this assessment it is shown that there has been no change in policy over the 30 intervening years, and that Trusts are free to create their own maintenance strategy to reflect local conditions and budgets.

The term 'maintenance' is examined and it is shown that it includes the managerial activities as well as the physical artisan tasks, thus forming a total system. The artisan tasks are then broken down into the two categories of reactive and planned preventative maintenance, and examples are given for each type.

The use of computers within maintenance departments is shown to be widespread throughout the NHS, and that the NHS has abandoned the development of a centralised form of software, forcing Trust's to purchase commercial packages to assist in the management of their data needs. Thus as a consequence there is no national data bank containing quantitative records relating to maintenance activity, or offering plant reliability statistics. The use of simulation models of the past is shown to be of little practical use to the NHS maintenance manager, even though there have been various attempts to develop such a system over the years.

The type of maintenance strategy to be adopted is considered, and an argument formed that shows that actuarial analysis relating to the relationship between plant and equipment operating age and failure is of little use to the NHS day-to-day maintenance manager. This then leads to the conclusion that the maintenance strategy should be targeted towards the system known as Reliability Centred Maintenance .

The assumption that plant and equipment failures are totally independent of managerial failures is discussed, as is the research which has shown that the vast majority of incidents are caused through managerial decisions, actions and failures to act. This has focused the thesis to consider the risks associated with the day-to-day activities of managerial systems, as opposed to those being generated from the physical plant and equipment.

Normal accident theory is discussed and the section concludes with the observation from its founder Perrow (1999) that the supporters of high reliability theory believe that if we tried harder we could overcome the proposed inevitability of accidents, and have a virtually accident free systems. This is in total contrast to normal accident theory, which states that, there will always be accidents due to the complex interaction between tightly coupled systems. Perrow also suggests that we need to examine organisations and systems that have not yet had accidents, in a similar manner that we would following an

3

accident. He states this in his belief that one would conclude that the system may well be just waiting for an accident to happen.

The creation of a suitable managerial system is then debated along with the two main techniques used to assess the risks associated with such a design. From this two themes have emerged relating to normal day-to-day activities (i.e. non-emergency). The first is the identification of the tipping point, where an organisation eventually fails to cope with an ever increasing workload, and the second is the need for monitoring/audit programs to be designed and implemented. A research model relating to an inherent failure of a management system due to the quantity of non-novel interruptions during normal day-to-day activities is examined. The Chapter then makes a parallel between the model's outputs and that of a typical NHS estates department. The designers of the model concluded their findings recommending that:-

> Future research could be profitably focused on helping organisations assess their capacity to handle interruptions and develop systems to signal when the level of unresolved interruptions reaches a critical level.
>
> (Rudolph & Repenning 2002)

It is then stated that this thesis will test the concept of this theoretical model against actual data from an NHS estates organisation.

Finally the need for, and development of, an assurance framework is discussed, along with the call for an integrated monitoring and audit programme to be implemented.

Chapter Three - Methodology

From reference to a focus group populated by ten NHS Trust estates managers, eight engineering topics were highlighted, as these were believed to be posing the greatest unaddressed perceived risks. Research instruments were then designed to assess the perceived risks being generated from these topics as an integral part of the first research objective which was to develop and implement a robust, transparent and repeatable methodology to assess if the managerial systems of internal control within the National Health Service (NHS) give adequate assurance that the day-to-day building service engineering maintenance activities conform to their respective national guidelines.

The methodology employed consisted of a number of Trust specific multi-professional focus groups assessing their managerial conformance to the relevant national standards. Use was made of a series of utility functions in place of the NHS 5x5 criticality grids in order to achieve greater repeatability of the groups' assessed risk scores. The output from each of the Trust's focus groups was the development of a risk profile.

Focus was then centred on the salient criteria that were considered to influence the risk profiles produced. The corresponding methodologies used to reduce the negative effects of these criteria were also presented. Although the criteria were discussed as individual topics, in practice they were all interacting to create a similar effect, that of reducing the negative impact of the criteria to an insignificant value.

Following the development of the risk profiles this Chapter then completed the second research objective to develop and implement a robust, transparent and repeatable methodology to assess the resource needs to correct the existing managerial system deficiencies to that required by their relevant National Health Service standard.. To tackle this, again Trust specific focus groups were formed to assess the managerial time and budget costs associated with each of the managerial gaps identified. The tools used to assist in these assessments employed industry standard Gant chart and Monte Carlo simulation software.

This led the research to develop a simulation program to address the third research objective to develop and implement a robust, transparent and repeatable methodology to assess the directly employed artisan revenue resources needed to ensure that the maintenance activities conform to the National Health Service standards. In response, reference was made to the Literature Review chapter which discussed the research model developed by Rudolph and Repenning (2002), relating to an inherent failure of a management system due to the quantity of non-novel interruptions during normal day-to-day activities, and a similarity was drawn between the model's outputs and that of a typical NHS estates department. It is then stated that this thesis would test the concept of this theoretical model against actual data from an NHS estates organisation, in order to predict the 'tipping point' of the department caused through a reduction of resources.

Before work began to gather the data necessary from Trusts to test the Rudolph and Repenning model, a number of assumptions were made that allowed the development of a simple estates model. This simple model was simulated using actual data for the planned preventative maintenance system with synthetic data to represent the reactive workload, which was based on the research of Al-Zubaidi and Christer (1997). The outputs from the simulation of the simplified estates model produced observations comparable to those of the original Rudolph and Repenning model. This gave confidence to the researcher to move the project forward by enhancing the software design and simulating a Trust's 'actual data'.

Finally the Methodology Chapter covers the creation of, an assurance framework under the fourth research objective to develop a robust, transparent and repeatable methodology to monitor the day-to-day activities of the maintenance department in line with the requirements of the National Health Service standards. As discussed within the Literature Review chapter, following a call for an integrated monitoring and audit programme to be implemented.

Five criteria are outlined within the Literature Research chapter that created a framework around which the monitoring system had to be designed. In summary the criteria required that the system must conform to international and national standards, reference mainstream academic theory, employ existing hierarchal staffing structures, job descriptions and software, be applicable to a multi-professional team and be revenue neutral against existing budgets.

The methodology presented for this fourth research objective conformed to all of the five criteria outlined above. This was achieved by the design of a 'driver' to push a common 'generic framework', which employed standard operational performance management techniques. The output of the process was a qualitative assessment of the maintenance department's efficiency and quality towards supporting the Trust in achieving its objective goals. This was presented as a series of performance index profiles against a base period, showing the actual achievement against a target value.

The introduction of the monitoring process into a department was designed to follow typical project management techniques. As each Trust and estates department differs in

their organisational structure, reporting arrangements and managerial systems, only a generic approach could be presented within this Methodology Chapter, as each project must be tailored to cater for each situation.

Chapter Four - Results

The methodology and subsequent results naturally split into the creation of four disparate research objectives each with an accompanying results section. From the 31 Trusts taking part in this research the results from the first research objective created demonstrable proof that there was a systematic failure of these organisations to conform to the recognised NHS national engineering standards. The results uncovered previously unknown latent conditions that were exposing the Trust, staff, public, patients and stakeholders to substantial/intolerable levels of risk.

The results from the second research objective identified in detail the resources (manpower, costs and timeframes) needed to close the managerial gaps identified within the first research objective. This information allowed the 31 Trusts (in all cases) to allocate their limited resources to 'where they would do most good'; and to project manage the closure of the gaps against a pre-agreed timeframe and budget allocation.

The results from the third research objective were created by simulating Trust specific data to assess the effects of varying resource allocations against subsequent estimations of generated risk. This allowed the senior managers within each Trust to judge what resources they considered appropriate to allocate to their local maintenance needs, when balanced against other cost/risk pressures within their organisations as a whole.

Due to the time constraints imposed on this research programme there are no results presented from the fourth research objective, that show the effects of the development, design, implementation and maintenance of the monitoring regime developed as part of this thesis.

Chapter Five - Discussion

This Chapter considers the research cited within the Literature Search Chapter, the techniques employed as created within the Methodology Chapter, and the robustness of the outcomes as tested within the Results Chapter. Following the successful analysis,

discussion and conclusion of the procedures the researcher then summarises the findings.

The output from the first research objective was shown to uncover previously unknown latent gaps within the managerial systems. In some cases these gaps were judged by the relevant Trust senior managers to be generating serious/intolerable levels of risk to their organisations, patients, public, staff and stakeholders. The output from this objective allowed the Trusts to refocus their limited resources to 'where they would do most good', Testimony to the value of this output was received by the Trusts requesting follow up studies to confirm progress in closing the gaps, and through national recognition and publication.

The second research objective outputs estimated the resources needed to close the managerial gaps identified. This gave a budget expenditure and risk reduction profile against a time line with which to project manage the closure of the managerial failings.

The third research objective outputs allowed the Trusts to determine what revenue resources were needed to keep the risks associated with maintenance to an acceptable level. As a direct consequence of this research, it is now possible for the first time in the 60 year history of the NHS for evidence to be shown of the consequences of varying resource in terms of the effects on risk.

The simulation model specifically developed for this shows that even for reasonably resourced organisation, peaks of risk naturally occurred within the red and amber zones with respect to time. This suggested that maintenance revenue budgets should be variable and not fixed year-on-year. The sensitivity of the simulation model was tested and this demonstrated that only negligible effects were experienced by: the use of interpolation in place of integration within the simulation software to speed its design; the sensitivity of the probability utility to the focus groups being particularly risk adverse; and the focus groups' judgement in assessing the ratio differences between the impact descriptors.

This research instrument also questioned if the simulation models created could be 'reverse engineered', to assess the trade group resource needs of a Trust if they had no

reliable raw data. This led to the development and use of synthetic data for available man-hours and reactive hours, and showed that it was possible to produce reasonable results from such information to within small tolerance limits of accuracy, when compared against actual values derived from original data.

The fourth research objective output which was designed to monitor the day-to-day managerial activities was not discussed in depth within the Chapter, as it was not possible to implement the methodology into a Trust within the acceptable time scale of this research programme.

Chapter Six - Conclusions

In this Chapter the whole research programme is summarised by stating that the main research question of this Thesis has been answered, and that from the 31 healthcare organisations analysed the results of this research conclusively show that, not one had an adequate managerial system of internal control to identify their estate department's non-conformances to national standards, and in all cases each Trust's own senior managers arising from multi-professional focus groups considered that the previously unknown systematic failings of their governance systems were exposing their organisation, patients, staff, public and stakeholders to substantial/ intolerable risk.

The Chapter then reviews the outputs from each of the research instruments specifically developed for each of the four research objectives, and discusses their novel academic contribution.

Chapter One   Introduction

1.1   Introduction

This chapter traces the history of the National Health Service (NHS) from its creation in 1948 to the present day, mapped in ten year timeframes. Included within this is a discussion relating to the government's strategy to assess the performance of the NHS against its conformance to national standards.

1.2   The first four decades

The history of the National Health Service (NHS) within the United Kingdom began with its formation in the year 1948 (see figure 1 over page) (The National Health Service 2009). Before this the poor went without comprehensive medical treatment, and relied on charity or home remedies. Some workers (but not their families) had access to doctors, but the great majority of hospitals charged for their services. In the 19th century social reformers tried to give free medical care to the poor, but the demand for these services outstripped the limited charitable resources. Local authorities of large towns created municipal hospitals as well as hospitals specializing in maternity, infectious diseases, care of the elderly, the mentally ill and those suffering from a mental handicap. The mentally ill were locked inside institutions and housed in appalling conditions. Many of the elderly who could no longer look after themselves were placed into Victorian workhouses. Following the Second World War, on the 5th July 1948, the NHS was created in a country where food was rationed, the economy was in crisis and there were shortages of building materials and fuel. The founding fathers of the NHS predicted that 'expectations would always exceed capacity', and as such the Service must always be in a state of constant flux. Their foresight proved to be correct as in administration terms it was found to be impossible to predict the day-to-day costs of this new Service as the demand rose, along with medical technology and consumer expectations. Inevitably the cost of the NHS began to exceed the initial estimates. In the second decade of its history (1958 – 1967) the NHS began to improve its management systems. Data in the form of Hospital Activity Analysis was developed to produce patient-based information.

Figure 1   *The six decades of the*
            NHS

1948 NHS formed

1958

Major
rebuilding
program

Hospital
activity
analysis data
created

1968

Rise in new medical
techniques

Strategic planning
introduced

Estmancode
published

1978

Medical techniques
outstripped resources

Performance indicators
introduced

Overspends, wards closed,
waiting lists increased

1988

Introduction of
internal market

Controls Assurance introduced

Labour government
elected, internal
market abolished

1998

Information strategy
developed

Corporate governance
adopted

Publication of An
organisation with a memory

Health Care Standards Unit
formed

2008

Alongside this, various reports were produced that covered numerous topics within the Service, ranging from medical staffing structures to the creation of major rebuilding programs to replace the war-torn and outdated healthcare buildings. A common trait within the reports was the acknowledgement of the complexity of the organisation and the need for constant change.

The third decade (1968 – 1977) of the NHS witnessed a continuing rise in the use of new medical techniques such as: endoscopy and CAT (Computerised Axial Tomography) scanning, kidney transplants and further developments in the field of drugs. New hospitals were brought into use to provide more centralised specialist services and local facilities. Debate continued to question how the NHS should be managed and organised. Managerial techniques such as strategic planning, long-range forecasts and reallocation began to emerge.

It was during this decade that the NHS recognised that very few types of public buildings suffered the continuous occupation and high degree of wear and tear of hospitals and other healthcare buildings.

11

This problem was compounded with the need for these buildings to be maintained with minimum effects on their occupants, and disruption to clinical procedures. Therefore during this decade the NHS published a document entitled Estmancode. This guidance for the first time gave a national framework around which maintenance could be systematically planned and introduced into the service (Department of Health and Social Security 1973). The document stated that the main objective of the maintenance organisation was: -

'The efficient and economical conservation of the National Health Service estate, subject to the overall needs of the service and the social constraints imposed on the National Health Service.'

Estmancode,

Chapter 1, Section 1, paragraph 3.1.1,

July 1973.

The national code introduced the then new concept of reactive and planned preventative maintenance into the service for the first time. reactive maintenance was the repair of the random breakdown of plant and equipment e.g. the replacement of light bulbs, repair of dripping taps, etc. planned preventative maintenance on the other hand was the pre-planned and agreed examination, checking and testing of a building's infrastructure. This preventative type of maintenance was developed with the specific aim of reducing random breakdowns and subsequent disruption to clinical activity, and to extend the working life of the installations.

The guidance also introduced the concept of Quality Control into the maintenance management function. This inherently focused attention onto developing frameworks for reviewing maintenance performance. Bi-annual returns to the NHS Headquarters were required, that were designed to show the extent of the planned preventative engineering maintenance schemes, as an indication of the maintenance effectiveness being achieved in each health premises. In this way the performance of each Region/Area/District could be judged according to the workload covered by planned preventative maintenance and their relative position in a 'league table' (Estmancode 1973 paragraph 2.2.1.1). Additional information relating to the maintenance function was accessed from reference to the national self financing engineering incentive scheme

(commonly referred to as the bonus scheme) operated across all NHS engineering departments (National Performance Advisory Group 1997).

The fourth decade (1978 – 1987) acknowledged that medical techniques had outstripped the available resources and that the NHS could not do everything that was now possible or expected (The National Health Service 2009). The medical advances by now covered all areas of NHS activity. For example, genetic engineering was developing its first drugs, magnetic resonance imaging was provided, and the number of orthopaedic operations along with heart and liver transplants rose rapidly.

Clinical budgeting was introduced with a realisation that better information was needed, therefore performance indicators were developed. Audit began to examine the rising costs following international concern over the costs of anaesthesia and surgery. By the close of the decade some health authorities had overspent their budget, hospital wards were being reduced, and waiting lists were increasing in length.

1.3 The fifth decade

The fifth decade (1988 – 1997) saw the introduction of the internal market within the Service. This split the NHS into 'purchasers' and 'providers' (The National Health Service 2009). Purchasers were health authorities and some family doctors. These organisations commissioned medical services from the providers, where the providers were: the acute hospitals, organisations providing care for the mentally ill, people with learning disabilities and the elderly, along with ambulance services. Providers became NHS Trusts with their own independent organisations and management structures. These organisations were in the main developed by splitting the former Areas and Districts into small independent statutory bodies.

The internal market created competition between Trusts as they bid against each other, thus improving cost consciousness within the NHS. As the engineering workforce was now split into Trusts as opposed to large Areas/Districts, the national incentive schemes (bonus schemes) were in the majority of cases disbanded. From discussions between the author and a number of Trust estates managers they state that this was done as a cost cutting measure and because of the inability for each independent Trust scheme to be self financing, partly due to the small numbers of engineering staff within each Trust. This stopped at a stroke any form of cooperation between engineering departments in

Trusts to compare maintenance strategies, notes, figures etc. Trusts were left to manage their own affairs.

Following the Cadbury Report (Cadbury 1992) a system called 'controls assurance' was introduced into the NHS in 1996. It was initiated to improve the internal controls (including the process of audit) throughout the organisation as a single initiative (Scrivens 2005 p95). In order to underpin these developments the Australian and New Zealand definition of risk management was adopted, which was defined as, 'the culture processes and structures that are directed towards the effective management of potential opportunities and adverse effects' (Standards Australia and Standards New Zealand 2004). Thus any control system should therefore focus on behaviour rather than on outcomes (Scrivens 2005 p 116).

Following the British general election in 1997 a new government was formed that pledged to abolish the internal market, but to keep that which was seen to work well. The government of the day published a white paper through the Department of Health entitled 'The New NHS. Modern Dependable' (Department of Health 1997). This encouraged partnership driven by performance, thus moving away from competition to collaboration. The white paper was based on six key principles, these are listed below: -

1. To renew the NHS as a genuinely national service, offering fair access to consistently high quality, prompt and accessible services right across the country.
2. To make the delivery of healthcare against these new national standards a matter of local responsibility, with local doctors and nurses in the driving seat in shaping services.
3. To get the NHS to work in partnership, breaking down organizational barriers and forging stronger links with local authorities;
4. To drive efficiency through a more rigorous approach to performance, cutting bureaucracy to maximize every pound spent in the NHS for the care of patients;
5. To shift focus into quality of care so that excellence would be guaranteed to all patients, with quality the driving force for decision-making at every level of the service;
6. To rebuild public confidence in the NHS as a public service, accountable to patients, open to the public and shaped by their views.

1.4   The sixth decade

The decade (1998 – 2007) began with the publication in September 1998 of: Information for Health: An Information Strategy for the Modern NHS' (Department of Health 1998) (The National Health Service 2009). This document introduced a strategic approach to the use of information technology. These Trust Boards were now faced with information overload from their respective departments, and had to find ways to separate those risks that would affect the organisation as a whole, and those that were significant at a local (departmental) level.

Documentation was issued by the Institute of Chartered Accountants (The Institute of Chartered Accountants 1999a), that related to 'Corporate Governance' and gave guidance for the internal control requirements of an organisation. It reinforced the principle that all controls should be subject to review. There have been various attempts to define the term Corporate Governance (Encycogov.com 2002), essentially it is intended to secure and motivate the efficient management of organisations for the betterment of its stakeholders and society as a whole. The Audit Office defines corporate governance as the bringing together of the topics of accountability, leadership, control and objectives (Audit Commission 2003).

Turnbull expected that a Trust Board should have an organisational wide view of the effectiveness of their internal control and publish their findings to their stakeholders. From this concept risks would be managed throughout the organisation and the Board would be informed of these issues and their control in what has become known as organisation-wide or enterprise-wide risk management. For this to function, risk management and internal control systems have to be entrenched throughout an organisation.

A further report observed that focusing on too many risks could hide the significance of the key risks. The report stressed that there needs to be a single integrated risk management system that combines both the operational day-to-day risks with those of a strategic nature (The Institute of Chartered Accountants 1999b). Also the cost benefit considerations of the managerial system must be considered during its design, to ensure that the system costs are in proportion to the risks that it is controlling.

15

In the year 2000 the Department of Health published a document (Department of Health 2000) which was a report by an expert group chaired by the Chief Medical Officer, advising the NHS (and others) to learn from adverse events within the organisation. The report described the poor systems that were in place to prevent serious incidents or occurrences; and stated that the position was such that: -

❖ Some failures were avoidable.

❖ Untoward events which could be prevented recur, sometimes with devastating consequences.

❖ Incidents which result in lapses in standards of care in one or more health organisations do not reliably lead to corrections throughout the NHS.

❖ Circumstances that predispose to failure, and which if addressed could allow risks to be minimised, are not well recognised.

The report then listed the estimated cost of these failures within seven statistically based statements. These are reproduced below:-

1. Research suggests that an estimated 850,000 (range 300,000 to 1.4 million) adverse events might occur each year in the NHS hospital sector, resulting in a £2 billion direct cost in additional hospital days alone; some adverse events will be inevitable complications of treatment but around half might be avoidable.

2. The NHS paid out around £400 million in clinical litigation settlements in the financial year 1998/99 and has a potential liability of around £2.4 billion from existing and expected claims; when analysed many cases of litigation show potentially avoidable causes.

3. There were over 38,000 complaints about all aspects of Family Health Services during 1998–99, and nearly 28,000 written complaints about aspects of clinical treatment in hospitals alone.

4. At least 13 patients have died or been paralysed since 1985 because a drug has been wrongly administered by spinal injection.

16

5. Over 6,600 adverse incidents involving medical devices were reported to the Medical Devices Agency in 1999, including 87 deaths and 345 serious injuries.

6. Experience from the serious incident reporting system run by one of the NHS Executive's Regional Officers suggests that nationally at least 2,500 adverse events a year occur which should be serious enough to register on such systems.

7. The costs to the NHS of hospital acquired infections have been estimated at nearly £1 billion a year, and around 15% of cases are regarded as preventable.

These figures relate to failures within the secondary care sector (e.g. hospitals); the expert group did not know the frequency or nature of similar failings within primary care (e.g. community services, GP practices etc). The report produced various tables of information but concluded by suggesting that:

.....nationally the costs to the NHS of extended hospital stays as a result of adverse events could be as high as a further £2 billion a year – five times the cost of clinical negligence litigation.

This estimation of cost did not take into account the human cost of the pain, suffering, trauma (both physical and psychological) of patients and their families, nor trauma to the staff involved. The report draws a number of conclusions, three of which are reproduced below: -

1. Information on the frequency and nature of adverse events in the NHS is patchy and can do no more than give an impression of the problem. Information from primary care is particularly lacking;

2. There is evidence of a range of different kinds of failure, and of the recurrence of identical incidents or incidents with similar root causes;

3. Case studies highlight the consequences of weaknesses in the ability of the NHS as a system to learn from serious adverse events;

The report goes on to state that research within the non-healthcare sectors has shown that for most unintentional failures there is a complex interaction between numerous topics, including human behaviour, technology, socio-cultural factors, organisational and procedural weaknesses (Department of Health 2000). It further stated that similar studies in the healthcare field were sparse, but that the available evidence suggested that a similar pattern of cause and effect relationships existed.

The report recommended a system approach as opposed to a person–centred methodology to analyse and reduce failure, not concentrating upon the human error as the main cause, but taking a holistic view of the elements surrounding an incident. This recognises that human error is influenced by 'upstream' systematic factors that include an organisation's managerial processes and culture towards risk management.

The introduction of humans into a managerial system can generate either an *active failure or latent conditions*, (Reason 1990) where active failures are caused by people's actions or inactions in carrying out work. These latent conditions lie dormant until combined through a set of circumstances to create the condition for a failure to manifest. These latent conditions are created by organisational and managerial systems and strategies. Unlike active failures, latent conditions are always present, and can be readily identified and removed. Attempts to deal with these issues can go to the heart of an organisation's culture, and some managers' personal values, beliefs, etc. Included within the report is also criticism of the NHS in relation to its ability to remember lessons learnt from past failures (Department of Health 2000).

1.4.1 Development and Conformance to National Standards

In order to develop a common approach that would eventually lead to an equal quality of service to the patient across the country, as opposed to a 'post code lottery' of care, the government stated that it was their job to set national standards that the service must comply with.

It was anticipated (Cabinet Office Strategy Unit Report 2002) that the focus on strategic risk would naturally evolve over time as the operational risks were identified, managed and reduced. In line with the Turnbull report the Department of Health required that Trusts sign a Statement of Internal Control that was to be produced alongside the

accounts of the Trust. To assist Trusts to complete their Statement of Internal Control the NHS produced two guides in the form of a framework (Department of Health 2002)(Department of Health 2003) These gave guidance to Boards on a method of demonstrating that they have been correctly informed regarding the totality of their clinical and non-clinical risks. It is the Board that is responsible for a Trust's system of internal control. Therefore it is the Board that sets the appropriate policies, and it is the Board that must seek regular assurance that the systems are functioning correctly.

Thus a Board must establish (and focus on) their main objectives in both an overall strategic direction and for each directorate. Then they must identify the main risks that could threaten their achievement. Having completed these tasks the design of their managerial system must be undertaken that would control and manage these main risks. After performing these high level tasks, the system of internal control must be embedded into the Trust's day-to-day operational activities and form part of the culture of the organisation. In this way behaviour is changed at all levels throughout the Trust, and there is less need for a plethora of controls to be forced onto operational managers (The Institute of Chartered Accountants 1999a p 19 and p 30).

Regulation is important within any society or organisation as it provides the scaffold through statute or guidance, within which public sector organisations including the NHS operate. Inspection, audit and the day-to-day monitoring functions provide assurance to the organisations senior managers and the public at large that the minimum standards are met. As such these three methods provide information relating to the levels of performance of an organisation's processes. This knowledge in turn helps to inform managers at all levels of their process efficiencies and overall effectiveness towards meeting their organisation's strategic aims and objectives. By feeding this data back into the policy makers forum, it helps to influence and guide revisions to the original regulations.

In 2003 the Audit commission published a further document discussing Corporate Governance. It stated that at the heart of many public service failures was poor corporate governance. The aim of the document was to investigate the relationship between the: 'quality of corporate governance arrangements in public sector organisations and the quality of services that they provide (Audit Commission 2003).

The report splits the elements of corporate governance into soft and hard systems. The soft areas are effective leadership and high standards of behaviour; whereas the hard areas are robust systems and processes. The commission goes on to argue that there is a direct relationship between organisations that have good corporate governance, and their ability to provide high quality services and deliver improvement. And that poor corporate governance creates serious service and financial failures. They then open the discussion into the areas of the social responsibility of organisations, the debate surrounding corporate manslaughter and the loss of public trust following an incident. It goes on to cite the Commission for Health Improvement (CHI) where this organisation found that, in a third of the organisations CHI had reviewed, there was a 'lack-of-connection', between the policies that the board had approved and what was actually happening in the wards, departments, etc.

The report then considers how regulation, inspection and audit contribute to the quality of corporate governance, in terms of decision making and information management, particularly focusing (in part) on the NHS. Recognition is given that the NHS has issued numerous guidance documents and promoted a programme of controls assurance. But still there is poor quality in relation to the information available and remains one of the key governance risks that the NHS has to overcome, before it can provide a robust basis for decision making. And that this was compounded by Trust boards not asking the correct questions, as the vast array of data was not always presented in a way that was meaningful to the organisation. This lack of clarity underpinning the decision making processes was not assisting in achieving effective accountability. And that most auditors reported a lack of confidence in the reliability and accuracy of the information reported. The report claims that this was found to be so badly managed that in 2002 CHI urged that action was taken in 198 published clinical governance inspections, as they found (amongst other issues) that half of the boards did not receive information from service activities.

1.4.2 Inspection

The role of inspection is discussed as it provides a framework to help ensure that minimum acceptable standards are met. This is an important part of the accountability process as it provides information (evidence) to boards, public and stakeholders. One of

the problems that the NHS is facing is the plethora of external agencies requiring information for their audit/inspection needs, Hence the call for integrated audit approaches. Also counterproductive is the centralist command, control and target setting culture. But the report recognises that total delegation to the 'coal face' is unrealistic therefore a compromise must be reached.

The report goes on to state that: 'Such assessments will become even more important as new organisations , such as foundation hospitals and children's Trusts are created and new public-private partnerships are entered into'. And that common governance challenges are the effective use of information together with the balance between local and national priorities, but that the infrastructure to deliver this is within a highly complex and taxing environment. Therefore, inspection needs to be integrated into a common coherent framework to create clearer accountability and greater assurance. These inspections must alert organisations not only to the judgments relating to past performance, but also to likely future events and the impact of problems that they will probably face. Together risk management techniques and internal control of managerial processes provide assurance to an organisation's managers that their objectives will be achieved (or not). These comments by the national audit office follow their observations that the financial position within the NHS is an area of great risk and that auditors continue to be concerned with the poor links between financial, management and service planning. Also performance management was identified as a risk for more than four in five health bodies.

1.4.3 Monitoring

Within NHS Trusts the creation of an audit committee is a statutory requirement. Underpinning this should be an assurance framework that allows for service commissioning, health inequalities, service provision, financial balance, partnership working, core standards, performance targets, developmental standards, workforce, patient involvement and effective governance. This should be complemented with escalation procedures for each performance level should they be breached. Thus a detailed assurance framework should be in place for each of the Trust's strategic objectives linked to a resource plan, with a named individual held accountable for its delivery. The Trust should also have in place a risk register for strategic and

departmental level risks leading to the creation of an overall Trust risk profile. This should be accompanied by action plans to manage these identified problems.

All of the above should form part of a Trust Risk Management strategy that engages the service uses, which ultimately are the stakeholders. The strategy should describe how the risks are identified, evaluated and managed in a structured way. From the various processes and committees within the Trust, there should be a reporting line through to eventually the Trust board.

1.4.4   Failure of good governance

The Audit Commission document goes on to discuss various inquiry reports, which show that poor governance arrangements, systems and processes failed to detect or anticipate service failures. The report cites various, 'warning signs', and that it was from an ignorance of these alarms that a number of serious service failures have occurred. These warning were in relation to:

❖ **Poor leadership**

The leadership of an organisation sets the 'tone' of the governance culture be it 'bad' or 'good'. Bad leaders create an air of self-delusion and can sometimes deliberately misrepresent information.

❖ **Poor information**

Information was accepted as being correct and was not challenged by management, this led to flawed decision making.

❖ **Lack of clarity**

Accountability arrangements were unclear and therefore weak. Policies, staff performance, unprofessional practice, etc were allowed to persist unchecked.

❖ **Poor working relationships**

These were exasperated by ineffective communication between professional groups across organisations supported by poor organisational structures.

❖ Closed culture

These consisted of inward focusing organisations with often hostile relationships with external departments/groups, which did not keep up-to-date with recent developments and did not adopt new working practices or techniques.

Corporate governance has been described as the brain and nervous system of an organisation, which when working well receives and generates information, which in turn enables action to be taken to anticipate danger, move to a more secure environment, learn from the experience and to progress forward. Systems that are not working well do not survive in nature and are eliminated. In designing a Trust's internal control methods HM Treasury suggested an appropriate managerial tool was to firstly identify its risks by either commissioning a risk review or undertaking a risk self assessment (HM Treasury 2004).

The foregoing led to the founding of the Health Care Standards Support Unit in 2004 to provide support to the then new Department of Health initiative on standards. Ellie Scrivens (Professor of Health Policy at Keele University) was appointed as the Director of Health Care Standards Unit for the Department of Health (DOH) to advise the DOH and NHS on standards development and use. With the creation of the Health Care Commission a new set of standards was developed. These were unusual in that they described the elements of a health care system as opposed to specifying a set of processes. They focused on the key mechanisms for safe healthcare, e.g. clinically effective; patient centred healthcare; good governance and accessible and responsive care. The standards were developed to reflect policy but also to allow for local conditions and constraints to be considered in their compliance and delivery (Scrivens p 123).

1.4.5   Creation of an Assurance Framework

Therefore fundamental to improving care is the development of robust managerial systems designed, implemented and monitored (audited) by multi-disciplinary teams. This can only be achieved by project teams that have been specifically trained in the various managerial techniques needed for the task, and who are led by competent managers capable of operating across customary departmental boundaries. When this is done then errors (latent conditions) will be uncovered. When these gaps in the managerial systems combine, they can produce a catastrophic incident. Therefore these

gaps must be identified, graded for severity and systematically closed across the whole system for such events to be avoided.

As control exists to reduce risk to an acceptable level, then it is necessary for an organisation's operational procedure to control its managerial systems. For assurance that these systems are adequate and working it is necessary that they are audited. The driving force behind the recent approaches to audit has been the desire to design a method that systematically appraises the whole system of an organisation. This includes both the formalised policies and procedures, and the informal norms of behaviour. Therefore Boards and their internal auditors are charged with establishing if their control systems are effective in order to give the required assurance to their stakeholders (NHS Audit Committee 2005).

In complex organisations such as NHS Trusts, the senior managers of a Board cannot know all of the processes and procedures practiced by the various departments. Therefore robust and transparent managerial systems (controls) must be developed, implemented and audited so that the Board can demonstrate that they are achieving their objectives to the required national standard, in a robust and continuous process. It is these systems that provide the assurance that a Board is doing its 'reasonable best'.

In attempting to define the statement 'reasonable best' this thesis recommends that reference could be made to the Corporate Manslaughter and Corporate Homicide Act 2007 that came into force on Sunday 6th April 2008 (Parliament 2007) Currently there is much discussion within the legal profession and industry itself concerning the impact that the new Act will have, and the way the courts will interpret what is reasonable. It is too early to quote any case law as guidance; therefore discussion must rely on past experience with existing legislation.

The new Bill significantly departs from the old law relating to corporate manslaughter; it looks at the collective responsibility of Trust's senior managers in place of attempting to identify a single individual responsible for the killing; thus overcoming some of the major failings of the old legalisation. Also it focuses on the governance of the day-to-day operational practices of the Trust's managerial systems.

The problem for the NHS is that (from its own figures) as an organisation it is responsible for approximately 15,000 deaths/year associated with an annual bill of £2b, most of which is considered to be preventable if good managerial systems of governance were in place (Department of Health 2000). As the responsibility for a Trust's managerial systems rests very firmly on the shoulders of its senior managers, it is probable that these people will be at direct risk of prosecution.

1.4.6 Need for reform

In the past, in order to prosecute a large company with complex management structures it was necessary to identify an individual at the top of the company's management organization that could be shown to have a 'directing mind', i.e. a person who embodies the company in their actions and decisions. This was known as the 'identification principle'. Only when this person could be identified and were themselves guilty of manslaughter, could a company be successfully prosecuted. It has been the identification principle that has led to the lack of successful prosecutions against large organisations.

The new Bill removes the need of the 'identification principle' to secure a conviction, by replacing it with a focus on the collective management failures at a senior level within a company. The new offence relates to the failures in the way that the organisation has been managed or organised, referred to as 'management failure'. This places the spotlight on the working policies and practices of the company as opposed to an individual acting negligently. The offence is designed to reflect the corporate failings by the senior managers (either individually or collectively) in relation to risk across the organisation as a whole, rather than concentrate on local issues.

As the offence is based on what an organisation has failed to do, the Bill gives a framework that allows it to be assessed against national standards, and a decision taken regarding the actions of its senior managers. It is this management failure that lies at the core of the Bill. This is interpreted as meaning that organisations are not liable under these proposals for the unpredictable rogue activities of its employees, or operational negligence causing death, but in the way that particular activities were managed and/or organised.

25

Where an organisation is found guilty of corporate manslaughter, individuals could also be prosecuted under existing health and safety law.

1.4.7 The offence

Section 1(1) of the act states that a Trust would be would be, guilty of an offence if the way in which its activities are managed or organised:

(a) causes a person's death, and

(b) amounts to a gross breach of a relevant duty of care owed by the organisation to the deceased.

The offence is targeted at the most senior managers of an organisation who play a strategic role as opposed to those managers at the more junior levels as defined within the Act in section 1(4) (c). The definition of a senior manager relates to a person that makes decisions regarding the organisation as a whole or substantial part of it, and plays a significant role. The word 'significant' captures those in a decisive or influential position, rather than a minor or supporting role. This research regards an Estates Manager as such a senior person, as they have managerial authority and take decisions over multi-million pound budgets, and are technically responsible for their departments. A major contributor to this definition when applied to an individual would be from reference to their job description (McShee 2007).

The offence is only to be applied in cases where 'gross failure' that causes death is proved. The term 'gross failure', is defined in terms of conduct that 'falls far below' what can reasonably be expected in the circumstances. It is not the intention of the bill to be used against organisations that are making proper efforts to operate in a reasonable manner where appropriate standards have not quite been met. This meaning is addressed within section 2 of the Act and explicitly includes reference to the subject of 'construction' or 'maintenance operations'.

In deciding if the breach of a 'duty of care' was a gross breach and that an organisation's conduct fell far below what could be reasonably be expected in the circumstances, a jury would be expected to consider two main issues. Firstly, how serious the failure was, and secondly, how much of a risk of death the Trust's failure posed. The jury may also

consider the culture of the organisation and its policies and working practices from reference to section 8(3) of the Act. McShee (2007) argues that the definition within the Act causes concern, as it leaves the jury to assess corporate culture which is subjective and not definitive.

1.4.8 NHS standards/working practices

The Healthcare Commission has issued a series of core and developmental standards that NHS Trusts will be assessed against. These audits are designed to act as an annual health check to assist in answering two questions: -

1. Is the organisation getting the basics right?
2. Is it making and sustaining progress?

At first sight, it would appear that if a Trust were charged with corporate manslaughter, then a reasonable defence would be to cite the Healthcare Commissions audit results. The problem with such an argument is that: the two questions are (by necessity) part of a very high level assessment of an organisation that is a statutory body in its own right. A Trust's results form part of a nation wide exercise to inform central government agencies in relation to strategic NHS policy at a national level.

Thus; they do not probe deeply into the detailed operational policies and procedures that are actually being practiced on a day-by-day basis by the Trust's staff. In order to answer the Healthcare Commission's two questions their approach is based on the central principle that (quote): "it is the responsibility of Trust boards to satisfy themselves that they are meeting the core standards and, where this is not happening to take appropriate steps to correct the situation".

In a similar way, the NHS Litigation Authority has issued a set of clinical standards that Trusts should follow (NHS Litigation Authority 2008). Whilst these standards tend to be more operationally biased in their structure, they are never the less still generic overviews of what is desirable. Therefore in a similar manner, relying on the defence of conformance to these clinical requirements on there own will still not demonstrate the detailed day-to-day actual operational practices and culture of an organisation. Similarly the now redundant Controls Assurance Risk Management Standards were only a high

level guide, that needed detailed assessments to unpin there respective levels of compliance statements.

This Thesis suggests that citing either/or both the Health Commission's and/or NHS Litigation Authority's audits as a defence would be unacceptable, as they are not designed to uncover the latent conditions within a Trust's day-to-day managerial systems that are generating unacceptable levels of risk. The only true measure of compliance is the detailed review of the operational systems when compared to the national codes of practice for each task or group of related tasks within each department. It is this type of evaluation that managers of all grades are fundamentally employed to undertake, not as part of an annual audit, but on a constant day-to-day basis.

Any gaps within the managerial systems must be assessed in terms of severity as part of a multi-disciplinary approach, resource needs estimated to close the gaps both immediately and ongoing revenue consequences and, the results conveyed up to the senior managers of the Trust, on a periodic basis. Accompanying this must be reasonable proof that: given the deficiencies within the management systems, the operational and senior managers are putting their existing limited resources to where they are doing most good.

If an organisation cannot demonstrate this level of commitment to the identification, prioritisation, management and audit to this level of detail then, this Thesis suggests that accusations of both 'gross failure' along with conduct that 'falls far below' what can reasonably be expected are likely to be made against the organisations senior managers. Therefore section 8 would appear to be indefensible in these cases.

Assuming that a Trust was following the national standards of good governance as promoted by government and NHS stakeholders (the general public), the Board and other senior managers within the organisation should: know of the organisation's failure to comply with legislation or guidance, and be aware of the impact posed by the failure to comply. Scrivens openly implies throughout her book, 'that there is still much work to be done before the NHS can with comfort and confidence state that it is conforming to systems of good governance' (Scrivens 2005). Within her recent book, for which she has

28

received the internationally recognised Baxter Award (Keele University 2005) for an outstanding publication and/or practical contribution to excellence in healthcare management in Europe, her closing sentence calls for greater research into audit and inspection (Scrivens 2005 p164). The remainder of this thesis will serve to uphold this view.

In February 2006 a handbook was published by the Department, to ensure that the basic building blocks of integrated governance were in place for implementation and roll-out between 2006 and 2008 (Department of Health 2006). The handbook confirmed that integrated governance did not supersede clinical, financial or any other governance but highlighted their importance, inter-dependence and inter-connectivity by acting to co-ordinate all governance domains. The thrust of the handbook is to define the rolls of executive and non-executive directors of Trusts, and the various committees that serve the Trust Board. Central to the strategy is the design, implementation and management of a robust and transparent Assurance Framework.

1.5   The research question

The Department of Health issued a report that highlighted the inadequacies of the systems of internal control within the NHS (Department of Health 2000). Although the report did not make direct reference to any managerial failures within NHS maintenance departments, the document was published against a backdrop of the research work of Perrow into why managerial systems fail and the numerous high profile engineering managerial system failures he cited (Perrow 1999), and the studies of Reason into the failure of managerial systems (Reason 2000).

Following the publication the researcher undertook a series of informal discussions with a number of experienced estates managers to discuss the possibility that there may be unknown latent conditions within the managerial systems of NHS maintenance departments that could affect patient safety. Ten participants agreed to attend an unstructured meeting to discuss the issue, which was chaired by the researcher who acted in a process consultation role to the group (Schein 1997, Schein 1998, Belton and Stuart 2003). There were three outputs from the meeting. The first was the development of the research question that this thesis addresses, which is:

'Do the managerial systems of internal control within the National Health Service (NHS) assure that the day-to-day building service engineering maintenance activities conform to their respective national guidelines?'

The second was the collective agreement to the question, 'Which areas under the direct control of the Estates Department are believed to be currently generating the greatest unaddressed perceived risk?' To answer this second question a generic structure of an estates department was designed by the researcher using his +30 years experience as an NHS estates manager, and presented to the group. Through face-to-face discussion the group 'informed the generic structure' and created a unanimous answer; it was those systems that have never been scrutinised by a managerial and technical audit during the normal course of events. These were the policies and procedures of eight engineering topics forming the very core of their activities.

The eight engineering managerial systems identified by the focus group as probably generating the greatest latent potential for developing into major risks are presented below:

1. Medical Gas Pipeline Systems

These installations feed directly to patient areas via distribution pipe work from central plant rooms. They normally supply oxygen, nitrous oxide or generate a partial vacuum. Prior to 2006 the NHS termed these systems as Piped Medical Gas.

2. Water

All water for use and circulated within a hospital distribution system is potable and is susceptible to water borne organisms including Legionella.

3. Lifts

Under this categorisation are all patient and goods lifts, including patient bed hoists.

4. Ventilation

Ventilation systems range from normal ducted air ventilation through to fully conditioned air and theatre quality air and extract systems.

5. Building Management Systems

These are computerised systems that monitor and sometimes control various building services plant.

6. Pressure Systems

A pressure system is defined under statute and governed by the Pressure Systems Safety Regulations (Health and Safety Commission 2000). In the case of a hospital it is not unusual to find main steam raising boiler plant and calorifiers.

7. Fire

This includes all fire detection devices, alarms, smoke control mechanisms, extinguishers, etc.

8. Electrical (LV)

It is common for hospitals to be fed via an 11KV (high voltage) supply, transformed down within the hospital site and distributed at 415/240V (low voltage) to the point of use within buildings.

In all cases the management system surrounding the high voltage supply was never seen by the focus group as posing a major problem because of the well trained and competent engineers in charge of the installation, coupled with the intense maintenance regime governing the plant and equipment. This was in stark contrast to the low voltage distribution system.

The third output formed as the participants discussed what should be done if the answer to their initial research question was:

> 'The managerial systems of internal control within the National Health Service (NHS) **do not** assure that the day-to-day building service engineering maintenance activities conform to their respective national guidelines.'

The meeting agreed that they needed to know as senior managers what managerial tools should be applied to: assessing the resources needed to close the identified

31

managerial gaps, what would be the revenue consequences of keeping these gaps closed, and finally how could they give assurance that the gaps remained closed.

Following the development of the research question and the identification of the most probable topics believed to be generating the greatest unaddressed risk; the researcher disbanded the meeting and began the research project by undertaking a detailed literature search.

Although the methodology employed to create and manage this initial meeting was totally informal, it did lay the foundations for this research project. The results and conclusions drawn during the following years have shown that their initial concerns were valid.

Chapter Two   Literature Review

2.1   Introduction

The previous chapter mapped the history of the NHS to the present day, where the service is now assessed against its conformance to national standards. It concluded by suggesting that the majority of NHS Trust estates' day-to-day maintenance strategies do not conform to these standards.

This chapter begins with a short description of reliability and risk, and the balance that must be achieved between these two topics when designing a maintenance strategy. The Department of Health's national guidance relating to the formation of a Trust's maintenance strategy issued in the year 2005 is discussed, and compared to that published in 1978. The term 'maintenance' is examined along with the use of computers within maintenance departments. The type of maintenance strategy to be adopted is considered, and the assumption that plant and equipment failures are totally independent of managerial failures is discussed,

The creation of a suitable managerial system is then debated along with the two main techniques used to assess the risks associated with such a design. A research model relating to an inherent failure of a management system due to the quantity of non-novel interruptions during normal day-to-day activities is examined.

Finally the need for and development of an assurance framework is discussed, along with the call for an integrated monitoring and audit programme to be implemented.

2.2   Reliability and risk

The use of the word reliability was first recorded in 1816, and was used by Samuel T. Coleridge in a letter to a friend (Saleh & Marais 2006). Since this date the word has come into general use. Its popularity to-day can be shown by entering the word into the search engine Google; this will produce over 9 million hits on the World Wide Web. With the extended use of statistics and adoption of mass production the discipline of Reliability Engineering was formed.

Society expresses its reliability 'wants' in the form of environmental standards and regulations, which reflect international, national and organisational needs. Most industries are complying with these requirements by designing and installing plant and equipment (assets) that match these needs. However it is not sufficient just to ensure that the assets are compliant at the point of commissioning and bringing into use. Managerial systems must be implemented to ensure that the assets continue to comply with their performance specification throughout their operational lifespan through to eventual disposal. Therefore the reliability (i.e. the probability of failure) of these managerial systems and assets is of profound importance, as failure is inherently linked to risk.

Risk is the product of the probability of an event occurring and the impact that that event will cause. There are basically two ways of discussing risk, classical or Bayesian. The classical view is a positivist view. This claims that risk can be measured and therefore recognises only that which can be scientifically verified. Alternatively the Bayesian view is a way of expressing uncertainty and therefore lies between the two extremes of positivism and relativism (where knowledge is not absolute) (Aven & Kørte 2003). The risk relating to managerial aspects does not have a history of quantitative data collection in the classical sense, leading to quantitative risk assessment, but is assessed by a multi-disciplinary approach adopting qualitative techniques that take the assessor's judgements into account. The balance between reliability and risk for the physical assets of an organisation will directly influence the maintenance strategy adopted.

2.3 National guidance for maintenance within the NHS

As stated within Chapter One, the first comprehensive guidance documentation relating to a maintenance strategy within the NHS was published in 1973 by the Department of Health, and entitled Estmancode (Department of Health and Social Security 1973).

It provided a framework for regularly reviewing maintenance performance at all levels, preferably by a multi-disciplinary team. The guidance called for the development of 'acceptable local standards of efficiency and safety' whilst seeking 'the advice of medical and other interests', in order to determine the local acceptable standards. It also recognised that reasonable flexibility must be employed to reflect the available annual budget in deploying limited resources to where they would do most good.

The framework developed called for the assessment of priorities, comparisons of costs between NHS staff and contract labour, investigations into the cause of plant and equipment failures, etc, All this work was to be under a continuous review of budget and performance conducted at local and national levels, and various points' in-between e.g. regional, area, district and sector levels. Horner termed this approach as a 'maintenance strategy'. (Horner, El-Haram & Munns, 1997).

Throughout the entire 1973 guidance documentation the word 'risk' was never used, as this did not begin to be applied within the service until the mid 90's. However from studying 'Estmancode' the basic building blocks for modern healthcare maintenance management practices were laid, as the document called for a combination of maintenance systems and managerial tools to be used that best suited local needs.

2.3.1 Maintenance strategy

In the year 2005 the Department of Health issued a set of national standards that emphasised local actions, local ambition and local innovation. Healthcare organisations were expected to be creative and innovative in forming new local solutions by challenging past practices when complying with these standards (Department of Health 2005) The use of standards has moved the NHS away from a service motivated by national targets to one where standards are the main driver. These describe the level of quality that health care organisations will be expected to meet, be they NHS Trusts, based in the voluntary sector or privately funded. An organisation's performance is now assessed against these standards and all judgements on performance are evidence based. The standards are split into seven domains:

1. Safety of patients
2. Clinical effectiveness and cost effectiveness
3. Governance
4. Patient focus
5. Accessible and responsive care
6. The care environment and amenities
7. Public health

Each domain is split into two categories defined as: -

Core standards: which bring together and rationalise existing requirements for the health service, setting out the minimum level of service patients and service users have a right to expect; and

Developmental standards: which signal the direction of travel and provide a framework for NHS bodies to plan the delivery of services that continue to improve in line with increasing patient expectations.

The core standards do not create any new guidance material for departments to follow but are based on standards that already exist. The national standards documentation states, **'Meeting the core standards is not optional. Health care organisations must comply with them from the date of publication of this document'**, (Department of Health 2005 p 24). This statement sets the maintenance strategy for all health care organisations, and takes account of both the safe operation of the physical engineering infrastructure when assessed against national codes of practice, and also provides an assurance framework to control and monitor the managerial activities associated with maintenance.

The statements above suggest health care organisations are placed in a 'strait jacket' in terms of the design of their technical and managerial systems, as compliance to the national standards is compulsory. This appears to be at odds with the previous statements made in 1973, where they were allowed to design their own maintenance strategy. This assumption is incorrect.

The national technical standards are flexible, in that local conditions are allowed to influence any guidance recommendations. It is for the local organisation to determine the maintenance necessary to comply with the sprit of the national guidance, so long as the basic premise is held that the systems are 'fit for purpose' and that the Health & Safety at Work Act 1974 is complied with. This requires that engineering systems are maintained to a reasonable standard. Therefore health care organisations are free to explore and adopt the various technical and managerial techniques that have been developed over the years to maintain their plant and equipment. Thus each organisation can develop their own maintenance practices to reflect their particular local environment. The issue of the National Standards by the Department of Health in 2005 has not altered

the 33 years of pre-existing guidance in relation to the maintenance strategy of the engineering infrastructure of the NHS healthcare buildings.

The Department of Health's section charged with developing, updating and issuing engineering guidance for use within healthcare premises is known as Estates and Facilities Management. The relevant national guidance is produced as a series of Health Technical Memoranda (HTM's).

2.3.2 Type of maintenance

Maintenance within the NHS estates function is categorised as either backlog or day-to-day maintenance. Backlog maintenance relates to the planned need for a major replacement of a healthcare building's 'as fitted' infrastructure due to normal wear and tear over the years, e.g. the replacement of 40 year old corroded steam pipes.

The scope of this thesis is concerned with the day-to-day maintenance of the 'as fitted' engineering infrastructure of a healthcare building. It does not address the specification, design and process of installation, commissioning, modification or replacement of these systems. The areas under examination are the day-to-day maintenance activities. These are defined as: 'The combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform a required function' (British Standards Institution 1993).

The activity undertaken in relation to day-to-day maintenance within an NHS healthcare building since 1973 was split into two distinct categories. Within the service these have been known as reactive and planned preventative maintenance. These categories are still in use to-day and are described below.

2.3.3 Reactive Maintenance

Reactive maintenance relates to small day-to-day breakdowns, where in the main the impact associated with an item failing is considered negligible. This type of maintenance strategy is associated with phrases such as, 'don't fix it till it breaks'. Research conducted by Al-Zubaidi into reactive maintenance within the NHS gives examples of this type of work (Al-Zubaidi 1997), and includes repairing blocked sinks, and leaking taps, or replacing failed lamp bulbs, etc. All healthcare buildings generate a reactive

maintenance profile. The larger premises create a day-to-day reactive workload that requires a team of staff from various trades and support managers to respond to these needs.

2.3.4 Causes of and degrees of reactive maintenance failure

An engineering item can fail in many ways, and is classified in terms of cause, time, degree and/or combinations of failure. The relative sub-classifications of failure are self explanatory and are: misuse or inherent weakness failure; sudden or gradual failure; partial or complete failure; catastrophic or degradation failure (Drummer and Winton 1974).

How often these failures occur is defined in two ways: mean time between failures (MTBF) and mean time to failure (MTTF), where the former relates to repairable items and the latter to those that are non-repairable.

The time-honoured view was that a batch of like items would have a high rate of failure early in their use (known as the burn in period), followed by a steady low constant rate of failure, ending with a wear-out period where the failure rate increases. This simplistic picture formed the traditional 'bath tub curve'. With the modern-day designs incorporating pneumatic/electro/mechanical devices manufactured to tightly controlled tolerances this curve is no longer applicable (Moubray 1997).

2.3.5 Planned Preventative Maintenance

This type of maintenance was designed to give assurance that plant and equipment was functioning to its design intent, and that some plant and equipment breakdown could be prevented, thus causing less disruption to the service provision. In 1973 technology was not as advanced as it is to-day. There were no desk top computers, calculators or hand held diagnostic devices; slide rules were the norm! Estmancode introduced a basic planned preventative maintenance system into the NHS for local officers to adapt to suit their specific needs. This used a system of regular inspections at set frequencies of time, for example - weekly, three weekly, six weekly and their subsequent multiples, through to multi-year frequencies (Department of Health and Social Security 1973).

Over the intervening years since this form of planned preventative maintenance was first practised within the NHS, industry has adopted other maintenance systems in an attempt to reduce overall costs and disruption to production. These systems have evolved two main themes, condition monitoring and total productive maintenance. Condition monitoring is sometimes known as 'predictive maintenance', and uses technology to monitor the 'health of plant' on a continuous basis to detect any abnormalities in an attempt to prevent breakdown or malfunction. Total productive maintenance (TPM) encourages a team working approach to the subject of maintenance and brings together the various stakeholders. However both themes follow the basic principles that were outlined within Estmancode back in 1973.

Due to a combination of suitable skill shortages, the increasing costs of repair and loss of production through equipment/plant failure, the reliability of equipment/plant (availability) is now more important than its maintainability (Sherwin 2000). Technology has responded to this need. Now it is common for plant and equipment to be designed so as to require minimum invasive maintenance. Also with the progression of computerisation, self diagnostic and/or remote monitoring/sensing systems have become the norm. Therefore since 1973 the need for dismantling parts of the equipment infrastructure to ensure that it is performing to the required standards is much reduced, and the tasks performed have nowadays mainly been condensed to examination and testing activities.

2.4   Computers

The use of computers within NHS maintenance departments is widespread and is now a central component of their managerial systems. In the 1980's the NHS began to develop a national computerised system for the management of its maintenance activity. This was known as Works Information Management System (WIMS). But after a number of years this was abandoned, and now there is no standard NHS maintenance software. Now each Trust must purchase its computerised maintenance management system (CMMS) from various specialist commercial organisations. These systems give managerial support across a number of topics namely: -

❖   Records reactive maintenance data

* Produces pre-planned schedules of tasks (planned preventative maintenance and Monitoring)
* Budgetary accounting of the maintenance function
* Movement of spares

All of the CMMS programs give facilities for data collection, with some expensive systems allowing for real time data logging. Yet despite the complexities of analysing this information in terms of plant availability and reliability, virtually all of the commercially available CMMS's lack any decision making capability to assist the maintenance manager to reduce costs and/or risk, and in some cases the systems function as no more than a diary (Labib 2004).

2.4.1 Maintenance simulation models

The overall aim for a maintenance manager is to optimise the maintenance function against set criteria applicable to an organisation's stakeholders. From the application of scientific techniques to determine this optimum level of maintenance, the field of Operational Research was born, and applied to maintenance during the Second World War. During the intervening years various mathematical models have been developed in an attempt to determine the optimum maintenance level that could guide an organisation towards this in effect theoretical target. The words 'lead towards' were used by Sherwin to emphasise that the optimum level was a moving target continuously influenced by new data/environments (Sherwin, 2000).

A modern acute hospital contains many thousands of different assets each made up of various different components, which between them exhibit differing reliability behaviours. Although the size of the NHS as a whole is extremely large, the sample size from which to conduct any form of actuarial analysis is small and restricted to individual hospitals only. This is due to past government policies that were discussed earlier within the introductory chapter, which discouraged the sharing of information between Trusts, and the abandonment of a centrally controlled national maintenance software programme for day-to-day maintenance.

In order for these mathematical models to function they must have access to a large data base of historical information relating to critical and non-critical failures. But as a

critical type of failure has serious consequences e.g. loss of life, then this body of data will not exist as preventative action would be taken immediately after the first incident. In the majority of cases if any non-critical data was available it would reflect minor consequences only. From this situation Moubray (1997) argues that it would not be cost effective to subject this information to actuarial analysis and detailed preventative maintenance action. Conversely this final point is argued against by Sherwin (2000), who states that savings however small all add up and therefore should not be ignored. I suggest that an open mind should be taken by the maintenance policy designer and that environmental conditions must dictate what is important and what is not, rather than following the wide sweeping statements of either Moubray or Sherwin.

Simulation models for maintenance need copious amounts of reliability data before they can be applied to a given situation. Unfortunately there is more often than not a complete lack of such suitable information. Thus although these models are academically sound, in the majority of cases they are of no practical value when placed within a normal working environment (Rausand 1998, Moubray 1997, Sherwin 2000).

2.4.2   Simulation of workload

The day-to-day maintenance demands for an acute hospital have been analysed and found to be random in nature (Al-Zubaidi 1997). This research was later used to develop a simulation model of a hospital maintenance department's day-to-day activity (Al-Zubaidi & Christer 1997). The model succeeded in showing the benefit and applicability of using simulation modelling to assess differing management policies, but their work developed a complex program requiring very detailed information for it to operate. They state that they had to make subjective assessments as there was little useful historical data available.

A similar study was undertaken within manufacturing industry but this time both preventative and breakdown (reactive) work orders were simulated in an attempt to analyse the various factors affecting the personnel capacity requirement (Mjema 2002). The model produced showed that for a decentralised maintenance organisation, (such as that now created by recent Trust reorganisations), then by allowing staff to work across the boarders of their decentralised department greater utilisation of staff is possible and throughput of work orders increases. The main reason for this increase in

utilisation and throughput is the stochastic nature of the work order demand for maintenance activities. As Mjema points out, although preventative maintenance work orders are predictive, the combined preventative and reactive work orders must be considered as stochastic. From my experience I agree with Mjema, and as will be seen later within this thesis this is an important statement.

Within the simulation studies cited, all were analysing the day-to-day work load and by default assumed that the technical and managerial practices were being performed to the required standard. Also they have not investigated the effects of reduced manpower availability and the risk this generates from the reduction in capability to undertake planned preventative maintenance duties. Therefore from a practical viewpoint whilst the studies have succeeded in terms of their own objectives, the scope of their simulations is too narrow to be of any practical use in the study being undertaken within this thesis.

From the forgoing and the historical evidence reviewed by this research, actuarial analysis relating to the relationship between operating age and failure is of little use to the NHS day-to-day maintenance manager. Therefore a system is needed that focuses on the future rather than the past. Historical information relates to past events, as opposed to the concepts of anticipation and prevention which attempt to gaze into the future to control organisational risks. Consequently any decision analysis tool for the NHS day-to-day maintenance manager must be based on a projected assessment of risk and supported by a planned maintenance methodology that is a forward looking system as opposed to a historical perspective. Such a system was developed in the 80's and is known as Reliability Centred Maintenance.

2.5   Reliability centred maintenance

This concept originated within the aircraft industry and has been applied throughout other organisations as a whole for more than 20 years Rausand relates a system to be the overall managerial and technical actions of the maintenance activity, and goes on to reference a number of reports/textbooks and military/non-military standards that have been developed over the years using the reliability centred maintenance philosophy (Rausand 1998). If implemented correctly the reliability centred maintenance system will ensure that the 'inherent reliability' of the system is realised. It will not improve the inherent reliability of a system as this is only possible by its redesign. Moubray

42

summarises the reliability centred maintenance methodology by stating that it is: 'A process used to determine what must be done to ensure that any physical asset continues to do what its users want it to do in its present operating context' (Moubray 1997 p 7). As such the reliability centred maintenance method provides answers to the following seven basic questions:

1. What are the functions and associated performance standards of the asset in its present operating context?
2. In what ways does it fail to fulfil its functions?
3. What causes each functional failure?
4. What happens when each failure occurs?
5. In what way does each failure matter?
6. What can be done to predict or prevent each failure?
7. What should be done if a suitable proactive task cannot be found?

To develop the answers to these questions a sequence of steps must be followed that have at their core the Failure Mode Effect and Criticality Analysis system of interrogating a plant/equipment item and associated management activity (British Standards Institution 1991). The reliability centred maintenance process in its totality forces a structured analysis of the consequences of each failure mode and integrates the operational, environmental and safety objectives of the maintenance function. As such the system relies on data that is developed in a descriptive and qualitative way, as opposed to quantitative historical information.

Following the output of such a reliability centred maintenance exercise, a system of planned maintenance is formed. This consists of tasks carried out at a set frequency to either prevent plant/equipment from failing and/or to monitor the performance against preset criteria (tolerance limits). Within the NHS this planned maintenance can be further subdivided into three parts:

Insurance inspections
These are undertaken for such items as steam pressure boilers, lifts, etc. From the work undertaken they are strictly planned preventative maintenance tasks as they all involve dismantling various components of the plant/equipment so that a

through inspection can be undertaken of the parts hidden from view during normal operating conditions. The aim of the inspection is to prevent any catastrophic failure of the plant/equipment and subsequent loss of life. This type of activity accounts for approximately 6% of the planned activities for a Trust's maintenance department. The frequency and activities required for this part of the programme are dictated to the Trust by external specialist engineers employed by the Trust's insurance company.

Other planned preventative maintenance activities

In the main these relate to large plant items such as ventilation systems, which would effect patient throughput or comfort if failure occurred. This type of activity accounts for approximately 8% of the planned activities for a Trust's maintenance department. These tasks are governed by the local plant and equipment's design, age and operating environment.

Monitoring

This activity records the performance of plant/equipment against their set points and tolerance limits. Examples are the recording of domestic hot water flow and return temperatures, to ensure that Legionella does not breed within the system and contaminate/kill vulnerable patients. This type of checking accounts for approximately 85% of the maintenance programme. The frequencies for monitoring are recommended by NHS national standards with an allowance for small deviations to account for local situations.

Although the reliability centred maintenance process presents a framework to build the planned preventative maintenance system around, it does not account for why incidents/accidents occur due to the failure of good governance as described within the introductory chapter which cited the Audit Commission (2003) page 22 findings. Therefore this research explored the concepts of Normal Accident Theory

2.6   Normal Accident Theory

Heimann (2005) observes that managerial problems are often more complex than a simple two state theory of good (operating), or bad (not operating). As often organisations create two types of errors, I) implementation of the wrong policy or, II)

failure to act when action is warranted. The cost/benefit ratios associated with each type of failure vary from organisation to organisation and the circumstances at the time, for each decision. Therefore it has been important to develop a general approach to studying these two types of failure that affect an organisation's reliability and the resulting consequences. He goes on to state that controls to contain the reliability of type I errors often come at the expense of the reliability of type II errors. Put simply organisations struggle between efficiency and effectiveness, and those concerned with effectiveness will aim to reduce type I errors, while those concerned with efficiency will reduce the likelihood of creating type II errors.

A series managerial structure is more effective at preventing type I errors, as each component must agree to the action before it takes place. An example within the NHS would be that a Trust would pass a draft policy through various committees for agreement, before accepting the document as a definitive statement of its intentions of how to proceed. But such a structure whilst going a long way to prevent type I errors relating to the adoption of the wrong policy, actually (as Heimann pointed out) creates type II errors of non-action whilst waiting for a suitable policy to be accepted (Heimann 2005).

2.6.1 Reliability through better components

High-reliability organisations promote a 'culture of reliability' which comes from a desire to improve decision making at the component level. Thus the structure of an organisation must be considered if reliability is to be improved. Heimann (2005) splits an organisation's structure into two separate camps, the systems strategy and the component strategy. The systems strategy modifies or changes the location of the number of components. A components strategy on the other hand needs a greater performance from its components while keeping the basic administrative system intact. However it should be noted that these are not mutually exclusive and an organisation could employ a mixture of the two systems.

It is not uncommon for a Trust to be faced with budget constraints, such that departments may be subjected to cuts in their revenue budgets. In such a case the Trust may reduce the number of components in series, thereby increasing the likelihood of a type I failure. To compensate the Trust may implement a component strategy to increase

the reliability of the remaining components. The choice between strategies depends on the tasks that the organisation conducts.

Organisations that handle repetitive simple tasks tend to adopt a component strategy i.e. increase the reliability of the components. Here repetitive tasks follow well laid down rules and operating procedures. In a perfect world Trust estate maintenance departments should be working within these environments. Adverse incidents would be from a direct infringement of/or derived from deviation from, these well defined and agreed rules. That stated, incidents will always occur according to Normal Accident Theory as proposed by Charles Perrow in 1984. (Perrow 2004).

Gephart (2004) observes that Perrow's work has influenced research in a diverse number of areas including: sociological study of risk analysis, the management of risk at the National Aeronautics and Space Administration, the safe management of nuclear weapons, sense making and control in organisations with complex technology, the nature of technology, and the management of industrial crises.

Normal accident theory has had a major influence on the way of thinking about complex organisations operating within hazardous environments (Sagan 2004). He claims that Perrow deliberately sought to shake the discussions around the safety of hazardous technologies, and to bring 'organisation theory' into the debate. As he stated, "The world is not becoming smaller; it is becoming more complex and interconnected". The breadth of Perrow's influence can be assessed by the number of citations of his work, coupled with the wide ranging journals across numerous subjects within which his research is cited.

The term Normal Accidents was coined by Charles Perrow in 1984 in his now famous book wherein he discusses the risks and dangers in society. Due to the book's popularity it was reprinted in 1999 (Perrow 1999). He states that his research was motivated by the desire to better manage these high risk systems (ibid p. 3); although he concludes with the observation that accidents within these systems are inevitable, and that we cannot effectively manage these systems. And those serious accidents are a result of the systems, and not the individual components. Thus by better understanding of these systems we may at the least stop blaming the poor operative, and start to question the

system that allowed the problem to occur in the first place (ibid p. 4). Normal Accident theory carried a warning: 'in high-risk technologies where the systems are complex and tightly coupled, catastrophic accidents are inevitable over time'.

It was this realisation that catapulted Perrow to the initial attention of the world, and his belief that: "new dangers were rooted in the organisational structure of organisations". He focused on two fundamental points in his argument. These were the interactive complexity and tight coupling of these organisations that confound the simple minds of humans in their attempt to envisage, assess their hazards and manage the risks that they collectively pose. The use of the probabilistic risk assessment system by the nuclear, petrochemical, aerospace, etc, industries bares witness to this (NASA 2002). Probabilistic risk assessment examines the way seemingly unconnected simple components and incidents can combine to create a catastrophic and previously unimagined failure or as Perrow predicts: normal accidents.

2.6.2 Linear and complex interactions

Perrow (1999) asks, "What kinds of systems are most prone to system accidents?" (pp. 72). To answer this question he put systems into one of two classes, linear and complex interactions. The most common are linear interactions where their failures and effects are well known. A typical example would be an expected production sequence that was recognized by the operators. To put it another way, where there are linear interactions from one component within the design, equipment, procedures, operators, supplies and materials and the environment of the system, with one (or more) components that precede or follow it. He goes on to suggest that say within a linear system only one percent of components could lead to a complex interaction and cause an accident. Whereas that figure would be increased to ten percent of components within a complex system would create the accident. If there are only four components within a complex system the possible number of interactions is 12. But in a system of 400 components where ten percent (40) could network in a complex way, any one could interact with the other 399 in millions of configurations, since the potential interaction pathways increase exponentially. Thus the danger lies not in the individual components, but within the system that allows them to interact.

Interactive complexity is explained as arising from a number of components in a given location interacting, either deliberately or not, with others. In these situations unintentional interactions can develop that were not seen as part of the design, which are baffling to the observer and are non-linear. However when these systems are transformational they have greater complexity and are more prone to accidents.

Complex interactions would in the main, be those not intended within the original design. Perrow attempted to define these two types of interaction as shown below, but was not completely satisfied with the result:

> Linear interactions (linear interactions in an expected sequence) are those in expected and familiar production or maintenance sequence, and those that are quite visible even if unplanned.
>
> Complex interactions (complex interactions in an unexpected sequence) are those of unfamiliar sequences, or unplanned or unexpected sequences, and either not visible or not immediately comprehensible.
>
> Perrow 1999 p. 78

With all of the problems associated with complex systems the obvious solution would appear to be simple, design all of our systems to be linear. Unfortunately complex systems are inherently efficient, as they have less slack, less tolerance of low quality performance, and allow for more multi-function components. Therefore complexity is desirable. Also when operating within a healthcare environment there is little choice in the design of the overall system in terms of complex or linear.

The major dimensions of a system have been described as complex or linear, but there are two other characteristics envisaged that are reasonably independent of the former two, and are split into tight coupling and loose coupling.

Tight coupling is described as the situation where system components having no 'slack', 'buffer' or 'give' (Ibid pp. 89 – 94). As an example, within a hospital operating room the surgeon's ability to perform the operation is tightly coupled to the decontamination of the surgical instruments. Any delay in the decontamination process subsequently delays the

48

operation because there is no slack in the system, no buffer in the form of a spare instrument pack due to the high costs of this equipment. Thus tight coupling is time-dependent, and intolerant of variations, as only one methodology can be employed to complete the job, as all tasks must be completed sequentially. Resources cannot be substituted for each other and failed equipment creates a shutdown.

Loosely coupled arrangements are the reverse of tightly coupled systems. In these systems time delays are tolerated, and little (if any) dilapidation occurs. Sub-quality components can be substituted to make-do, and tasks can be completed out of sequence.

The recovery situation after a failure has occurred from each of these two different structures is also different. In the tightly coupled system the system redundancies and buffers must all be built into the initial design, and thus be thought through well in advance of the inevitable accident occurring. In loosely coupled systems, again the reverse is true, substitutions can be made and off-the-cuff solutions found that can bodge the system until the correct processes are re-established.

Perrow describes how normal accidents are created as a result of, 'unanticipated interactions of multiple component failures'. He then goes on to demonstrate and describe through numerous examples, how high risk systems that were tightly coupled and experiencing unanticipated interactions have caused catastrophes, particularly in poorly run organisations. In these cases of poor regulation, poor quality, poor training, etc, there is more chance of failure within the areas of: design, equipment, procedures, operators, suppliers and materials, and environment components. Therefore in these cases there are two causes of failure, the component type and the managerial type. Such failures can interact in unexpected ways particularly in badly run organisations, as the rate of such failures would be much grater than would normally exist.

From studying the arguments made by Perrow (1999 pp.88 table 3.1) relating to the differences between linear and complex systems the researcher reconstructed parts of his work to form Table 1 to highlight these when applied to the field of maintenance within a healthcare environment.

49

Table 1 *Complex vs. linear systems relating to a maintenance environment within a healthcare setting.*

Complex	Linear
Proximate maintenance tasks	Segregated maintenance tasks
Many common-mode connections of components not in sequence	Limited common mode connections
Limited isolation of failed components	Easy isolation of failed components
Personal specialisation limits awareness of interdependencies	Less personal specialisation
Unfamiliar or unintended feedback loops	Few unfamiliar or unintended feedback loops
Many control parameters with potential interactions	Control parameters few, direct and segregated
Indirect and influential information sources	Direct on-line information sources
Limited understanding of some processes	Extensive understanding of all processes

Perrow defined the act of maintenance as following a linear interaction with its components. Although there exists a crossing of interpretation boundaries when analysing a particular function between complex and linear, the researcher disagrees with his suggestion when applied to the healthcare environment. Within a hospital the departments operate in a complex interaction with other departments, in a tightly coupled high risk environment, where one section of the organisation depends on the output of another in a timely and controlled manner to exacting pre-described and agreed standards. Failure of any department (component) to perform can lead to catastrophic accidents.

As previously discussed, where departments are tightly coupled, failures and/or delays in one section of the organisation can spread throughout the wider system in a domino effect, as independent time dependant processes and tasks that rely on specialised staff or equipment are compromised. These inevitable failures can be affected by the extent of the organisations potential to transform what is a risky business, into a catastrophic failure, i.e. a normal accident as defined by Perrow.

To better understand where engineering maintenance sits within the healthcare environment, and how it can contribute to a catastrophic failure with loss of life, a description of an operating department is presented to aid the discussion.

2.6.2 Operating Department

This is a designated area where the planning, preparation and provision of an environment is created for surgical care within an operating theatre. The building design and engineering services create a space within a hospital that is unlike all others. It relies on its own specialist lighting, electrical supplies, air conditioning, piped medical gasses and extract ventilation. It is supplied with specialist drugs, equipment and instrumentation, used by highly trained clinical and surgical staff and is the designated place for surgical procedures.

In order for a theatre to function it relies on the complex interaction of communication, exchange and transfer of resources with a number of other hospital departments. These in turn form close tightly coupled interactions with other departments both within and external to the hospital, all undertaking a series of specialised tasks. Some of these departments (but certainly not all) are described below.

Surgical wards are where patients are prepared before surgery and are attended to afterwards. The sterile services unit wash, maintain, decontaminate and prepare the surgical instruments. Histopathology analyse biological specimens at the pre, inter and post patient operation. Radiology takes images of the patient, again pre, inter and post operation. The High Dependency Unit is where high risk patients receive immediate post operative intensive care.

For the operating theatre to be effective and safe all of the departments must function satisfactorily, both with regard to their designated tasks, and in their ability to interact with the theatres. Waring, McDonald and Harrison (2006) found that a major role for the managerial support team within the operating theatre departments was to co-ordinate all of the activities delivered by the support departments for each surgical procedure. They observed that each day the team would communicate via the telephone or in person with the other departments, to ensure that all of the necessary supplies information and in some cases even the patient was available in readiness for the surgical team. This despite weekly meetings with these departments to schedule their activities in relation to the man-power resources, information needs, equipment requirements, etc, for the pre-planned theatre lists for the following week. The aim of these weekly meetings was to

anticipate the various needs and constraints, and devise strategies for overcoming any shortfalls.

From reference to Normal Accident Theory the system surrounding a typical hospital operating theatre just described is typical of a complex interactive, tightly coupled, high risk structure. By combining the concepts of Perrow and the findings of the organisations employing the probabilistic risk assessment system, there is great potential for even seemingly insignificant failures to combine, escalate through the organisation and disrupt the theatre in unexpected and catastrophic ways. From viewing the theatre this way the next question to ask is, "how can the maintenance department affect the smooth running of the theatre?"

As the theatre depends on the engineering services and equipment to be correctly maintained any sudden failure or deviation away from the recognised tolerance limits would seriously disrupt its services. However within such an environment of interdependence with other departments, any similar maintenance problem could/would create a knock-on effect and be transmitted through the organisation and eventually to the detriment of the theatre unit.

A typical example of such an incident that the researcher has had personal experience of (and which is not unusual) is for a surgical instrument decontamination plant to fail. These machines are expensive and therefore it is rare for there to be spare capacity (redundancy) within the Sterile Services Unit department to cope with such a vital piece of equipment to be out of action for more than a few hours. Again through cost it is rare for spare instrument sets to be available; therefore the pressure is on the maintenance department to bring the machine back into service as rapidly as possible.

The modern day decontamination units are a complex electro/mechanical/pneumatic device, which inject steam at high pressure into a chamber in a series of pulses, then clears and dries the chamber and its contents by drawing a vacuum. The sequence of operations is programmed by the automatic devices within each machine, and must control the temperatures, pressures and timeframe of operations to within strict tolerance limits laid down by rigorous national standards. Any deviation outside of these accepted limits fails the machine, and it must immediately be removed from service.

To repair the machine takes the work of a specifically trained maintenance technician who has the necessary nationally recognised academic decontamination qualifications, and competency assessed practical skill levels for that specific machine. As the machines are complex sometimes it takes many hours to trace and repair a fault, particularly if it is intermittent. Once repaired the equipment must be re-commissioned before being brought back into use. The cycle times for these units run into hours (not minutes), therefore the total down time from the fault originally occurring to the machine finally being brought back into use could be considerable, that is if all goes well. Delays in waiting for a suitably qualified technician to arrive, problems tracing the fault, or trouble obtaining a spare part all add to the total machine down time. The longer the SSU is without access to the machine the greater the probability that the theatre will be affected. Such a sequence of events was termed an 'operational breakdown' by Waring (2006), where one department fails or steps out of sequence with another, due to the inadequacies of another department. He divided these breakdowns into three categories, delays where items were delivered at the wrong time, miscommunications, or missing transactions where items failed to be exchanged. (Waring, McDonald & Harrisson 2006)

The example above describes the risk generated from a highly visible equipment failure. Other failures occur that are not so highly visible (latent conditions). The failure to undertake planned preventative maintenance tasks in accordance with the nationally agreed inspection and test frequencies and standards are an example of such failure. Such defects can (and do) lay dormant for long periods without being discovered and are managerial failures. These managerial failures combine with normal component failures to create an even greater potential for a catastrophic failure.

Nothing is perfect, in the areas of design, equipment, personnel, procedures, supplies or the environment. Much can be done to make these systems safer, however accidents can never be totally eradicated. To improve safety, managers and designers create safety buffers, alarms, redundancies in systems, etc. But small failures are a natural part of life and go on regardless. Occasionally these failures combine in unexpected ways, overcome the system defences and an adverse incident occurs; now thanks to Perrow we term this a 'normal accident'. If the system is tightly coupled the speed of the incident escalating through the organisation can be incomprehensible to the operator(s). The way

universal views are created from 'sense making', and are then accepted is a central feature of Perrow's normal accident theory; where sense making relates to the perception that an operative has of the situation. Different operatives and groups can have conflicting interpretations of events using diverse methods. It is the act of sense making that attempts to converge these views into one all embracing acceptable interpretation. Problems with such sense making create further errors, such that these lead to or exacerbate existing accidents.

2.6.3 People problems

One of the central themes in Perrow's work has been the relationship between power and risk. He claims that: 'the issue is not with risk but power; the power to impose risks on the many for the benefit of the few' (Perrow 1999 p.306). Risk assessments have always been purchased by the people in power, as important decisions are always accompanied with risk assessments, which assess the likely benefits against the risks incurred. But as power became more centralised so did the knock-on effects of any decision. This growth in centralisation in the private sector and the technological revolution meant that more and more of society were affected by the decisions of the few. This led to the need for government to step in. The necessity for this was because personal control was being eroded by a small number of people in powerful positions.

This was from the realisation that accidents are not only the result of system failure but also from organisational failure. It is wrong to continually blame 'operator error' as the problem, as capitalism and human greed play their part. But why would managers not put safety first? In answer to this question Perrow argues that it is because the 'harm and consequences are not evenly distributed' (ibid p.370).

As an example, the career span of a senior manager is relatively short compared to the gestation period of a normal accident, and few managers are punished for putting profits ahead of safety. The difficulty of proving the case against this culture is witnessed by the low number of corporate manslaughter charges that have been successful and the need to introduce a new act of Parliament. Also managers learn that it is unusual for a major incident to occur, therefore it is highly unlikely that their misdeeds would be discovered. Thus the personal risk to them is very small whilst the risk to society as a whole is large. This problem of 'learning the wrong thing' is not confined to the board room of an

organisation (the elite), but cascades down to those at the sharp end, the operators. They learn to cut corners for their own benefits, and more-often-than-not this works. It is not often that corner cutting 'bites back'.

People that follow an accident investigate what happened and what was the likely cause. If A followed B then they recommend that procedures are put in place to ensure in future that B follows A. What should happen is that they should investigate if this reversal could have occurred 'before' the incident by comparing the actual operating practices against the nationally and locally recommended and agreed operational procedures (Ramanujam & Goodman 2003).

2.6.4   Defence in depth

Sagan (2004) describes how Redundancy Theory can assist in designing an increase in overall system reliability by using components arranged in parallel, even if these components were themselves unreliable, i.e. a system that is more reliable than its parts! (Sagan 2004). Perrow 1999 states that the addition of parallel systems has been the main line of defence (p.73). The difficulty comes not in the design of the parallel systems but in ensuring the complete independence of the components. He goes on to explain three ways the employment of redundancy can 'backfire' to create a system of low reliability. Firstly, the system can become more complex and as such introduce hidden errors. Secondly, with the introduction of additional safety components human nature can dominate, and a lackadaisical attitude can develop that leads to attitudes (cultures) that assume, 'the other guy will spot if there is a problem', creating a diffusion of responsibility. Thirdly, the sense of inbuilt safety no-matter-what leads managers to push up production levels to produce unsafe conditions.

As an organisational theorist himself Perrow states that, 'time and again warnings are not heeded, risks are taken unnecessarily, inadequate workmanship allowed and dishonesty and outright lying are all not unusual' (ibid p.10). From his work he calls for more research into the redundancy concept in organisational theory, citing that engineers have shown the way by designing parallel systems to improve the independence of components in highly complex technical systems. The 'baton' thrown down by Perrow to the academic world has been picked up by many, one of whom is Sagan (2004).

2.6.5 The redundancy problem

In engineering terms it can be easily shown that designing safety systems in parallel greatly increases the reliability of the safety system, due to the employment of redundant components. This is also the case when parallel defensive systems are inbuilt into a managerial system. But when the 'people problem' is brought into these parallel managerial defensive systems it creates a 'redundancy problem' (Sagan 2004). Sagan split the problem into three parts, these are:

   a. Catastrophic common mode error
   b. Social shirking problem
   c. Overcompensation problem

a.　Catastrophic common-mode error

Sagan argued that adding redundancy increases complexity, and goes on developing this observation further by stating that the addition of additional components to create the desired redundancy can unknowingly create a catastrophic common-mode error, i.e. a fault that causes all components to fail. Just because the architect of the system designed the redundant components to be completely independent, does not necessarily mean that in reality they are. Sagan demonstrates this by reference to the aircraft and nuclear power industries. Aircraft designers successfully argued that putting two engines instead of three on a Boeing 777 would result in lower accidents, as the ability of a failed engine to interfere with the others would decrease. Then in a nuclear power plant, where a safety device added at the last minute broke away and caused a pipe blockage. Sagan goes on to argue that in complex systems it is difficult to predict probabilities of failure, and as such it is hard to know when to stop adding redundancies. The argument ends up in a discussion of, "who guards the guards? and "who audits the auditors?".

b.　Social shirking problem

This is where the reliability of the system is compromised by the diffusion of responsibility through what is termed, 'social shirking'. Again Sagan (2004) develops this observation further by looking at the 'translation problem' which he states is rarely examined within technical writings. The translation problem is when the pure mechanical theory of redundancy is attempted to be converted into the managerial systems of complex organisations. Redundant components within a mechanical system are

inanimate objects with no personal feelings or aspirations; they are thus totally unaware of the other components within the design. But in organisations the theory has to cope with people that consist of individuals, groups or departments that form the back-up system, and that are totally aware of each other.

Intuitively one would assume that increased redundancy within a managerial back-up system should lead to greater reliability, but in practice it has been shown in many cases to fail. Sagan (2004) presents examples where this has occurred. This failure is due to the individual or group being aware of the other redundant systems and shirking off unpleasant duties, or just becoming less observant assuming that someone else will spot any problems.

c.   Overcompensation problem

The addition of additional safety devices can cause managers to assume that they can push the system to ever increasing levels of output or to take chances that all will be well, when otherwise they would have halted the process/task. This 'sweating' of the organisations assets, pushing the system to its limits, actually reduces the reliability as the system is now operating within an entirely different scenario than was originally intended. Again Sagan presents a number of high profile examples where this has occurred with catastrophic results (Sagan 2004).

2.6.6   Lessons to learn about redundancy

The forgoing has shown three problems that can occur when designing redundancy into a complex system. The common 'knee-jerk' reaction to an incident, of increasing the safety systems, based on the proposition that this will increase the systems reliability has been shown to be false in some cases. This must be viewed in parallel with the realisation that low probability adverse events will occur from time to time, as this is natural and normal. Once it is accepted that there are limits to how reliable a complex tightly coupled system can be, it should be accepted that these systems are not perfect and that accidents will eventually happen, and that no matter how hard one tries, this fact will not alter. Knowledge of the ways that redundancy is effected by common-mode errors, social shirking, and overcompensation will not lead to a definitive set of equations to calculate how many and what type of redundancy systems to employ for a given

system and situation. But this awareness should help in their initial design and in-post incident investigations.

It has just been described how a healthcare environment is a complex tightly coupled interactive system, and previously how from its own reports it is wasting approximately £2 billion per year and causing 30,000 to 40,000 needles deaths per year. In recognition to this the National Patient Safety Agency and health care organisations are looking towards high reliability organisations to enhance patient safety (Fear 2006). The understanding of high reliability organisations as systems as opposed to the person centred approach is in line with the thinking of Reason (1990) and others.

2.6.7   High Reliability Organisations

When describing high reliability organisations it is common to cite nuclear powered submarines and aircraft carriers as typical examples. Fear (2006) describes and compares the key points defining high reliability organisations with a healthcare setting as:

❖ High reliability organisations work to high levels of safety despite operating in high risk environments.

❖ Healthcare systems have many features of an high reliability organisations they are complex; they require distinct disciplines to work together; they often have to make decisions quickly in the absence of adequate information; and errors can have significant human and financial costs.

❖ The main reason why healthcare organisations fall far short of being classified as high reliability organisations is that high reliability organisations do not fail.

❖ For an high reliability organisations to succeed, its processes (which include sensitivity to operations, preoccupation with failure processes and strong commitment to resilience) and its process elements (which include auditing, avoiding quality degradation, risk perception) must be consistently applied at all times.

❖ Another part of high reliability organisations is having the 'big picture' – in other words not everyone knows everything so communication between staff is vital. But more work is needed to understand how exactly high reliability organisations achieve reliability.

❖ Healthcare organisations are applying high reliability organisations processes to develop reliability checklists. This may be able to bring about improvements in both safety and performance.

High Reliability Organisations
and Safer Health Care
National Patient Safety Agency
Fear 2006

Even though Fear (2006) published in a peer reviewed document, he has made a fundamental mistake in his key point relating to the reliability of high reliability organisations, they 'do' fail. This is amply demonstrated by Roberts and Tadmor (2002). In their article they describe the catastrophic failure of the USS Greeneville nuclear powered submarine, which when quickly surfacing hit a local fishing vessel and killed nine people. To quote Perrow (1999 p. 330): "accidents cannot be entirely avoided". Fear should have stated that "high reliability organisations rarely fail". That stated, the lessons that can be learned from these high reliability organisations are being applied into other less glamorous industries, even where the effects of failure are not so catastrophic.

Communication is a main theme running throughout a high reliability organisation, as it is recognised that no one person can possibly know all of the actions that are taking place within such a complex and dynamic system. Therefore staff groups are not only task focused but appreciate the need to continue feeding the big picture. The system is fragile when resources are reduced, and as soon as key communications are lost, the high reliability organisation fails. Fear (2006) states that health care organisations are applying the lessons from high reliability organisations, but from the researcher's experience this approach is only in the early embryonic phase of development. He goes on to cite the work of Roberts and Tadmor (2002) and their work of applying lessons from the non-medical industries (Roberts and Tadmor 2002).

Perrow discusses the work of other researcher's that have contrasted Normal Accident Theory with High Reliability Theory. He argues that the supporters of high reliability theory believe that if we try harder we can overcome the proposed inevitability of accidents, and have virtually accident free systems. This is in total contrast to normal accident theory which states that there will always be accidents due to the complex interaction between tightly coupled systems. He also suggests that we need to examine organisations and systems that have not yet had accidents, in a similar manner that we would following an accident. He states this in his belief that one would conclude that the system was just waiting for an accident to happen.

2.6.8 Managerial System

From analysing a selection of the major incidents that have occurred throughout the world, one can determine that these illustrate the mistake of assuming compliance to 'good industry' and/or 'national standards'. In the majority of cases the disasters resulted not from quantitatively calculated risks occurring, as statistically they always can, but from inadequate standards of operational management, e.g. Bhopal, Three Mile Island, etc, as referenced by Reason (Reason 2000). They stand as evidence against the assumption that plant and equipment failures are totally independent from managerial failings, as the collapse of managerial systems can lead to hazardous conditions quite separately from the type and condition of the infrastructure. This has been recognised throughout the world in the Healthcare industry. There has been a rapid increase in National Patient Safety type organisations throughout the world in recent years. They emphasis, either in whole or in large part, the need to focus attention towards managerial system design, organisation and operations in relation to assessing the risk that can be generated from these systems (Emslie 2005).

As a result this thesis concentrates upon the managerial systems in place rather than the inherent reliability of the plant infrastructure and its engineering design and components. As such the research is operating from the hypothesis that the majority of the risks inherent within a Trust's maintenance regime lie within the management system, whose origins can be traced to decisions taken some time before an accident.

2.6.9 Managerial system failures

These can be classified as either active or latent. Active failures are actions committed by those operatives at the 'sharp end'. They are the immediate cause of the incident; but they inherit the accident sequence as opposed to being its instigators. Latent failures are the result of decisions taken by senior managers within the organisation. These latter failures lie dormant within the managerial system for long periods, seldom producing adverse consequences by themselves until triggered by an active failure (Ramanujam, Goodman 2003). At this point they combine to breach the system's many defences. Major managerial decisions are formed from a combination of economic, political and financial constraints, in the same way that engineering designs are always a compromise. Hence it can be taken that these types of decisions will include some negative safety consequences for some part of the management system. Thus we cannot prevent all latent conditions but we can expose their existence before they combine to breach managerial systems' local defences (Reason 2005).

The technological and design advances made since the birth of the NHS, particularly with regard to the engineered safety features have made the majority of installations proof against single failures, either human or technical. For a serious incident to occur it requires an unlikely combination of a number of contributing factors. Therefore instead of concentrating on the last incident and trying to develop local fixes the focus of risk managers is now being directed towards identifying and eliminating the worst latent problems. To avoid inconsistency when reviewing a managerial regime, the 'worst credible' case must be considered. By adopting this approach the aim of assuring that patients, staff, contractors and others would be safe in the event of the worst credible incident would be achieved by the implementation of a suitable assurance framework (Tweeddale 2003).

2.6.10 The design of a Management system

A managerial system must be designed to cope with both the steady state normal day-to-day activity and to adequately respond to an emergency situation (disaster planning). This thesis only researches the estates' managerial systems for the normal day-to-day activity.

The joint Australian/New Zealand risk management standard shows a common graph that depicts the rise in risk as the available budget decreases for a steady state situation (Standards New Zealand 2005). This approach is complemented by that of Tweeddale as he argues that if there is a change in resource availability and costs are cut with subsequent drastic reductions in maintenance, supervision and training, the risks are likely to rise in a manner that cannot be totally predicted (Tweeddale 2003 p 205). He goes on to argue that any changes to the design of a managerial system must assess the probable increase/decrease in relative risk, and the assumptions made from such a study must be verified before, during and after any such action.

Risk within an NHS Trust estates' engineering department arises from two sources, the as fitted engineering infrastructure and equipment, and human error. The latter topic of human error itself can be split into two sub-topics; these are the person approach and the system approach. The person approach focuses on the actions of people 'at the sharp end' and the respective errors of individuals, targeting their forgetfulness, inattention or moral weakness. The systems approach centres upon the managerial systems surrounding their activities and attempts to build defences to avert errors or reduce their effects (Reason 2000).

This thesis excludes both the inherent risk that is being generated from the engineering plant and equipment, and that created by the person approach, but it does investigate the risk created from the managerial system governing the estates' actions in relation to some engineering topics.

Within process plants some equipment is fitted with very reliable shutdown systems for safety reasons. These are designed to activate in the event of a given condition(s). Within the specification these systems are required to be periodically tested to ensure continued safe operation of the installation, as an untested system cannot be relied upon. This thesis recommends that this philosophy must be duplicated when designing and operating a managerial system. It is well known that people are not machines; they have lapses of attention and do not follow rules precisely, they do not make important decisions well when under stress, they are emotional, etc. Therefore the managerial systems must be designed to account for such human behaviour. Kletz makes remarks to the effect that to blame human error for an accident is about as helpful as 'blaming

gravity for falls' (Kletz 1997). The principle that should be adopted where any single human error could lead to a catastrophic incident is that sooner or later it will happen, as the people involved will be nothing more than human (Tweeddale & May 2001).

Differing tools are available for the assessment of risk in relation to a management system. Therefore managers must be aware of these tools, when to use them when assessing risk, and what the limitations are of each methodology. Aven and Kørte point out that 'engineers in general are not familiar with the details of alternative risk theories' (Aven & Kørte 2003) They also go on to argue that it is not sufficient to find fault with a particular technique and reject it, but that an alternative must be found that is less defective and can be seen as useful by the decision makers. This statement is also supported by Backlund and Hannu (2002) who go on to observe that a study must develop a detailed specification with the need to analyse and interpret the data produced before any management action is taken. But all agree that risk assessments should only be undertaken by those with experience based on the technology and managerial systems in use, and the environment that the systems will be operated within.

2.6.11 Risk assessment of a managerial system

Qualitative risk assessments are not precise, but the actual process of collectively questioning and analysing the system, develops a detailed insight into the separate components that create the risk. This in turn allows the team to concentrate on those areas of concern, and to develop risk reduction measures where they are most needed. However equal attention must be given to both the high-risk and low-risk components. This statement is supported by the work of the National Aeronautics and Space Administration. employs risk assessment techniques for their engineering and managerial systems on a daily basis. For many years they relied on the Failure Modes and Effects Analysis methodology for the identification and assessment of risks, until the Challenger disaster on 28th January 1986 (NASA 2002). From this period onwards they began to use a system known as probabilistic risk assessment.

The failure modes and effects analysis approach begins with a process to be investigated, whereas the probabilistic risk assessment starts with an adverse outcome and employs a 'top down approach' that first models the outcome, and then investigates all combinations of process failures that could create this event. The probabilistic risk

assessment system was developed within the nuclear industries and is now used in other fields such as petrochemical, offshore platforms and defence. The system is now a mature and credible method, and has repeatedly proven its ability to uncover design and operational weaknesses not found by other means. The process has showed the importance of not only examining low-probability and high-consequence events, but also high-consequence scenarios that emerge as a result of a combination of high-probability and almost trivial events. The National Aeronautics and Space Administration states that 'contrary to common perception, the latter is sometimes more detrimental than the former'.

Marx and Slonim (2003) reported that the Joint Commission on Accreditation of Healthcare Organisations has required (since July 2001) that each accredited hospital develop at least one proactive risk analysis at least annually. The Joint Commission on Accreditation of Healthcare Organisations recommended that Failure Mode Effects Analysis was one tool that could be applied to the task (Van der Schaaf, T. 2002). The National Patient Safety Agency has recommended a range of tools to assess risk within a managerial system; amongst these are Failure Mode Effects Analysis and probabilistic risk assessment (National Patient Safety Agency 2005).

2.6.12 The disruptive role of non-novel events

Marx and Slonim (2003) have incorporated both engineering equipment and human components into probabilistic risk assessment within the healthcare environment, thus transforming the approach from an equipment probabilistic risk assessment to a sociotechnical probabilistic risk assessment. Thus it has been recognised that probabilistic risk assessment has the potential to improve patient safety in healthcare, and can build on the results of Failure Mode Effects Analysis and other systems, but further research must be undertaken particularly in the area of human probability data (Wreathall and Nemeth 2004). However both the National Aeronautics and Space Administration and Marx *at al* state that whilst probabilistic risk assessment is an advantage, it is a very complex system to attempt to employ.

Major disasters such as Chenoble and Three Mile Island have long interested organisational theorists as they provide the opportunity to study managerial processes that are not in equilibrium. Insight that has emerged from the literature over recent years

is that major disasters do not have proportionally large causes. It is now recognised that small events can link together in unexpected ways to create disproportionate and disastrous events. It is suggested that as production technologies become more sophisticated, interconnected and interdependent then the likelihood of chain reactions, where one event can reverberate throughout a system and cause numerous malfunctions, greatly increases the chance that a minor everyday event will lead to a major disaster.

The analysis of disasters often focuses on the novel as opposed to the non-novel events that are sensed and resolved. Rudolph and Repenning state that in many cases both the novelty and quantity of interruptions to existing and established routines play a significant role in precipitating disaster, yet the role of quantity has received little attention (Rudolph & Repenning 2002). To investigate the interconnections of the quantity of small non-novel events and organisational crises, they constructed a dynamic mathematical model to simulate an ongoing stream of interruptions. They go on to cite the 1977 air disaster at Tenerife and the downing of a passenger jet by USS Vincennes, as in both cases these incidents were considered to be caused by such a stream of non-novel interruptions.

2.7   Rudolph and Repenning model theory

The model was developed by building on in-depth analysis of a number of existing case studies and relevant psychological theory. But unlike formal models they developed the system from grounded theory, whose creation in the 1960's has been credited to Glaser and Strauss (1968). In their words, "The model synthesised the causal processes found in existing analyses of disaster and experimental studies of human performance under stress, to create a laboratory for studying and theorising about the connections between quantity and disaster."

The simulation and analysis provided a new characterisation of how organisational systems respond to an ongoing stream of non-novel, survival threatening interruptions. The purpose of their research was to develop a general explanation of how small events can create a crisis. Central to understanding the role of quantity in organisational crises was the use of the Yerkes-Dodson Law that posits the relationship between stress and performance as an inverted 'U' shape. The result of their research is a theory explicitly

linking stress and performance concerning the role of interruptions in organisational collapse.

The model was based on a number of key assumptions. One of these assumed that the organisation was faced with a continual (and potentially varying stream) of non-novel (as opposed to novel) interruptions. This is very similar to an estates department's reactive and planned preventative maintenance workload, as this consists of small risk tasks arriving in a random order. Their focus on non-novel did not imply that all such events were created equally or that they could be resolved without significant cognitive effort. Therefore they varied the number of mental steps that must be undertaken for various interruptions. This is again similar to the reactive maintenance problems that are generated within healthcare premises. Although made up of small individual and unrelated tasks they all require varying degrees of management time and engineering skill to complete them.

They also assumed that resolving the pending interruptions was required for the organisation's survival. This is also equivalent to an estate department's need to close all reactive and planned preventative maintenance tasks, not always immediately but within a reasonable time span. Thus it is natural for there to be a constant source of tasks pending and waiting for action.

To define the net interruption rate, a distinction was drawn between interruptions and errors. Interruptions were treated as exogenous inputs as they were created outside the system, and errors were classified as endogenous i.e. created inside the system. As errors arose then more effort was required to resolving the interruptions. As the number of interruptions increased then the probability of errors arising also increased. This is very similar to the activity to be found within an estates department, where because of the varying nature and quantity of the day-to-day maintenance work that is generated, errors are created by the operatives under stress.

This approach created a simple 'stock and flow' structure for the Rudolph and Repenning model that simulated the ebb-and-flow of the interruptions that were being experienced. This developed a feedback mechanism from which past performance could be linked into the Yerkes-Dodson Law, which suggests an inverted U-shaped

relationship between stress and performance. Rudolph and Repenning recognise the controversial history that this law has attracted, but accept that the curve is an appropriate depiction of the relationship between stress and performance within the scope of their model. The historical vicissitudes of the Yerkes-Dodson law have been traced by Teigen (1994) and an introduction into stress/performance research is presented by Mendl (1999), who describes the law as 'a general description of the relationship between stress and cognitive function' (Mendl 1999; Teigen 1994).

When their model is ran forward under varying conditions it shows the point where the performance of an organisation decreases and the stress levels increase, as the non-novel interruptions increase. This research will show that their model very closely mirrors the activities experienced by an estates department when subjected to varying levels and rates of interruptions. Rudolph and Repenning conclude their findings by recommending that:-

> Future research could be profitably focused on helping organisations assess their capacity to handle interruptions and develop systems to signal when the level of unresolved interruptions reaches a critical level.
>
> (Rudolph & Repenning 2002)

This thesis will apply the findings of the above theoretical model to an estates department's actual data and determine this critical level.

From the forgoing, an integral part of any managerial system must be the topics of monitoring and audit to give assurance that the agreed maintenance strategy was being adhered to and that the critical level was not close to being breached. That in turn would form the overarching assurance framework of the total management system.

2.8   Assurance framework

Monitoring is part of the day-to-day activity for a manager, for without monitoring issues are not being controlled. Therefore the monitoring role (internal governance control) must be subjected to the same performance criteria as any physical engineering safety device and must follow the same reporting structure as financial or production performance. Thus internal control is explicitly linked to risk management (Spira & Page,

2003). The national guidelines for NHS Trust Boards recognise this in the Integrated Governance Handbook 2006. An audit should be the second check on the monitoring activity and not the only backstop (Tweeddale 2003 p 295).

The variables to be monitored within a managerial system must be department specific, and be based on specific hazards and hazardous scenarios of the working environment, as well as the more generic requirements of national good practice guidelines. These variables will by default include both qualitative and quantitative assessments of performance. However not everything needs to be monitored (Dobson, Lock & Martin, 2006).

It is a fallacy to assume that if an organisation has a low rate of accidents and incidents that these figures can be used as an indicator that the managerial system is sound. Tweeddale (2001) claims that there is sufficient evidence to support this observation, and goes on to cite Flixborough and its low level of lost time through injury figures. He then states that this is not an isolated example and that numerous other cases could be quoted. Also it must be recognised that the limitation of historically determined risks is that the computations may not be relevant for future predictions. The management system requires a range of decision support tools and indicators to continuously assess risk against agreed levels, and give assurance as to the stability and robustness of the managerial system, i.e. audits must be undertaken (Tweeddale 2001).

2.9   Audit

The need for audits arises when accountability cannot be continued by informal relationships of trust alone, but must be placed on a formalised footing, made visible, transparent and be subjected to independent validation. Audit is generally only visible following a failure, and has become the central plank to ways of discussing the administrative control of an organisation. This in turn usually leads to calls for more policing (audit), but rarely for a detailed analysis of why the system had failed. The traditional style of such audits concentrated on basic tasks and employed engineering-based quality control techniques with statistically grounded methods of analysis. But these systems based audits could become rituals governed by a mentality that created a mind set concerned only with compliance against a pre-agreed process.

In contrast the modern audit looks at the systems of control in relation to corporate governance issues. As such they indirectly act on the systems of internal control, as opposed to the first order activities. Thus audit has become a 'control of control', where what is being assessed is the quality of the internal control systems, rather than the tasks themselves. Therefore this style of audit assists in demonstrating the excellence of control and not the quality of the output. As inbuilt into the audit would be questions that related to the relevance and efficiency of the system, "o/k you are doing it, but why are you doing it?", "is it the correct thing to do?", and finally, "could it be done better?"

The transparency of the audit process, how the audit was conducted, by whom and into which areas is an integral part of the quality of the audit process itself. Just as importantly is the need for transparency of the audit findings within a government body such as the NHS. This has now been created through the Freedom of Information Act. Whatever the merits of the act, it has exposed the NHS managerial systems to the gaze of the public.

One of the objectives of a modern audit relates to continual organisational improvement. Russell (2004) states that: very few top quality managers have received an adequate level of quality management training. He goes on to argue that even now (the start of the 21st Century) quality management principles are not normally taught to university business students, and that the same could be said for other students going into other professions e.g. safety, health and environment (Russell, 2004).

There is a distinct difference between the tasks of monitoring and audit. The former is associated with what should be normal managerial day-to-day activities of checking that various tasks have been completed to a pre-agreed timeframe and output standard. The latter relates to a periodic (normally yearly) in-depth examination conducted at two levels, strategic and operational (Donaldson & Armstrong 2000). Strategic level audits investigate whether the key organisational requirements are being met, such as financial and budgetary constraints, service levels to be achieved and key performance indicators. Operational audits confirm (or otherwise) compliance with topics such as the agreed planned maintenance regime, procurement methods, staff competencies, etc. In order to audit these topics clear statements of procedures and the methodology for assessing outputs need to be developed. These assessments can be qualitative and

quantitative. From the audits conducted by BSRIA (BSRIA, 2000) into a sample of organisations other than NHS estates department's common failures were found across a range of headings such as:

- ❖ Maintenance organisation and resourcing
- ❖ Planning and management of operation and maintenance
- ❖ Health and safety issues
- ❖ Financial management control
- ❖ Records

They also investigated different types of audit regimes to establish if business critical systems were being addressed. Within their research they investigated different types of business, and reported that each required a unique focus and criteria for determining business critical aspects. Therefore one of the purposes of an audit is to give assurance that the relevant standards are being achieved. As a managerial system may appear satisfactory, the only means of really knowing how well it is understood and operated is to test it. There are numerous standards and guidelines issued that organisations should be complying with. Much information has been published discussing the integration of these standards, but there is comparatively little documentation relating to how to actually build up such an audit system (Beckmerhagen *et al.* 2003).

Within the NHS there are two types of audit; the first is an internal audit that is undertaken by a specialist department, the second is an external audit where this activity is contracted to an independent company. The latter is normally commissioned to undertake work where a high degree of technical speciality is required.

Karapetrovic and Willborn (2000) have recognised that specialist standards and corresponding management systems have many critical characteristics in common. Thus a generic integrated management system should be able to be designed that could offer guidance to all managers of an organisation to present them with a clear view of the complex managerial tasks which they should be performing to identify and manage risk. This is also applicable to auditors, where from the design of a systematic, well designed and executed integrated audit, all participants would be aware of the overall goal of the audit. A short history of the problems that organisations are facing due to the

lack of such an integrated system has been written by Karapetrovic and Willborn (1998). Beckmerhagen *et al.* (2003) cites a number of examples where organisations have benefited from such an integrated audit approach. They go on to state that to undertake such integrated audits, auditors must learn new skills and adopt a new kind of audit, to one that "assists management to integrate function-specific management systems". The design of the new audit must be developed from a multi-professional team. Thus a truly integrated audit would be one where the organisation is assessed via a group, as opposed to an overall audit based on a number of separate audits undertaken by individual specialists. (Karapetrovic & Willborn, 2000; Karapetrovic & Willborn, 1998)

In response to the need for an integrated audit system ISO 19011:2002 was produced which replaced six previous ISO standards. It gives guidelines for quality and/or environmental management systems auditing (British Standards Institution, 2002). The standard is presented as a generic process as opposed to a procedure, which can be applied to other types of audits such as other management system audits.

2.9.1 NHS Trust audit committees

To promote the modern role of audit committees the NHS has published an updated handbook that makes explicit reference to the assurance framework that has been specifically designed to shape the local audit committee's responsibilities in relation to risk assessment and management activities (NHS Audit Committee, 2005). Reference within the document has also been made to the strands of governance that bind together the financial and clinical risks, thus emphasising that today's audit committees have a responsibility for analysing the risks and controls that affect all of a Trust's aspects of business. They must focus on the risks, controls and related assurances that deliver the organisation's objectives (the assurance framework), in particular the Statement of Internal Control that is included within the Annual Report and Accounts. The handbook recommends that the committee concentrate their efforts where the inherent risk is high and their control mechanisms are critical, and also where the residual risk is high and the situation needs monitoring.

Part of the audit process is to ensure that when senior managers make decisions that affect the available resource levels for maintenance, they do so with the demonstrable understanding of the effects their decisions will have on the risk to the organisation, be

these positive or negative in nature. However as the Healthcare Commission is charged with auditing the national standards, any maintenance strategy adopted must be 'evidence based'. In this way assurance can be demonstrated, and arguments presented as to why certain actions (or non-actions) were taken when assessed against the relevant standards.

In order to practice good governance it is necessary for managers to be constantly appraised of the performance of their organisation when assessed against measurable output targets. This information must show the history of performance over the years, and the progress trends against projected aims. This necessitates the design implementation and maintenance of a day-to-day monitoring system, or more correctly termed: an assurance framework. The creation of such a framework in relation to this thesis must by necessity conform to the five criteria headings shown below:

a. International standards
b. National standards
c. Existing systems
d. Multi-professional
e. Revenue budget

a. International standards

The inherent design of such a system must match the recognised international standards through to locally set values and principles. The international standards would be in the form of ISO 9000 series which is a family of principles for quality management systems, and ISO 19011 which relates to the integrated audit of systems. As the structure is dealing with processes within the NHS then reference must also be made to the Australian/NZ Risk Management standard, because the NHS risk management process is based on these guiding principles. All of which must be underpinned by reference to mainstream academic theory. Without compliance to these criteria then whatever system was developed could not be regarded as 'fit for purpose'.

b. National standards

These begin with the national codes of practice issued by the NHS. For the eight engineering systems analysed within this research, and take the form of Health

Technical Memorandum issued by the Department of Health. These in turn are closely linked to the national standards issued by the Healthcare Commission and those published by the NHS Litigation Authority.

c.   Existing systems

Use must be made of the existing staffing structures and job descriptions, as there would in all probability not be the resources to employ additional staff for monitoring purposes. In addition any computerisation of the monitoring system should use existing hardware and software.

d   Multi-professional

Although this research is dealing with engineering systems the electro/mechanical/pneumatic processes form part of a complex interactive tightly coupled managerial system that affects either directly or indirectly the delivery of patient care. As such the need for monitoring spans much further that that of the engineering management team. Therefore the monitoring system must be applicable to a range of professions within the service.

e   Revenue budget

The researcher has set the criterion that the final design must achieve a zero revenue impact on the Trust's budget. If the monitoring system failed to attain this then the researcher considers that there would be great difficulty in persuading Trust's to adopt such a system.

2.10   Discussion

This research programme began with the publication by the Department of Health that outlined the severe managerial failings within the National Health Service, that were creating unacceptable conditions of patient safety (Department of Health, 2000), this was followed by an informal meeting chaired by the researcher, which consisted of ten experienced NHS estates maintenance managers, as discussed in section 1.5. The first output from this meeting was the following research question:

'Do the managerial systems of internal control within the National Health Service (NHS) assure that the day-to-day building service engineering maintenance activities conform to their respective national guidelines?'

The literature search showed that no study had been published that could adequately answer this question, Therefore the aim of this research is to determine if unknown latent conditions exist within the NHS maintenance department's managerial systems that could cause intolerable/substantial harm to patients; and if the managerial systems do not give such adequate assurance, determine how this could be achieved. From this aim the following four research objectives were formed:

First research objective

Develop and implement a robust, transparent and repeatable methodology to assess if the managerial systems of internal control within the National Health Service (NHS) give adequate assurance that the day-to-day building service engineering maintenance activities conform to their respective national guidelines.

If the response to this first research objective was positive, and that the managerial systems **do give** adequate assurance then the research was to cease. However if the response was negative and that it was shown that the managerial systems were inadequate and **do not** give suitable assurance then the second, third and fourth research objectives were to be undertaken:

Second research objective

Develop and implement a robust, transparent and repeatable methodology to assess the resource needs to correct the existing managerial system deficiencies to that required by their relevant National Health Service standard.

Third research objective

Develop and implement a robust, transparent and repeatable methodology to assess the directly employed artisan revenue resources needed to ensure that the maintenance activities conform to the National Health Service standards.

Fourth research objective

Develop a robust, transparent and repeatable methodology to monitor the day-to-day activities of the maintenance department in line with the requirements of the National Health Service standards.

2.11 Chapter Two Summary

This chapter began with a short description of reliability and risk, and the balance that must be achieved between these two topics when designing a maintenance strategy. The Department of Health's national guidance relating to the formation of a Trust's maintenance strategy issued in the year 2002 was discussed, and compared to that published in 1973. From this assessment it was seen that there has been no change in policy over the 30 intervening years, and that a Trust is free to create its own maintenance strategy to reflect local conditions and budgets. The term maintenance was examined and it was shown that this included the managerial activities as well as the physical artisan tasks, thus forming a total system. The artisan tasks were then broken down into the two categories of reactive and planned preventative maintenance, and examples were given for each type.

The use of computers within maintenance departments was shown to be widespread throughout the NHS, and that the NHS had abandoned the development of a centralised form of software. This has forced Trusts to purchase commercial packages to assist in the management of their data needs, and as a consequence there is no national data bank containing quantitative records relating to maintenance activity or reliability statistics. The use of simulation models of the past was shown to be of little practical use to the NHS maintenance manager, even though there have been various attempts to develop such a system over the years. The type of maintenance strategy to be adopted was considered, and an argument developed that showed that actuarial analysis relating to the relationship between operating age and failure is of little use to the NHS day-to-day maintenance manager. This led to the statement that the maintenance strategy should be targeted towards the system known as Reliability Centred Maintenance.

The assumption that plant and equipment failures are totally independent of managerial failures was discussed, as was the research that showed that the vast majority of incidents are caused through managerial decisions, actions or inactions. This focused

the chapter into the risks associated with the day-to-day activities of managerial systems, as opposed to those being generated from the physical plant and equipment.

Normal accident theory was discussed and concluded with the observation from its founder Perrow (1999) that the supporters of high reliability theory believe that if we try harder we could overcome the proposed inevitability of accidents, and have a virtually accident free systems. This in total contrast to normal accident theory which states that: there will always be accidents due to the complex interaction between tightly coupled systems. He also suggests that we need to examine organisations and systems that have not yet had accidents, in a similar manner that we would following an accident. He does this in his belief that one would conclude that the system was just waiting for an accident to happen.

The creation of a suitable managerial system was then debated along with the two main techniques used to assess the risks associated with such a design. From this debate two themes emerged relating to normal day-to-day activities (i.e. non-emergency). The first was the identification of the tipping point where an organisation eventually fails to cope with an ever increasing workload, and the second was the need for monitoring/audit programmes to be designed and implemented.

A research model relating to an inherent failure of a management system due to the quantity of non-novel interruptions during normal day-to-day activities was examined, and a parallel was made between the model's outputs and that of a typical NHS estates department. It was then stated that this thesis would test the concept of this theoretical model against actual data from an NHS estates organisation.

Finally the need for and development of an assurance framework was discussed along with the call for an integrated monitoring and audit programme to be implemented.

Chapter Three   Methodology

3.1 The research methodology

The literature search chapter showed that there were no comparable studies undertaken or relative data available to address the research question:

> 'Do the managerial systems of internal control within the National Health Service (NHS) assure that the day-to-day building service engineering maintenance activities conform to their respective national guidelines?'

As discussed within section 2.10, in order to address this research question four research objectives were formed, for each a separate research instrument methodology and process was designed.

Central to the methodology of each research instrument was the use of 'Expert Elicitation' techniques and 'Expert Judgement'. In cases where there is a lack of suitable information or data, (as in this case), and/or the creation and collection of the data needed is prohibitively expensive and/or time consuming to collect, (as in this case) then, 'Expert Judgement' is employed as it characterises the 'state of knowledge' (Meyer & Booker, 2001). This term is also referenced as expert opinion, subjective judgement, expert forecast, best estimate, educated guess or expert knowledge; where an expert is regarded as a person who has a specialist understanding of a particular subject by his/her peers (Meyer & Booker, 2001). The extraction of the expert's knowledge is via the use of expert elicitation techniques, this has been defined as:

> 'Expert elicitation refers to a systematic approach to synthesize subjective judgments of experts on a subject where there is uncertainty due to insufficient data, when such data is unattainable because of physical constraints or lack of resources. It seeks to make explicit and utilizable the unpublished knowledge and wisdom in the heads of experts, based on their accumulated experience and expertise, including their insight in the limitations, strengths and weaknesses of the published knowledge and available data.'
>
> *Expert Elicitation*
> *RIVM letter report 630004001/2008*

Although the origins of expert elicitation can be traced back to the 1950's, the research work undertaken by Meyer and Booker discuss at length the difficulty of choosing the correct elicitation technique and components for a particular project. They state that, 'unfortunately, there is little comparative literature on the methods, and the few sources that exist may provide conflicting advice'. They suggest that the standard methods and components should be listed, compared and reasons for their final choice be recorded and agreed with the project sponsors and participants before undertaking the study, particularly if these are to be tailored to suit a particular situation (Meyer & Booker, 2001). Following their advice, an overview of the elicitation process is presented below with particular reference to this research project.

3.1.1 Overview of the elicitation process followed

The methodologies employed for the elicitation of expert judgements are either formal or informal. With the informal approach experts are asked to provide opinions off the 'top of their heads'. This is subject to bias and skews the judgment from a given reference point. Formal methods for gathering the responses from experts involve the selection of the experts, designing the elicitation methods and determining the mode in which expert is to respond. This latter approach has two major advantages. Firstly more time can be spent in gathering the information in a controlled environment. Secondly documentation and records are developed showing how the judgements were created for future auditing and analysis.

There are a range of elicitation processes that differ in areas such as: interaction between experts, number of meetings and their frequency, time spent preparing for the meetings and the time each meeting takes and the response modes used. Meyer and Booker state that even with the variety of the elicitation processes there are only three basic situations, these are:

> **Individual interviews**
> In this case each expert is interviewed alone in a face-to-face situation, this can extend to a series of meetings over a period of months where the interviewer travels to each expert. The advantages of this methodology are that it is regarded as the best method for gathering detailed data, and avoids the problems with

group dynamics. The disadvantages relate to the extended length of time it takes to gather the data, labour intensive, expensive and it suffers from the lack of interactivity with other experts.

Interactive groups

In these situations the experts meet in a group with face-to-face interaction. The meeting can follow the traditional unstructured approach or can be structured to prevent spontaneous interaction. The advantages to this type of approach are that it gives a more accurate data set, and generates a greater quantity of ideas. The disadvantages are that it creates the potential for group think bias and the logistical problem of scheduling the meetings between the experts.

Delphi situations are where the experts operate in isolation from one another and submit their responses to a moderator, who then redistributes them to the other participants but keeping the authors identity secret. The experts are then allowed to revise their own judgments if they wish in the light of this new information until (if possible) a consensus is reached. The advantage of this system is that it avoids the potential of group dynamics influencing the outputs. However the main disadvantages are that there is less synergism created between the participants and is normally the most time consuming of the three systems.

3.1.2 Modes of communication

In combination with each of the three elicitation processes above communication can be face-to-face, telephone and mail (either post or e-mail). However it would be common for the face-to-face mode to be used for the individual or group interactions and telephone and/or mail for the Delphi system.

The face-to-face mode being time consuming, labour intensive and expensive but with the advantages of eliciting detailed information. This, whilst the telephone method can be used to discuss limited amounts of information it, is normally less expensive than face-to-face interaction and quicker than the mail mode. Alternatively the mail mode is frequently used within the Delphi method to extract simple data from a large sample

population and is cheaper than the face-to-face mode but has the lowest response rate of all of the modes.

3.1.3 Choice of Elicitation process

The data to be collected consists of a substantial amount of detailed technical responses across a range of eight specialist topics. Also the subject matter directly influenced patient safety therefore the timescale for the process had to be kept short, and costs had to be minimised. Therefore an interactive group (focus group) via a series of structured meetings with the researcher acting in a process consultation role (Schein 1997) was chosen as the vehicle for the data gathering exercise as opposed to other expert elicitation methods (Meyer & Booker, 2001; Ayyub, 2001) as speed of response and face-to-face interaction amongst the group were high priorities for the researcher.

3.1.4 Selecting from the response modes

There are various response modes relating to each of the three elicitation methods of individual interviews, interactive groups and Delphi. A brief outline of these is given in table 2 which highlights the advantages and disadvantages of each.

Table 2   ***Differing response modes and examples showing their advantages and disadvantages.***

Response mode	Example	Advantages for experts	Disadvantages for experts
Physical quantity	Temperature, Pressure	Little difficulty in understanding	Decomposition of the question is needed
Probability estimate	15% of an event occurring	Commonly used in conjunction with established analysis techniques	Very time consuming and most experts are not good at estimating probabilities
Odds ratio	1 chance in 10,000 of an event occurring	Convenient form of estimating rare events	Small sample of events effect the estimates
Probability distribution	Min 15, mean 20, max 45, from a triangular distribution	Same as probability estimate	Same as probability estimate
Continuous scales	End points represent	Most people appear reliable when using	Requires time to develop the scales to

Response mode	Example	Advantages for experts	Disadvantages for experts
	extreme values separated by linear divisions	these scales	fit a particular application
Pairwise comparisons	Comparing A to B is reciprocal of B to A	Most people reliable as only two quantities are discussed	Time consuming to discuss all comparisons and comparing more than seven things is difficult
Ranks or Ratings	Either or both numerical or descriptors	Experts find ranking easy to use	As above
Bayesian updating	Combination of different sources of information	Can combine various sources of data	Requires assumptions about distributions

3.1.5 Designing and Tailoring the Elicitation

Tversky and Kahneman (1974) have extensively researched the inability of people to accurately predict the percentage probability of an outcome. The effect that this problem has had on various projects and situations has been the subject of much criticism and research (Conrow, 2003). Therefore the response modes relating to the assessment of percentage probabilities were rejected. As there was a lack of information relating to these topics then the Bayesian methodology was also rejected. The timescales needed to undertake a comparison or rank and rating study eliminated this method. These arguments reduced the response modes to: physical quantity and/or continuous scales.

3.1.6 Selecting the type of Aggregation

In order to form a single answer to a question from the responses of various experts then an aggregation process must be employed. The types of process fall into one of two methods; behavioural or mathematical. The former relies on the experts reaching a consensus during the elicitation session. The latter employs mathematical processes to form the single consensus output. As the choice of elicitation process was an interactive group, the aggregation system employed was the behavioural process. In this case the averaging of the expert's responses was taken to obtain a single output. Averaging in this way gives equal weight to each participant's expertise (Ortiz, Hora, Meyer & Keeney, 1991).

3.1.7 Designing the individual questions

The proposed accomplishments of the project were stated at the end of chapter two by defining the research purpose (goals). This in turn dictated the topics to be analysed and from the literature search it was shown that the issues could have an intolerable/ significant impact on the patient; and that related sources of information from experimental, observational or validated computer models were not reliable and/or not available (Ortiz, Hora, Meyer & Keeney, 1991).

The questions under each topic have to be designed to reflect the overall purpose (goal) of the research project. They have to be sufficiently specific to allow the expert to answer without ambiguity if possible, and the magnitude and level of detail of the data required to answer the question must be assessed. In addition the ordering of the question sets must follow a logical pattern and they must be phrased in a non-leading way (Ortiz, Hora, Meyer & Keeney, 1991; Meyer & Booker, 2001; Ayyub, 2001; Øien, 1998).

3.1.8 Selecting the experts

If the study is to focus on the experts answers then a different type (and usually number) of experts are needed compared to the type of study that focuses on the experts problem-solving techniques. For applications studying risk/reliability and decision analysis the former type of expert is needed. This type of expert needs two types of skill. The first being substantive expertise that comes from knowledge and experience in the question field. The second relates to the type of response required from the expert to answer the question known as normative expertise, e.g. percentage compliance to a national standard. Also to ensure that the question (problem) is viewed from a number of angles it is advisable to employ experts from a diversity of backgrounds. As the selection schemes used to chose the participants could skew the final result, careful consideration must be given to the composition of the expert group, their motivation for participating and their accessibility (Meyer & Booker, 2001) (Ayyub, 2001).

3.1.9 Designing and Tailoring the Elicitation

The forgoing paragraphs have summarised the main parts of the expert elicitation process and reduced them down to the design suitable for this research objective, this being both the adoption of a structured multi-professional interactive group (focus group) and individual interviews, both generating single answers of physical quantity and/or

continuous scales to a series of detailed questions via a consensus agreement. Where the topics under investigation are divided into a number of discrete questions, each is specifically designed to be easily answered, by experts with substantive and normative knowledge.

3.2 Development of the first research instrument

The first research objective is a similar problem to that faced by Ramanujam and Goodman (2003). In their research they collected data from 80 organisations, which had undergone numerous mergers and reorganisations, but where their core operating policies and procedures had remained the same. This is very similar to the situation of NHS Trust estates departments. Their research assessed the quality of internal control as evidenced by the organisation's compliance with rules and regulations. Following their lead, critical elements (questions) were developed based on the respective NHS codes and the eight engineering topics identified by the initial focus group (see section 1.5). Each critical element represents a major discrete component within the guidance in the form of a question. Three examples of the critical elements developed are shown below:

a. **Water temperature control** It is essential to check the temperature settings and operation of all water mixing devices routinely, at least half yearly.

b. **Ventilation fire containment** Regular tests, at intervals agreed with the local fire prevention officer, will need to be carried out in order to demonstrate the continuing efficiency of the fire detection and containment systems. Records of these tests should be kept.

c. **Lightning protection** Where lightning protection systems are provided they shall be subject to a yearly planned preventive maintenance inspection and report.

The critical elements were formed from the 'exact wording' (i.e. copied) from the relative standards where possible, in order to avoid misinterpretation of the actual meaning of the standards.

Following the formation of a series of critical elements, it was necessary to determine a Trust's actual managerial activity against these critical elements, in this way any gaps in

managerial performance would be identified. In the words of Ramanujam and Goodman (2003) these gaps are the: 'uncorrected deviations from procedures and policies that have no direct adverse consequences', i.e. they are latent conditions. In order to capture the expert's response mode to the question of conformance to national standards, a series of utility functions were developed.

3.2.1 The need for and design of Utility Functions

The normal method within the NHS of assessing risk from reference to the standard NHS 5x5 criticality grid shown below in table 3, where descriptions of Impact and Likelihood are matched against the Trust's understanding of these terms (Department of Health 2004).

This thesis argues against the use of a 5x5 criticality grid to assess the risk of any managerial non-conformances because:

❖ There is an inability of like professional and managerial groups to rank risk on a consistent basis therefore this is not a reliable system (Conrow, 2003).

❖ It is unclear how to score management systems that are only partially compliant to the required standards.

Table 3   NHS qualitative risk assessment grid

		Probability of Failure				
		Rare	Unlikely	Possible	Likely	Certain
	Insignificant	1	2	3	4	5
	Minor	2	4	6	8	10
Potential Consequences	Moderate	3	6	9	12	15
	Major	4	8	12	16	20
	Catastrophic	5	10	15	20	25

There are various risk assessment techniques in common use (British Standards Institution 1991), but they are time consuming to implement, and rely on an assessment of the probability of failure of the system as a key component of their methodology.

Tversky and Kahneman (1974) have extensively researched the inability of people to accurately predict the percentage probability of an outcome. The effect that this problem has had on various projects and situations has been the subject of much criticism and research (Conrow, 2003). For this reason the use of probability assessments was rejected, and in their place the research of Keeney and Raiffa into the development and use of Utility Functions was employed (Keeney & Raiffa, 1976).

The recognised algorithm dictating the relationship between risk and its components percentage probability of failure and the impact of that failure is:

$$Risk = probability \times impact$$

By holding the impact constant then risk is proportional to the percentage probability of failure. Unfortunately it is not possible to directly measure the percentage probability of failure of a NHS managerial system relating to maintenance. From this the researcher made the assumption that for a managerial system that relied on conformance to national standards, then risk would be at a maximum when the percentage conformance was zero (i.e. no conformance), and at a minimum when the percentage conformance was at (or near) maximum (i.e. 100%) . Therefore the researcher suggested that risk was inversely proportional to the percentage conformance to the national standards for these scenarios.

$$Risk \propto \%Probability \propto 1/\%Compliance$$

Thus by assessing the percentage of compliance (a physical quantity) of a managerial systems critical elements to the respective national standard, then this would form a proxy assessment of the systems percentage probability of failure, and thus of risk. As the two end points were fixed at zero percentage and full (100%) compliance the mid points could be assessed utilising expert elicitation techniques ((Keeney & Raiffa, 1976; Meyer & Booker, 2001; Ayyub, 2001).

Following the observations above the researcher solicited the assistance of five experienced estates managers, to act as experts within the field of NHS maintenance needs and NHS national standards requirements. Their role was to develop a series of

generic utility functions that Trust focus groups could use as a template to aid their judgements relating to the risk posed by a managerial system.

As the process involved the elicitation and collection of 'detailed data' the individual interview face-to-face mode of communication was used as the preferred option in this case, as opposed to use of the telephone or mail, as discussed at the beginning of this chapter. The use of individual as opposed to group elicitation was to remove the possibility of group dynamic effects, and preserving the individual's anonymity denied the participants the opportunity to converse between themselves and compare/influence their observations. The combined outputs of the five experts were then aggregated by the researcher to form the generic utility functions. It was these generic utility functions were to be employed by all Trusts within the analysis, using the diagrammatic method (USA Department of Defence, 2002) when assessing the risk generated by the inadequacies of their managerial systems.

The five experts were judged by the researcher to have equal levels of experience and knowledge of the eight engineering maintenance topics. Each expert was individually interviewed by the researcher and initially asked to review the critical elements and assess the level of impact should each critical element fail in service, by making reference to the NHS 5x5 criticality grid descriptors for potential consequences of failure (impact). Each expert observed that over 95% of the levels of impact assigned to the critical elements were level five, i.e. there would probably be a catastrophic failure. Following this observation each expert agreed that the most important utility function to design first would be that relating to level five.

The researcher then asked each expert to choose a critical element from those attracting an impact score of five at random. Then they were asked to sketch a curve between two axis of risk and percentage compliance to national standards. The experts were instructed that the curve was to represent their assessment of the rate of change in risk as the percentage compliance of the topic under scrutiny rose from zero compliance up to a full (100%) compliance in steps of 10%. This was repeated for ten critical elements. In all cases the general shape of their curves followed a common outline of an exponential form. The researcher then repeated the process two weeks later with each expert (Meyer, Miniszewski & Peaslee, 1989). The curves from each expert were

aggregated using a simple averaging of the scores (Ørtiz, Wheeler & Breeding 1991). The resulting curve suggesting the relationship between risk and percentage compliance took an exponential shape and when subjected to a simple curve fitting exercise an excellent closeness of fit was formed from the algorithm shown below:

Curve algorithm for an impact of five = $(1-e^x)$

Where: x was the product of impact value (five), percentage compliance and a constant.

The researcher then formed the curves (utility functions) for impact of one, two three and four from reference to the algorithm for this initial curve. A further interview took place with each expert to determine their assessment of the five generic utility functions, in terms of their ability to act as templates to guide Trust focus groups decisions, relating to the risk posed by the inadequacies within their managerial systems. Agreement was received from all five experts as to the curves general use. As agreement was given by all five experts, the researcher considered that further development work in this area was not warranted. These five utility functions are shown below in figure 2.

Figure 2   The series of five generic Utility Functions created for and used to develop the Risk Profile.
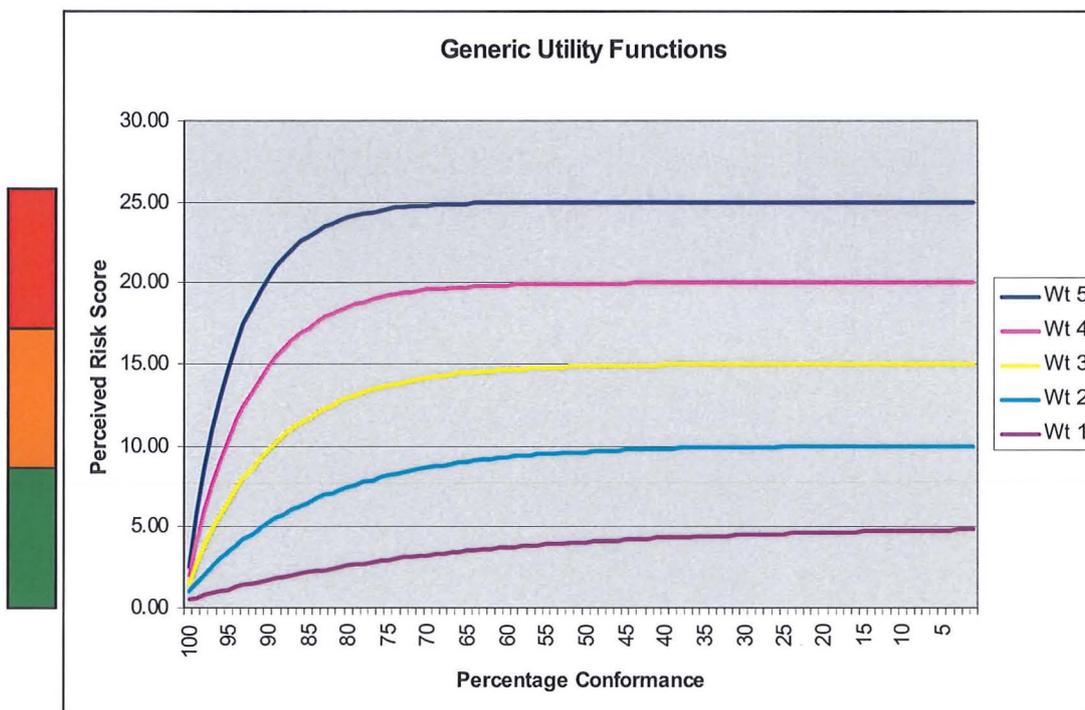
Each of the utility functions was assigned a weighting to reflect the impact of total failure of the management system (i.e. 0% compliance). As the management system improves for a given weighting then the value of the function decreases, thus reducing the value of the perceived risk score, until total compliance of the management system with the recognised standards is achieved at 100%.

The utility functions are shown against the commonly used NHS traffic light system of red, amber and green to denote the seriousness of the risk. The focus group consisting between eight and twelve people from the Trust analysing the management system then followed two steps:

1. Assessed the impact if the management system under investigation were to fail, assuming zero percentage compliance, i.e. there was no management system in place. This determined which one of the five utility functions to employ.

2. Compared the existing management system against the recommended standards across the X axis, and read the assessed risk score from the Y axis.

A trait of the utility functions is that for management systems that are considered to have the greatest impact if failure were to occur (numbers 5 and 4), then their percentage conformance must be very high before the level of risk that they are generating begins to fall below their respective maximum scores. However, this rate of change decreases as the weighting decreases culminating in a gradual decrease in risk score for utility function number 1. The five functions were developed to link into the NHS 5x5 criticality grid definitions for user impact. Therefore, by using the standard algorithm below for risk a simple conversion of the utility function scores was made.

Risk = probability x impact

As the consequences of failure, and the assessed risk score, were known from using the utility functions, then the 5x5 criticality grid value for likelihood was found. This allowed Trusts to enter the resulting scores into their overall risk register for comparison and prioritisation against other identified failings within their organisation. A comparison of the values generated from the five utility functions, and the resulting 5x5 grid scores

developed from the conversion methodology employing the simple equation above, was made to determine the relationship between the two. The graph below (Figure 3) shows pictorially the closeness of fit.

Figure 3   ***A pictorial representation of the relationship between the Utility Functions and the NHS 5x5 criticality grid of perceived risk scores.***
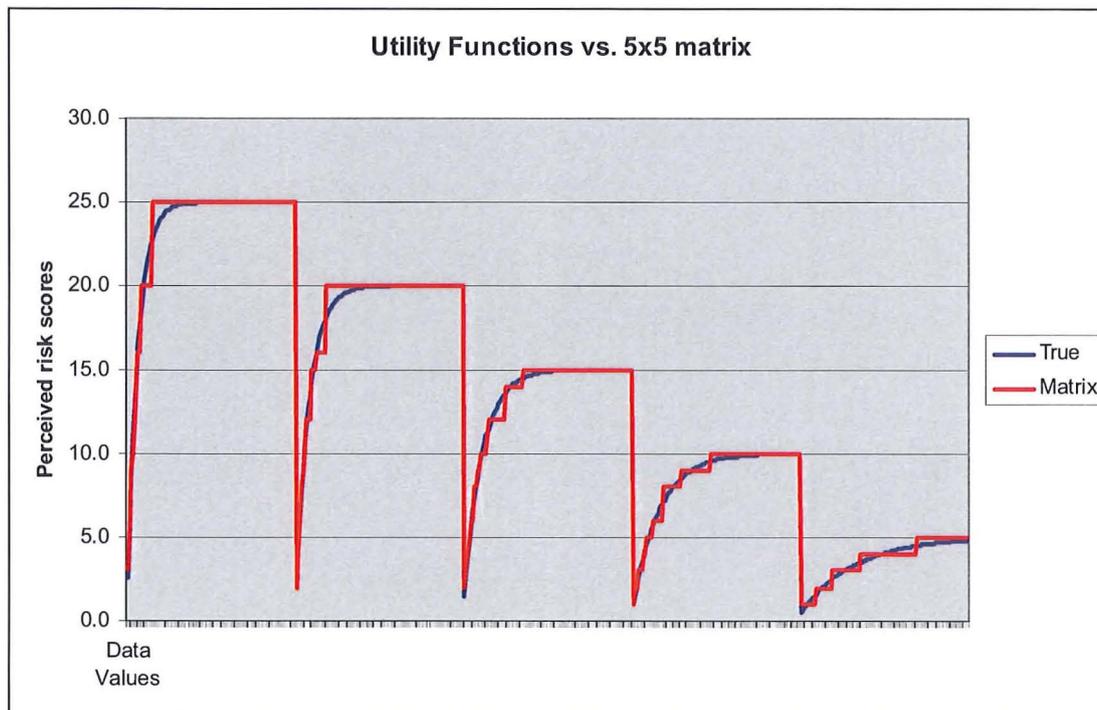
For each of the five utility functions the percentage compliance was varied from 100% down to 0%. This produced the curves shown in blue and labelled – 'True'. Corresponding 5x5 criticality grid values were computed, coloured red and superimposed upon the former and labelled – 'Matrix'. The advantages of using the utility curves were:

❖ Transparency and generation of repeatable assessed risk scores

❖ Ease of understanding and use by the individual(s) and/or evaluation team(s)

When this research began the first ten participating Trusts used a series of six utility functions as opposed to the five shown above. Since these first ten risk profiles were

developed, the researcher has reduced the number of utility functions used by focus groups down from six to five. Since this reduction there has been no change in the process of uncovering any latent conditions by each group, but it has made it easier for them to simulate the utility function shapes with the five criticality grid descriptors for impact. The impact descriptors are the same as those for the weighting used within the model to determine which utility function curve to chose when assessing the risk being generated from a particular critical element. It is the output from these initial ten risk profiles that are presented within the results section.

The focus group employed a desk top exercise only (no on-site inspections/audit) normally spanning one or two consecutive days. The researcher read the wording of each critical element from an overhead screen via a spreadsheet layout. The screen layout used is shown below in table 4.

Table 4   ***The spreadsheet layout for the assessment of risk scores for each critical element.***

Critical Element	Weighting	Percentage Compliance	Probability	Impact	Assessed Risk Score
No 1	5	95	5	3	15
No 2	1	100	1	1	1
etc	4	75	4	5	19

The focus group then discussed the critical element's meaning in these terms, and agreed the impact that would most probably occur if there were no management system in place to regulate the element; i.e. 0% compliance with the existing managerial system.

The level of impact so determined was assigned a weighting on a scale of one to five for impact. This weighting exercise choose which one of the five utility function curves to adopt when assessing the perceived risk that was being generated from this particular critical element. The group then agreed the percentage compliance that the department is believed to be actually achieving from its existing managerial system for this particular critical element. This ranged from 0% to 100% compliance. Algorithms within the overhead display then automatically calculated the resultant assessed risk score and display it on the screen alongside the conversions of the score into the NHS 5x5 criticality grid components of probability and impact.

The score for probability was subject to rounding errors due to the conversion into an integer format. This thesis considers that this slight error was acceptable as it was the assessed risk score that was used within the final analysis. The probability and impact components are only shown to aid the group develop a sense of understanding and trust in the model's design. The display is dynamic in nature, thereby allowing the group to challenge their assumptions and analyse scenarios of 'what if', e.g. what if the percentage compliance was actually 70% when they estimated only 60%. In this way they experienced for themselves the robustness of the model's outputs against their own perceptions of the level of perceived risk that was being generated. All of the critical elements were subjected to this type of assessment. On completion the researcher then reconfigured the data to rank the assessed risk scores into a 'Risk Profile' for their department. Figure 4 below shows a typical Risk Profile for a NHS Trust estate's department.

Figure 4   A typical Risk Profile for a NHS Trust estate's department
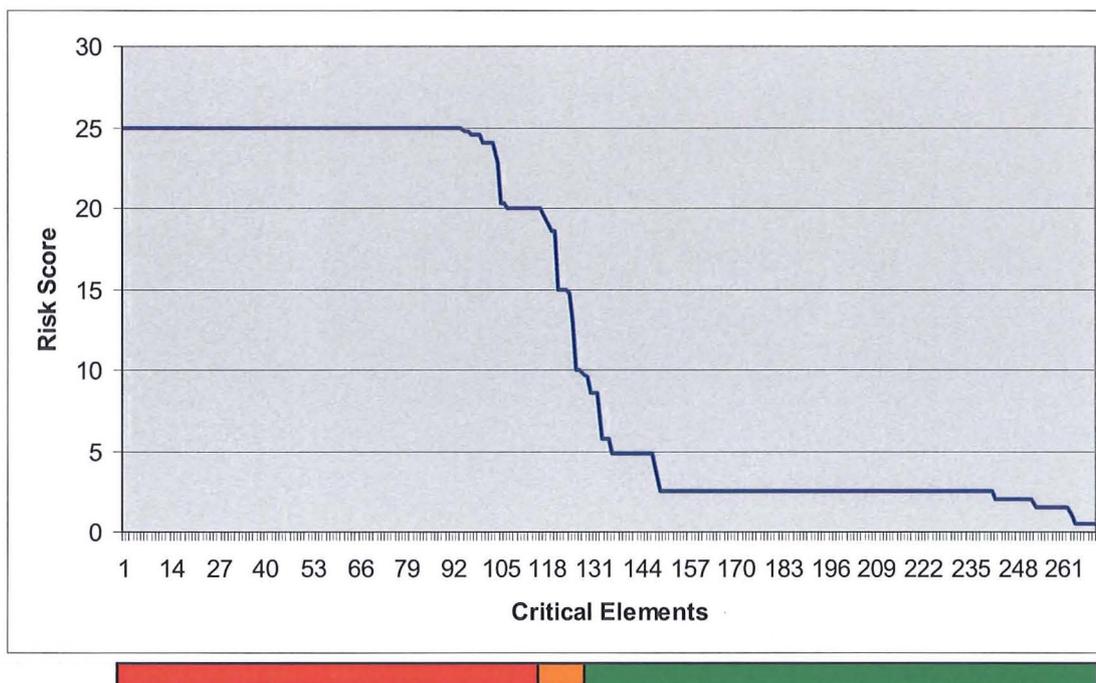
The risk profile was shown alongside the NHS Traffic Light system to give the group a perspective of the scale of assessed risk they believed was being generated. Again at this point they had the option of going back into the body of the spreadsheet to challenge

91

and question some of their assumptions, and see the effect of any changes on the resulting risk profile.

When the group had agreed the final version of the profile, the researcher then presented the profile and salient features to the totality of the estates managers and the shop-floor staff within a large group meeting. The researcher then solicited comment. On receiving general agreement on the output the researcher then held individual meetings with the Trust's specialist advisors to seek their comments and confirmation of the assessed risk profile.

3.2.3 Procedure

For each NHS Trust taking part within this research the process involved the participation of a multi-disciplinary team drawn from the Trust's own senior managers, each acting in the capacity of a specialist advisor. The titles of these managers were typically: -

- ❖ Director of Finance and Chief Internal Auditor;

- ❖ Facilities Manager, Estates Manager and his/her management team;

- ❖ Risk Manager and the Corporate Governance coordinator;

- ❖ Control of Infection officer, Fire officer and the Trust's Microbiologist;

Individual presentations of the research techniques, critical elements and anticipated output were given to each specialist advisor. The reason for individual meetings is to eliminate as far as possible any group dynamics influencing their decisions, comments, etc (Cook & Slack, 1991).

Next a group presentation was given to the estate's senior managers and the shop floor staff. Following this presentation approximately four volunteers were requested from the shop-floor to work alongside the estates managers within a focus group, where the researcher acted in a 'process consultation' role to the group. The specific aim of this focus group was to compare the department's actual activity against the critical elements

identified from the relevant standards, and assess the risk relating to any non-conformances. Following these risk assessments the group was disbanded.

It was considered by the researcher a necessity to form such a group to produce the output, as no single manager would know all of the activities/non-activities being undertaken, and that the situation needed a group interaction to stimulate an informed debate to quickly structure a collective opinion. The management of the focus group and the mechanisms employed to control the main influencing factors affecting their collective judgement, are discussed later within this chapter.

3.2.4 Confirmation of the assessed Risk Profile

An individual meeting with each of the Trust's specialist advisors was then arranged. Within these meetings the facilitator presented an overview of the work to-date, and summarised the findings/discussions. The specialist advisors were then invited to comment and to challenge the assumptions made, on the resulting assessed risk scores and the overall profile by making reference to the base data. When agreement had been reached the assessed risk profile, was confirmed, and the output was regarded as a reasonable representation of the assessed risks that the Trust, patients and staff were being exposed to. A short presentation was given by the researcher to the Trust Board members, of the methodology and the output from the multi-disciplinary group. The Board then decided what course of action to take next.

The number of active participants within each of the 31 Trusts contributing to the data collection for this first research objective was between eight to twelve people within the focus groups, and involved six to eight specialist advisors. Thus between 496 and 620 experts were consulted as part of this first research objective data gathering exercise.

3.2.5 The main influencing criteria on the risk score.

When designing the methodology used to expose the latent risks within the estates department's managerial systems, reference was made to the criteria contained within the literature search and the researchers personal experience that would influence the Trust's assessment of their risk score (Meyer & Booker, 2001). Table 5 below lists these main influencing criteria.

Table 5 *The main influencing criteria upon the risk score.*

Influencing Criteria
Actions of the researcher
Group dynamics
Understanding of methodology
Shape of utility functions
Knowledge of existing managerial systems
Assessment of weighting
Cross-over of skills
Anchoring

a.   Actions of the researcher

The methodology adopts a series of utility functions, which assists the focus groups to assess the risk that their managerial systems are generating. These utility functions were designed to mimic the rate of change of risk with respect to the percentage compliance of a management system. As this technique was new to the Trusts' involved, the researcher helped the groups to manage the assessment process.

Although the term researcher is used here, the actual methodology employed Process Consultation (Schein 1997, Schein 1998). This system differs from facilitation techniques, in that the researcher is not regarded as an 'expert' being engaged to solve some predetermined problem. Nor is he/she acting in a 'patient-doctor' role, where he/she is being asked to determine what the problem is, and to then to develop a cure. By adopting the process consultation role, the client (the Trust) is guided into retaining both the diagnostic and remedial initiative. This type of approach is much praised and practised by Belton and Stewart (Belton and Stewart 2003) in their work relating to similar managerial focus group activities.

Throughout the development of all the risk profiles, the same researcher (the author) conducted all seminars, focus groups and data analysis. Thus, continuity throughout the study could be regarded as constant. The main role of the researcher was to: -

❖ Build a relationship with each focus group i.e. a psychological contact

❖ Train each group in the basic concepts and use of the methodology.

❖ Bring to the focus groups some 'content knowledge', of the technical and/or organisational issues surrounding the managerial systems.

❖ Refrain from directly contributing to the discussions and decision making processes, and remain strictly neutral.

Therefore, the researcher's responsibility was to help the focus groups to perceive and understand the task in hand, by guiding them towards developing a common understanding of the situation.

b. Group dynamics

The methodology employed to develop the risk profile for a Trust split naturally into four discrete parts:

Initial presentations

This first part described the objectives and processes for the development of the risk profile. Focus group meetings are attended by all of the departments' engineering officers and artisan staff. One-to-one discussions were reserved for the Trust's specialist advisors. As the initial meetings were only to introduce the overall concept and to describe the output of the model, there was little room for any adverse group dynamic activity (Cook & Slack, 1991).

Workshop

Representatives from the department's engineering and artisan staff attended this focus group session. The total numbers within a group ranged between eight and ten representatives, normally evenly split between the managerial and shop floor staff. It was in this arena that the group assigned weightings and percentage compliance values to each critical element in turn. From these assessments the overall risk profile was formed. The workshop created a potential forum for adverse behaviour to develop, but the participants were advised of this possibility, together with the need for the researcher to monitor and control the proceedings, whilst being totally impartial throughout.

In all cases the outputs from the groups were ratified by the Trust's specialist advisors before being finalised. None of the focus groups' risk assessments were rejected by the specialist advisors, therefore, showing the success of the process to control the effects of group dynamics.

Final presentations to specialist advisors

These meetings took place between the researcher and each of the Trust's specialist advisors, on a one-to-one basis. Thus the group dynamic influence was automatically removed, as previously discussed.

Final departmental meeting

The last presentation was to the total departmental workforce as a single group. The researcher advised the group of the salient points raised in the development of the risk profile. This final meeting was more of an advisory presentation to all people within the department, but was also used as an opportunity for any last minute discussion.

c. Understanding of methodology

Without understanding the basic concepts of the methodology, then risk could not be assessed with any degree of confidence. Therefore, a detailed explanation of the methodology was presented by the researcher to all of the participants, and their views/agreements sought 'before' the creation of the risk profile began.

d. Shape of utility functions

When using the utility functions to create the risk profile, the participants within the workshops and specialist advisors were encouraged to alter the weightings and percentage compliance values to see the effect on the risk score. In this way the sensitivity of their responses were immediately demonstrated. This gave the groups confidence in the robustness of the final risk score developed (Armstrong 1985), and as such they have reallocated existing resources to close the identified managerial gaps.

e. Knowledge of existing managerial systems

This criterion is only applicable when assessing the percentage conformance of the existing managerial systems when viewed against the national standards required.

Therefore, the main contributors to this part of the model were the engineering officers and artisan staffs within the workshop groups. By combining both types of personnel together within a single focus group, the lack of knowledge of one part of the group was counter-balanced by the other.

f.   Assessment of weighting

An assessment of the weighting took into account the 'as-fitted-infrastructure' and corresponding managerial system; combined with the perceived effects on the patient should that unique system fail. As these three components affected each Trust differently, then within the sample group of Trusts studied the weighting for each critical element could not be expected to be similar from Trust-to-Trust. An alternative interpretation for such a skewed weighting distribution would be to suggest that some Trusts did not assess the importance of the critical element correctly. That in turn is to suggest that the combined experience of multi-disciplinary teams within a number of Trusts totally failed to appreciate the seriousness of the risk being posed by their managerial systems. It is highly improbable that one (let alone a number of) highly qualified and experienced multi-disciplinary focus groups could produce consistently incorrect assessments.

g.   Cross-over of skills

Each of the Trusts formed a focus group to assess the risk generated from the managerial systems practiced. These consisted of a core number of engineering officers and artisans, supported by specialist advisors. A common trait between each member of the groups was that each possessed a certain level of technical and managerial knowledge/skill of the other professions. This gave an inbuilt 'safety mechanism' that checked each assessment made regarding both the weighting and percentage compliance of each critical element.

h.   Anchoring

Anchoring has been explicitly defined as 'bias in judgements under uncertainty' (Tversky & Kahneman, 1974). As the entire risk profile is based on qualitative and not quantitative judgements, the effects of anchoring on the robustness of the risk profile are explored below. Two forms of anchoring are discussed, firstly that caused by suggestion and secondly by a near experience.

97

Two disparate groups within each Trust assessed the risk that was being generated from the managerial systems within the estates departments. The contribution that each of the two groups made to the process are discussed below: -

The focus group was composed of engineering officers and representatives from the shop floor. They (as a group) had extensive and detailed practical experience of the Trust's as-fitted-infrastructure and its probable impact of failure on the patient population. This was complemented with a good appreciation of the existing managerial systems and their compliance to the national standards and guidelines.

The specialist advisors all possessed excellent theoretical knowledge within their specified areas, and had all practiced their profession in the field for a number of years, at a high level of managerial seniority within the NHS. These facts in turn gave them a high level of understanding of the probable impact that the failure of any critical element would have on the patient. They did not have (either as a group or as individuals) a high degree of familiarity with the existing managerial systems, or their compliance with national standards or guidelines.

Table 6 below summarises the contribution that each of the two groups made to the assessment process for both the weighting and percentage compliance components of the utility functions.

Table 6   **The contribution of each group to the overall assessment of risk.**

Group	Knowledge of 'as fitted infrastructure'	Appreciation of probable impact of failure on patients	Percentage compliance of the managerial systems
Focus Group	High	Medium	High
Trust specialist advisors	Low	High	Low

The skills within the two columns to the left of table 6 influenced the weighting of the utility functions. The levels of expertise within the right hand column had an effect on the assessment of the existing managerial systems compliance with national standards and guidance.

The focus group initially assessed the risk of each critical element using the utility functions. This assessment was then tested by the researcher and the group by varying both the weighting and percentage compliance values on the risk score to the group. This greatly aided the discussions and belief that their final assessment of risk was robust. The specialist advisors were then presented (on a one-to-one basis) with a summary of the findings of the focus group. This data was supported with the option for them to view the detailed analysis developed, and to undertake simulations on the assessments. Therefore by combining the skills of both groups a robust assessment of the overall risk profile was developed. It has been suggested by Mussweiler and Strack (2000) that:

> The amount of knowledge a judge has about a target determines the magnitude of the anchoring effect. Thus, a small distribution about the mean is developed when a high knowledge base exists, the reverse being true when a low level of expertise is available.

The focus group created the first estimates of risk. Table 6 shows the considerable skill base that the focus group members possessed. The output from the workshop created an anchor for the specialist advisors, but as the second group held a very high knowledge base, the magnitude of the anchoring effect was probably negligible.

The outputs from the workshops were questioned by the specialist advisor groups throughout the development of all of the risk profiles. But these discussions centred more on areas of clarification rather than on direct challenges to the risk score assessed. In all cases none of the risk assessments developed by the focus groups were altered by the specialist advisors. This agreement by the two groups supports the work of Mussweiler and Stack (2000).

The second form of anchoring referenced as a near experience, would be created by an 'incident or near miss' occurring in recent times relating to one or more critical elements. This would heighten awareness of both groups within a Trust, which in turn could influence judgement. No known anchoring effects were discussed within the focus groups during the development of the risk profiles.

The results from this first research objective are shown within chapter four of this thesis and confirm without doubt that:

> The managerial systems of internal control within the National Health Service (NHS) **do not** give adequate assurance that the day-to-day building service engineering maintenance activities conform to their respective national guidelines.

Therefore the researcher developed methodologies to address the second, third and fourth research objectives.

3.3 Development of the second research instrument

The second research objective developed a robust, transparent and repeatable methodology to assess the resource needs to correct the existing managerial system deficiencies to that required by their relevant National Health Service standard, via risk treatment plans with the specific aim of reducing the assessed risks to levels acceptable to the Trust. Not all of the 31 Trusts participating within the first research objective considered that they needed to undertake such a detailed study to assess their resource need as required within this second objective. Therefore the number of Trusts contributing data to this second research objective was reduced from 31 to 27 Trusts.

For the output of this second objective the researcher developed and presented a hypothetical assessment of the likely resource needs to close the managerial gaps identified within the red and amber zones, to the Trust estates senior managers (experts) again using expert elicitation techniques, for their discussion and adjustment to meet their local conditions. These outputs were then subjected to a Monty Carlo simulation as recommended by HM Treasury to allow for uncertainty within the expert's resource assessments (HM Treasury 2003). Only these latent conditions within the red and amber zones were analysed as it is they that were believed to be posing the greatest risks. The three resources included within the study are outlined below:

Manpower

It was assumed that a competent engineering officer was assigned to the task of closing the identified gaps for 100% of his/her time.

Outsource Cost

These figures related to any specialist services that were procured from consultants and/or contractors to assist with the closure of the gaps.

Timeframe

Two types of timeframe were developed. The first related to an estimate of the 'actual' time the engineering officer would expend in completing the managerial tasks to close the gaps. The second referred to the estimated 'process' time that naturally occurs to complete all of the tasks to close the gaps. This included delays, such as co-ordinating diaries for a meeting, contractor set up time before undertaking a task, etc.

First the manpower and outsource costs were assessed in parallel (as explained below), then the overall timeframe was estimated by combining the 'actual' and 'process' time scales together using an industry standard Gantt Chart software package.

The managerial gaps within the red and amber zones were grouped into the eight topics e.g. water, electricity, etc, as discussed in section 1.5. Then, they were further subdivided into practical clusters. This approach can be exemplified using the piped medical gas topic as an example: -

Assume that a major potential problem had been identified with the accuracy and quantity of the 'as fitted' drawings for the site. Also, assume that not all of the isolating valves were numbered correctly. From a practical viewpoint these two critical elements were combined into one work task; because, as the drawings were brought up to date, then the renumbering of the isolating valves would also be undertaken.

Once the practical clusters were formed, then each cluster in turn was sub-divided into work tasks and split between engineering officer and outsource tasks. Continuing with the Piped Medical Gas problem just described, table 7 below shows an example of a partially finished assessment.

101

The man day estimations were restricted to three days or less. If it was considered that the time to complete the task would probably be greater than this, then, that particular task was again further subdivided until all of the tasks showed less than three days. The reason for splitting these tasks up to identify those of less than three days was that, in the opinion and experience of the researcher, people have great difficulty assessing more than three day's output with any reliable degree of accuracy.

The time scales assigned to each task were estimated using professional judgement and experience only, and represent the mean 'actual' time that the engineering officer would be expected to take. Unlike the construction industry that has national norms associated with artisan tasks (e.g. a bricklayer can lay 1000 bricks/day in mid-summer), there is no such central reference for the type of managerial tasks described above.

Table 7   *Example of a partially completed assessment for Piped Medical Gas*

Tasks	Engineering Officer (man days)	Outsource Costs (£)	Confidence Level (L, M or H)
PMG Record Drawings			
Engineering Officer			
Develop specification & tender documents	2		H
Issue tenders	0.5		H
Tender analysis	1		H
Meeting with the preferred bidder	1		H
Management of contractor on site assume 2hrs/cuircuit for 15 circuits say 4 days	4		M
Mop up session with contractor	1		H
Outsource Cost			
Survey 15 circuits say 1 circuits/day at £300/day for 2 men		9000	L
Develop CAD records say 4 circuits/day at £250/day		1000	M

3.3.1 Procedure

Against each of the estimations a level of confidence was assigned by between five and eight engineering officers operating within a focus group. The reasons for the use of a focus group as opposed to other forms of expert elicitation processes have been previously discussed at the beginning of this chapter. These confidence levels range

between Low, Medium and High, and represented the assessor's view of his/her accuracy. These levels were used to conduct a Monte Carlo simulation of the final agreed means, as described later in this chapter.

Once all of the clusters had been assessed in this way, their values were transferred into a Gantt chart where their respective process times were added. The tasks and sub-tasks were linked together where necessary, and the resources levelled to avoid over allocation and work peaks. The resulting documents were presented to the estates department senior officers for comment to ensure that any local site anomalies were accounted for before the Monte Carlo analysis began. Following agreement between the parties then a standard Monte Carlo simulation was undertaken.

As each task/sub-task was allocated a level of confidence, a probability distribution was automatically assigned. The shape of the distribution was determined from the level of confidence. This research used a triangle style distribution as this is commonly used for rough modelling, and has been found to be acceptable to Trusts. The spread about the mean of the distributions differed for each of the three levels of confidence as shown below:

Low   -20%, +20%

Medium   -20%, +25%

High   -20%, +30%

The reason for the skewed distribution for the medium and low levels was to reflect the group interaction effects and the ability of people to accurately predict the effort required to complete difficult and/or unfamiliar tasks. This research work created these distribution limits from experience in the field, and they have been accepted by all of the 27 Trusts and between the 135 to 216 engineering officers taking part within this second research objective. A typical example of this form of final output from an off-the-shelf software package is shown in figure 5 below. This shows that the analysis estimates that 409 man-days of an engineering officer's time were required to close all of the gaps that fell within the red and amber zones identified within the risk profile.

Figure 5   *Typical Monte Carlo output for the estimated totality of an*
engineering officer's actual time needed to close the managerial gaps identified
within the risk profile.
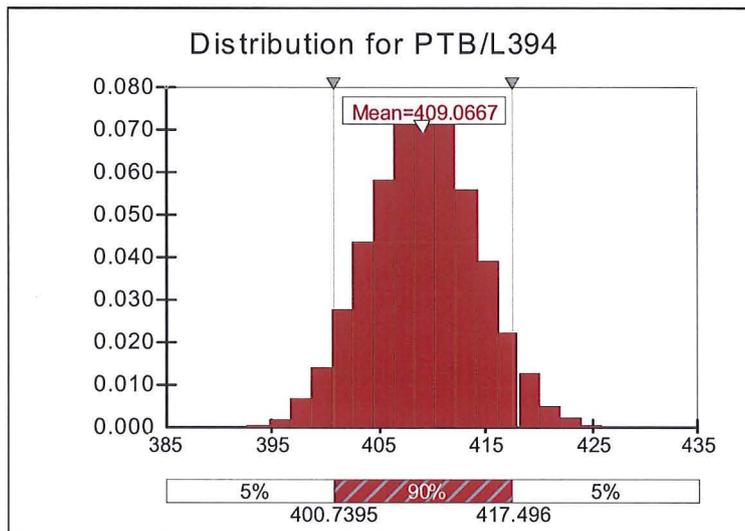
This section only assessed the resource needs to 'close' the identified gaps. It did not attempt to estimate the revenue resources required to 'keep' the gaps closed. From the +30 years experience of the researcher as an NHS estates manager it was considered that the bulk of the essential resources involved to keep these gaps closed would relate to the artisan activities of both contract and directly employed labour. The methodology employed to assess this type of revenue resource needs is described in the next section of this chapter that addresses the third research objective, which is to:

> Develop a robust, transparent and repeatable methodology to assess the directly employed artisan revenue resources needed to ensure that the maintenance activities conform to the National Health Service standards.

3.4   Development of the third research instrument

Section 2.4 discussed the lack of an NHS centralised maintenance data base from which to draw quantitative information relating to the maintenance function within healthcare organisations. This section also highlighted the research work of Labib (2004) and the inability of computerised maintenance management systems to have any decision making capability. Section 2.4.1 discussed the role and need for maintenance simulation models and the dynamic nature of the maintenance function (Sherwin, 2000), and the

problems of trying to use historical information (Moubray, 1997). This is followed by section 2.4.2 that cites the work of Al-Zubaidi (1997) in relation to the dynamic and unpredictable nature of reactive maintenance within a healthcare environment, and the work of Zubaidi and Christer (1997) in relation to the need, design and use of maintenance simulation models within a healthcare setting, together with the research work of Mjema (2002) employing simulation techniques within a manufacturing situation facing similar problems. Section 2.7 introduces the research work of Rudolf and Repenning (2003) and their simulation of non-novel interruptions on an organisations outputs (similar to the environment within healthcare maintenance organisations) which was developed using grounded theory (Glaser & Strauss, 1968).

From reference to the above, the development of the third research instrument began with the design by the researcher of a 'proof of concept model' based on the research work of Rudolf and Repenning (2002) incorporating the findings of Al-Zubaidi (1997) and Zubaidi and Christer (1997). From the success of this concept model, detailed discussions were then held with each of the eight Trust's participating within this section of the research. These discussions were undertaken within eight separate focus groups (one for each Trust). In each case the participants were a multi-professional group consisting of the Trust's senior managers and were regarded as 'experts', e.g. the estate's, risk, fire and internal audit managers. The use of focus groups and the reasoning for employing this type of expert elicitation technique has been discussed at the beginning of this chapter.

These focus groups were formed to develop a clear understanding of the problem to be solved which was the output from the third research objective:

> Develop a robust, transparent and repeatable methodology to assess the directly employed artisan revenue resources needed to ensure that the maintenance activities conform to the National Health Service standards.

It was agreed by the group that the research instrument to be used to create this output was to be a simulation model based on the successful demonstration by the researcher of the 'proof of concept' model.

The group then defined and agreed the inputs and outputs required by a detailed simulation program, and the assumptions and algorithms to be employed, i.e. effectively forming a brief for the researcher to follow. The researcher then converted these requirements into a computer simulation program, tested the program for correct operation and produced the outputs for analysis. The individual focus groups then discussed these outputs against their original brief and used this information to inform their decisions, in relation to the allocation of their limited trust wide resources to where they would do most good. This in effect allowed the focus group members to determine the Trust's risk appetite, and form an output specification for their maintenance department.

The following paragraphs discuss in detail the development of this third research instrument, the techniques employed, experts involvement and outputs, beginning with the creation of the 'proof of concept model'.

3.4.1 Procedure

Within the scope of this research the generation of tasks for the day-to-day operations and maintenance of an NHS estates department involves data from two discrete areas. That data required to comply with national policies which take the form of Planned Preventative Maintenance programs, and the data concerned with local reactive maintenance needs of the Trust's buildings. The rate at which these reactive maintenance tasks are required to be completed on a week-by-week basis over the course of a calendar year is not linear. It has been stated by Al-Zubaidi and Christer (1997) that this work flow into the department is 'random in time, random in nature and random in location' (Al-Zubaidi, 1997; Al-Zubaidi & Christer 1997).

The workforce to complete these combined planned preventative maintenance and reactive tasks is a hybrid mix of a Trust's directly employed labour (e.g. electricians, fitters, plumbers, etc) and contractors. For the most part contractors are rarely used to undertake reactive maintenance tasks, due to their lack of local knowledge in terms of both the 'as fitted' infrastructure and protocol relating to their interaction with medical/clinical staffs and patients. Therefore their workload (in the vast majority of cases) relates to some planned preventative maintenance tasks only.

106

Eight Trusts participated within the scope of this third research objective. From interviews with their respective estates managers it was agreed that in the vast majority of cases when resources were scarce (e.g. lack of immediately available manpower), the reactive maintenance tasks are completed as priority issues over those scheduled within the planned maintenance programme; this leads to some planned maintenance tasks not being completed.

When a planned preventative maintenance task is undertaken on an item of plant and/or equipment and the results are found to be within the acceptable limits of the specification, then the risk posed to the organisation, etc, is regarded as being acceptable. As time progresses to the next planned inspection/test/examination the risk will naturally begin to rise up to the next time period for inspection. At this point the risk will have risen to its acceptable maximum; it is this value that is termed the residual risk. This 'saw tooth' type rise and fall between these inspections would in the normal cause of events continue throughout the lifespan of the plant/equipment.

If the next planned inspection were not undertaken then the risk would continue to rise with time and the plant/equipment would be classed as being in potential failure mode. The level of the impact of the failure, should it materialise, would depend on the type and location of the plant/equipment that had actually failed. As the purpose of a planned maintenance programme is to reduce the likelihood of a building's plant and/or equipment failing to perform within a predefined set of limits, then the inability to detect any such plant/equipment failure will increase risk to the organisation. Two types of planned preventative maintenance task are undertaken within the NHS. The first is an invasive inspection e.g. the changing of a filter. This first type of planned preventative maintenance task is a true planned preventative maintenance action, as it prevents the system from deteriorating beyond a pre-determined tolerance level. The second type of planned preventative maintenance is a non-invasive inspection. This addresses the condition of the system and is designed to record readings, for example temperature and/or pressure. The term planned preventative maintenance that is assigned to these non-invasive tasks is historic and dates back to when planned preventative maintenance was first introduced into the NHS. In these modern times this term should not be used as these types of actions do not 'prevent' any deterioration of the system. They merely confirm (or not) that the system is operating between acceptable limits. Figure 6 below

was derived from a similar sketch within the international guidance for the Australian/New Zealand risk management standard (Standards New Zealand 2005). It depicts the fall in perceived risk as the available budget (resource) is increased. The fall in risk is assumed to level as the budget for the organisation approaches the ideal, where the ideal point is regarded as the budget allocation that funds all of the tasks necessary to match both national and local standards. At this point the remaining risk experienced by the organisation/stakeholders, etc is that of the residual risk and that any budget increase after the ideal point will have negligible impact in risk reduction. The figure also depicts the level of perceived risk where an organisation under-funds their maintenance activity below the ideal point, and where they are not putting their limited resources to where they will do the most good.

Figure 6   *The effects of varying budget allocation with respect to the perceived risk.*
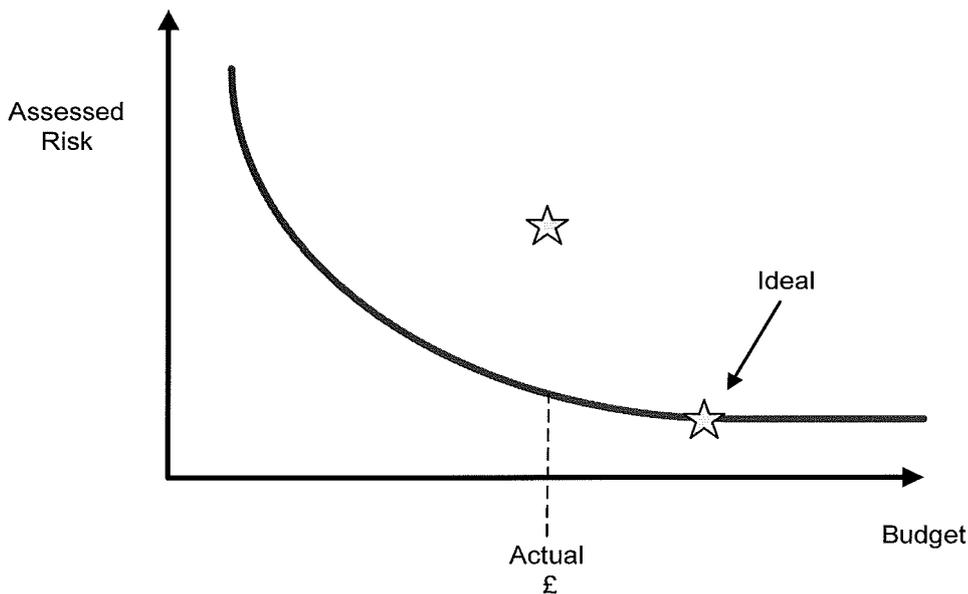
As shown within the results chapter of this research, as resources are withdrawn from the planned preventative maintenance system (invasive and non-invasive) the risk rises until the overall level becomes unacceptable to the Trust, i.e. the 'Tipping Point' is breached.

Rudolph and Repenning (2002) developed a dynamic model to simulate an organisation and investigate the elements affecting this Tipping Point. Their model which was

described within the Literature review chapter of this thesis, attempted to mimic the actions of an organisation that was so overwhelmed by a series of non-novel interruptions (tasks) over a short period of time, that for those involved the levels of stress exceeded the Tipping Point (as theorised by the Yerkes-Dodson Law), and they were unable to clear all of their tasks within the allotted time, which led to a disaster occurring. From reference to the Rudolph and Repenning model, a simple representation was designed by the researcher as a 'proof of concept model' to reflect the situation that occurs within a typical NHS Trust estates department when resources are withdrawn and the assessed risk begins to rise. This was achieved by keeping the same terminology but making the following two fundamental assumptions. In summary these were:

a. Within the original model a time delay was developed to mimic the period between the need for the organisation to increase their rate of resolving interruptions/minute, and the organisation realising that they needed to increase this rate. As the timeframe for the estates simulation is operating with a timeframe of weeks as opposed to minutes, the need to include this realisation time was removed and treated as a constant.

b. The NHS has for a number of years operated a national productivity scheme to monitor the output of the artisan workforce. If their performance output fell then the shop floor wages were affected accordingly. This monitoring scheme was totally disbanded in the year 2006 from all NHS Trusts. From the evidence from two large acute hospital sample sites it was observed that these productivity levels were unaffected by workload pending. Therefore based on this evidence the effects on stress have been neutralised within this simulation programme.

The two assumptions above reduce the original Rudolph and Repenning model to one of a simple stock and flow structure as shown below.

Figure 7   The basic stock and flow structure of interruptions in organisations as reproduced from the Rudolph and Repenning paper (2002).

Interruption Arrival
Rate (inflow)

Interruption Resolution
Rate (outflow)
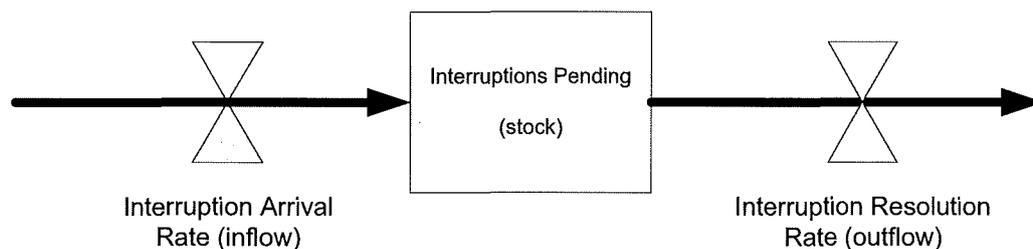
In the simplified estates model the Interruption Arrival Rate (inflow) was the number of man hours required to complete all of the tasks for a given week. This arrival rate was not constant but varied dependant on the summation of the reactive maintenance needs and the planned preventative maintenance tasks for the week, again assessed in hours. The Interruption Resolution Rate (outflow) is the rate that the workforce could complete the set tasks, measured in hours/week. This rate again varies from week-to-week as it is affected by staff holidays, sickness, unfilled vacancies, etc.

Any tasks that were unable to be completed for a given week remain as stock in the system (Interruptions Pending). During normal operations both the inflow and outflow varied independently, and slack time was developed that allowed the stock of tasks to be reduced. If the manpower of the organisation was reduced the stock of tasks built up. It was stated earlier that the reactive maintenance tasks were always completed before any planned preventative maintenance tasks, therefore this build up of stock was actually a backlog of uncompleted planned preventative maintenance tasks. If planned preventative maintenance work was not completed then the risk posed to the organisation, staff, stakeholders, etc, began to increase.

Therefore a Microsoft 'Excel 2003 macro' was designed by the researcher to simulate the simplified estates 'stock and flow' model just described so to observe the effects of risk for varying levels of manpower availability. This was chosen as the preferred software media for a number of significant reasons, but namely the ability to demonstrate transparency of the model's algorithms and assumptions to the Trusts involved as this software is commonly used throughout NHS organisations (Armstrong 1985).

In order to give a sense of authenticity the simple estates 'proof of concept' model was developed from actual data relating to the planned preventative maintenance tasks for the fitting and plumbing trades from the 650 bed Hospital referenced as site S within this thesis in order to preserve their anonymity. This gave an average workforce need of 56 hrs per week over a year, but the need varied between 104.5 and 31 hrs/week. In order for a level of risk to be computed should a planned preventative maintenance task fail to be completed, values of impact were assessed, using the NHS 5x5 criticality grid. A cumulative probability distribution profile was also assessed for each task, should the frequency of inspection be missed.

For the dummy data it was assumed that a hospital's planned preventative maintenance system accounted for approximately 40% of the total day-to-day workload, the remaining 60% being reactive maintenance. From the work of Al-Zubaidi and Christer (1997) they stated that the reactive maintenance workload for an acute hospital was random, therefore for the purposes of the test a reactive workload was developed with a mean of 84 hrs/week over a year, but varying between 97 and 84 hrs/week. The combined planned preventative maintenance and reactive workload formed the 'dummy' input to the simulation model. Eleven simulations were run for varying constant levels of manpower availability from 70 hrs/week to 170 hrs/week in steps of 10 hrs for a two year timeframe. It was assumed that at the start of each simulation the whole planned preventative maintenance system had previously been fully resourced and that the only risk that was being generated was to the level of the residual risk. Thus a common starting point was formed for each simulation. The resulting levels of risk generated for each step down in manpower availability were recorded, and the graph in figure 8 obtained.

Figure 8   *Output from the simulation of reducing manpower availability and the effect on the organisations level of risk for the simplified estates model.*
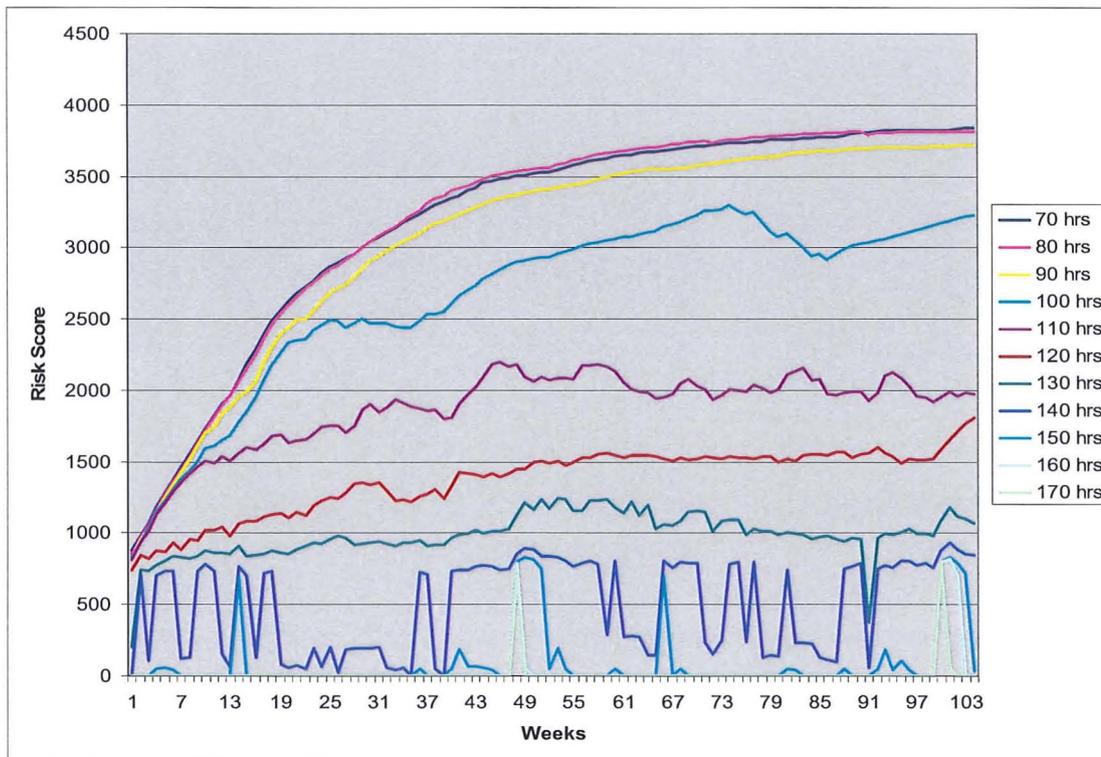
Rudolph and Repenning (2002) developed the central notion within their analysis that 'organisations can have tipping points beyond which their response to interruptions changes dramatically'. In the simple estates simulation model the term interruptions was changed to refer to maintenance tasks to be completed. As the available manpower resource reduced the tipping point was breached and the estates response was seen to change dramatically.

In the original Rudolph and Repenning model the interruption arrival rate was varied whilst the organisation's manpower to deal with the interruptions was constant. In the dummy estates model the reverse was true. The interruption arrival rate (maintenance tasks per week) was held constant but the manpower availability to undertake the maintenance tasks was varied. At a maximum manpower level of availability, the peaks and troughs of the maintenance tasks arriving and the corresponding peaks and troughs of the reactive maintenance needs created opportunities for the system to regain a state of relative equilibrium. In the beginning, as the manpower availability was reduced with

112

each simulation run the ability of the organisation to deal with the incoming tasks was not significantly effected, as any increase in resulting risk from the stock of uncompleted maintenance tasks was negligible. However once the available manpower was reduced to between 140 and 150 hrs per week the system quickly began to descend into crisis, in a similar way to the original Rudolph and Repenning model. They observed that from the system dynamics of their model:

* ❖ The Tipping Point only needs to be exceeded for a moment to create disaster.

* ❖ The wider the variation in arrival rate the more likely the tipping point threshold will be crossed.

* ❖ The system collapses rapidly once the threshold is exceeded.

* ❖ The model implies that even systems in low-variability environments face the possibility of a quantity induced crisis.

Identical observations to those of Rudolph and Repenning were made from varying the parameters of the dummy estates model in relation to the input (maintenance tasks) and available manpower, simulating the system and recording the rate of rise of risk.

Based on the results from the simulation of the dummy estates data, it was assumed that for a fully resourced workforce, the only risk that would be generated would be the residual risk from each individual asset. This assumption was found to be totally false when a Trust's actual 'data' was simulated, as demonstrated and explained within the results chapter of this thesis.

Following the apparent success of the simple estates 'proof of concept model' to predict the effects of varying resource levels and resultant generation of risk, it was decided by the researcher to enhance the model and use it with a Trust's 'actual data'.

Sterman eloquently describes the impossibility of being able to 'verify and validate' a model's ability to mimic reality as all models are wrong, and are in his words, "limited,

113

simplifications of the real world", and they, "differ from reality in ways large and small, infinite in number". He goes on to illuminate his observations by citing a number of past studies in this subject. By accepting this premise, then the model created for this thesis was subjected to a rigorous testing programme designed specifically to uncover errors, so that a greater understanding of the model's limitations could be formed. This is in stark contrast to the problem often suffered by model designers and their clients which is referred to by Sterman as 'confirmation bias'; in these cases results are selected that are favourable to their preconceptions (Sterman, 2000). As the various parts of the simulation process are described within this methodology chapter, the assumptions and known limitations at the time are highlighted. Others were uncovered only following the simulation of 'actual data', which gave a greater understanding of the processes at work. These latter observations are described in detail within the results and discussion chapters.

The whole model process and output results were designed to simulate the current time horizon, and assumed that there would be no significant change within any endogenous elements that could affect the model. Figure 9 below gives a pictorial overview of the procedure employed to simulate the effects of varying the available manpower resources for an estates department's artisan directly employed labour force, and to analyse the resulting rise (or fall) of risk to the organisation, staff, patients and stakeholders. The figure is followed by a brief explanation of the various steps throughout the process.

Figure 9   An overview of the simulation process

```
                    ┌──────────────────────┐
                    │  Assesses the effect of │
                    │   resource variation    │
                    └──────────────────────┘
         ┌──────────────────┼──────────────────────┐
         ▼                  ▼                       ▼
┌─────────────────┐                    ┌──────────────────────┐
│ Establish PPM   │                    │                      │
│ requirements    │                    │  Gather actual Trust │
└─────────────────┘                    │  available manpower  │
         │          ┌──────────────────┐   and reactive       │
         ▼          │                  │   maintenance data   │
┌─────────────────┐ │ Construct software│                     │
│ Develop PPM     │ │  model and test   │                     │
│ schedule and    │ │                  │                      │
│ resource need   │ │                  │                      │
└─────────────────┘ │                  │                      │
         │          │                  │                      │
         ▼          │                  │                      │
┌─────────────────┐ │                  │                      │
│ Assess cumulative│ │                  │                      │
│ probability     │ │                  │                      │
│ distribution    │ │                  │                      │
│ profile         │ │                  │                      │
└─────────────────┘ └──────────────────┘                     │
         │                              └──────────────────────┘
         ▼
┌─────────────────┐
│ Gauge impact    │
│ descriptor ratios│
└─────────────────┘
         └──────────────┬──────────────────┘
                        ▼
              ┌──────────────────┐
              │    Simulate      │
              └──────────────────┘
                        │
                        ▼
              ┌──────────────────┐
              │ Presentation and │
              │ interpretation of│
              │ results          │
              └──────────────────┘
```
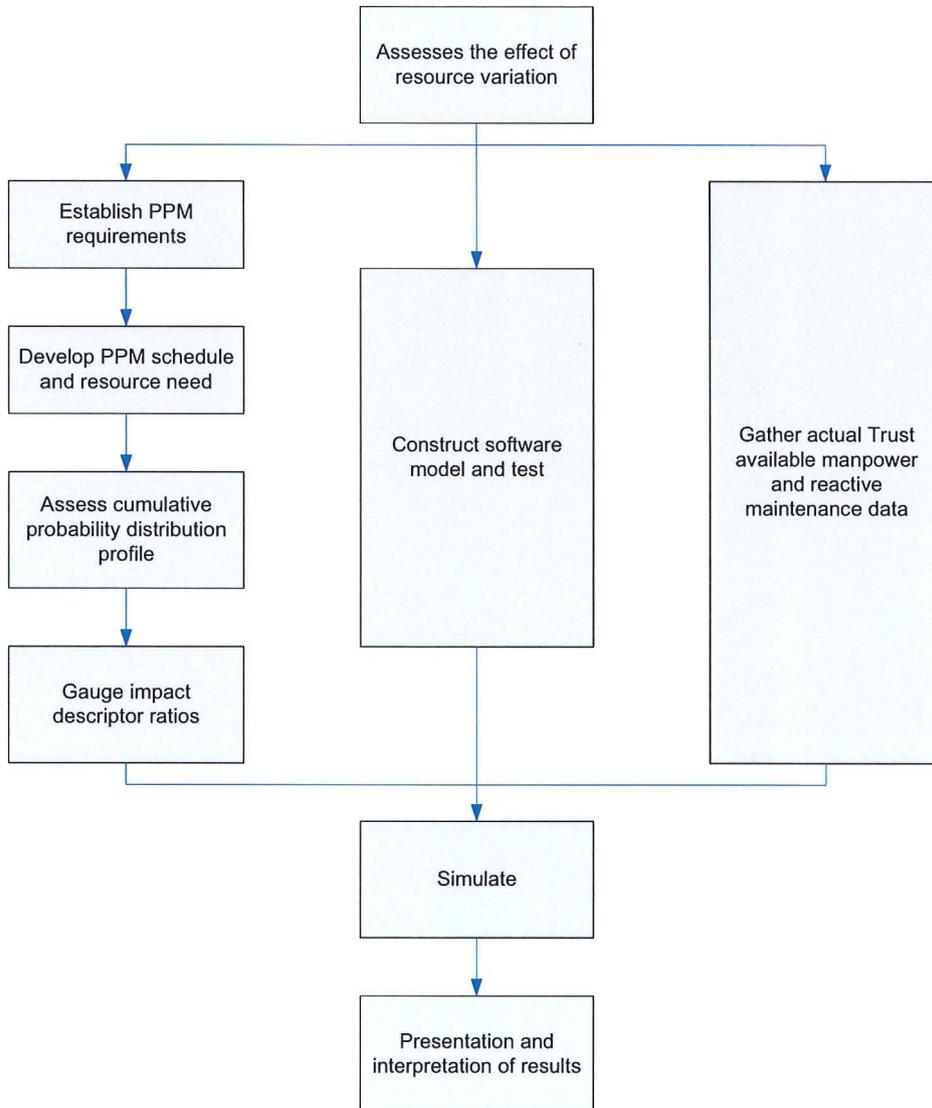
Reading figure 9 from left to right, three pieces of work were begun in parallel. The first established the planned preventative maintenance requirements of the organisation from reference to the relevant national codes of practice and any local periodic maintenance needs. These tasks were then arranged into a yearly schedule alongside an assessment of their relevant resource needs and cumulative probability distribution profile. This was followed with a judgement of the impact that would befall the organisation should the asset actually fail. It was from the multiplication of the probability and impact figures that risk was computed for each asset on a week-by-week basis.

Secondly the software was constructed that would simulate the variation in resource availability. Before the simulation was undertaken the program was subjected to various tests to determine its accuracy and robustness. The third part of the project began to gather the Trust's actual data in relation to the available manpower and the reactive maintenance completed on a week-by-week basis.

Once these three areas had been completed the actual data was simulated. The output was then presented together with an interpretation of the results. Throughout the procedure extensive use was made of focus groups comprising of the organisations estates and other senior Trust specialists (e.g. Control of Infection, Chief Nurse, etc). The following paragraphs expand this brief overview to describe the process in detail as shown within table 8.

Table 8   Overview of the simulation process

Process Topic
Focus groups
Establish planned preventative maintenance requirements
Develop planned preventative maintenance schedule and resource need
Assess cumulative probability distribution profile
Gauge impact descriptor ratios
Construct software model and test

a.   Focus groups

As all of the data generated in relation to the probability of an asset failing and its relative impact on the organisation, staff, patient, stakeholders, etc, were based on qualitative as opposed to quantitative data, Trust specific focus groups were again employed. The participants within these focus groups were specific to each Trust participating within this research, and consisted of estates senior managers and other Trust specialist advisors e.g. Control of Infection managers, Chief Nurse, Risk Manager and Internal Auditor. The inherent problems surrounding the reliability of these groups to produce consistent and robust outcomes have been discussed and addressed previously, within this methodology chapter when describing the development of risk profiles..

As in the case for the development of the risk profiles, members of the focus groups were specifically selected for their expertise within the area under discussion. Prior to any workshop undertaken, a full and frank explanation was given of the aims of the

project, methodologies, timescales and probable outcomes together with all of the assumptions made throughout the whole process. Specifically explained was the fact that the expert elicitation technique employed was a heuristic tool NOT a scientific tool, that was being used, in the words of the US Army this process is used, "to explore vague and unknown issues that were otherwise inaccessible" (U.S. Army Corps of Engineers, 2000).

b.   Establish planned preventative maintenance requirements

Reference was made to the various technical standards (which included manufactures recommendations) and local maintenance needs, national NHS (and other) codes of practice, together with the Trust's Estate and Clinical strategies. This was undertaken by the Trust's local estate managers and officers with reference to the Trust's specialist advisors such as Fire, Control of Infection officers, etc. These basic actions began to form an audit trail from tasks undertaken on the shop floor up to statutory requirements, cross linking to the strategic directions and objectives of the organisation.

c.   Develop planned preventative maintenance schedule and resource need

For each Trust site every asset (plant and/or equipment) was individually assessed, and the resources needed to undertake each individual planned preventative maintenance task in terms of man-hours and frequency of inspection were agreed by the focus group from reference to the national and local planned preventative maintenance needs. Then the planned preventative maintenance schedule was designed to produce an even workload (as far as possible) throughout the year.

d.   Assess cumulative probability distribution profile

The NHS 5x5 Criticality grid employs the use of numbers (1 to 5) to depict the various levels of probability and impact. In the NHS system these are multiplied together to give a value of risk.   As the simulation program requires mathematical manipulation to determine the variation of probable failure rates with alterations in resource availability then this grid has been rejected for the following reason. There are four types of measurement; the level of measurement dictates the arithmetic and statistical operations that can be undertaken on the figures (Wharrad, 2004). These four types are: Nominal, Ordinal, Interval and Ratio.

117

The NHS 5x5 Criticality grid has been constructed using un-calibrated ordinal scales for both the probability and impact axis, thus are unable to be mathematically processed (Department of Health, 2004). Conrow (as discussed earlier within the Literature Review chapter) makes great criticism of these types of scales as they lead much misinterpretation and confusion (Conrow, 2003). This stance is supported by the work of Stamper who states that, 'Failures of organisational behaviour stem, not, infrequently, from a failure to use measurements correctly' (Stamper 2001). Therefore the formation of Utility Functions to determine the probability and impact scales was adopted. Keeney and Raiffa have described in detail within their much cited work the fundamental development and mathematical use of utility functions (Keeney & Raiffa, 1976).

e.   Development of the probability utility functions

In this case the failure distribution of components must be determined in order to create the probability distributions. The maintenance engineers were regarded as the 'experts', and their judgement was sought relating to the remaining lifespan of the equipment before it began operating beyond its design tolerance parameters, should the planned preventative maintenance tasks not be undertaken. In other words how long would it be before the equipment failed? To this end they were asked to respond to the same question posed by Øien when undertaking his research into the optimisation of maintenance models:

> 'Assuming that you had not carried out this planned preventative maintenance task, how long do you think the remaining lifelength of the component would have been?'

Øien 1998

By 'component' Øien means the piece of equipment, and 'lifelength' is the remaining time frame before the equipment was deemed to fail. As Øien points out, 'this is beneficial as the answer to the question may be given in years, i.e. the term probability is not involved' (Øien, 1998).

Noortwijk, Dekker, Cooke and Mazzuchi (1992) also employed expert elicitation techniques in their research to obtain lifetime distributions needed for maintenance optimisation, due to the scarcity of reliable data. As in this case, the maintenance

118

engineers that they worked with had little or no knowledge or experience with statistical failure models and only had a limited time to spend during the elicitation process. Due to time constraints and the powers of concentration of the experts, they had in one elicitation session to gather information relating to several component lifetime distributions, effects of planned preventative maintenance, consequences of failure and any other relevant information. For these reasons the use of a well-known failure model was rejected as experts would be asked (in their words), 'to fit a parametric distribution to their subjective probabilities is a lot to ask for even from statistically trained experts'; again as in this case. They go on to claim that:

> From a practical point of view, it is more comprehendible for statistically untrained experts to use a discretised version of the continuous probability distribution function (viz, a histogram) thus replacing the concept of the probability of failure in a fixed time interval'
>
> Noortwijk, Dekker, Cooke & Mazzuchi, 1992

Copying their research technique, this research also employed the use of a simple histogram asking for failure probabilities from only two different time intervals. As in their case this research found that this approach found favour with the experts. To this end the probability function descriptions (continuous scale) and their rank order from the NHS 5x5 Criticality grid were used by the focus groups as an aid to assess the probability of failure over time should each asset's associated planned preventative maintenance task not be completed. Table 9 reproduces that part of the grid that was used for this purpose.

119

Table 9   **The ranking and descriptions for the probability scale from the NHS 5x5 Criticality grid**

	Probability of Failure Rating				
Rating	**1**	**2**	**3**	**4**	**5**
Failure descriptions	**Rare**	**Unlikely**	**Possible**	**Likely**	**Certain**
	Non or minimal remedial action required and/or new/recent upgrade. Estimated time to failure may be circa > 10 yrs	Normal wear and tear. Sound, operationally safe and exhibits only minor deterioration. Estimated time to failure may be circa<10yrs	Reasonable physical damage/ deterioration. Reassignment of life may be acceptable based on technical tests or residual robustness. Estimated time to failure may be circa < 5 yrs	Major physical damage/ deterioration. Failure apparent/asses sed as imminent or unacceptable built environment. Not appropriate to reassign life. Estimated time to failure may be circa < one yr	Failure occurred. Unacceptabl e built environment. Not appropriate to reassign life. Estimated time to failure may be circa < six months
	Fire/ statutory Complies with mandatory fire safety requirements and statutory safety legislation	**Fire/ statutory** Complies with mandatory fire safety requirements and statutory safety legislation with minor deviations of a non-serious nature	**Fire/ statutory** Known contravention of one or more requirements - which falls short of - "B".	**Fire/ statutory** Dangerously below "B".	**Fire/ statutory** Dangerously below "B".

The methodology employed to assess this increase in the probability of failure is best illustrated by the example in figure 10 below which is a cumulative probability distribution profile for a fictional asset.

120

Figure 10 *The perceived increase in the probability of failure of an asset as the number of planned preventative maintenance tasks not completed increases.*
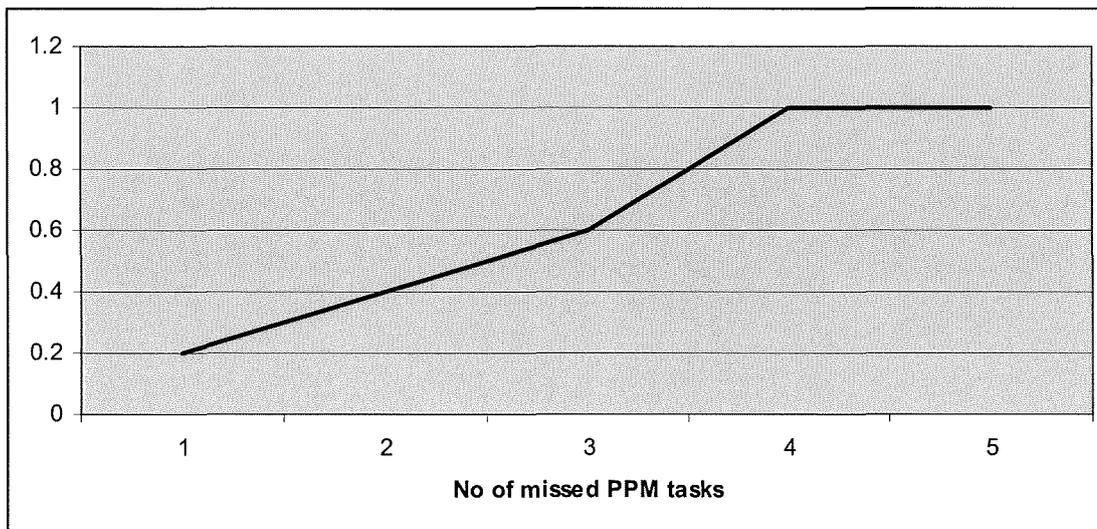

Figure 10 shows that if a relative planned preventative maintenance task was completed at the scheduled frequency (1) then the remaining risk level would remain at the accepted residual risk. In the example the focus group had estimated that if the first scheduled planned preventative maintenance task was not completed at the scheduled frequency, then the risk would rise from Rare (rating 1) up-to Unlikely (rating 2) from reference to the probability scale in table 9 above. If the relevant planned preventative maintenance tasks continue not to be completed the focus group estimated that the probability of failure would continue to rise from Unlikely (rating 2) on to Likely (rating 4) and level out at Certain (rating 5) for both the 4th and 5th time that the planned preventative maintenance task remained uncompleted. Thus the graph in figure 10 is showing the focused group assessed cumulative probability distribution profile for the asset in the example should the relevant scheduled planned preventative maintenance task not be completed.

As time progresses the probability of failure can only increase (not decrease), however where the group considered that the increase was extremely negligible they were allowed to assume a flat line increase between points. Also it was agreed within the group that the simulation should only need to consider five missed planned preventative maintenance tasks, as they considered that after this any probability estimation would level to a flat line. These rules allowed the generation of 72 possible cumulative

121

probability distribution profile shapes to be available following the five failure descriptions within table 9. The descriptions within table 9 were only used as an aid for the focus group to determine five points on the graph. The intervening positions were determined using interpolation between the points. This system for calculating the probability value during the computer simulation does give rise to an error when using the interpolation process against what can be assumed to be the 'true value' that would form a curve between the set point values. Therefore it was necessary to estimate the size of this error and to determine its significance once the results of the simulation run was completed. As with the development of the probability utility functions the design of the impact utility functions made reference to the NHS 5x5 Criticality grid as an aid only, but in this case the Analytic Hierarchy Process was adopted to create the impact utility function (Saaty, 2001). There are various methods available to conduct a multi-criteria analysis for the comparison of descriptors such as those employed within the NHS grid for potential consequences, but as Emblemsvåg and Kjølstad state:

> 'The greatest advantage of the Analytic Hierarchy Process concept, for our purpose, is that it incorporates a logic consistency check of the answers provided by the various participants in the process'.
>
> (Emblemsvåg & Kjølstad, 2006).

Table 10 below shows the design of the NHS grid relating to impact.

Table 10   NHS 5x5 Criticality grid relating to impact

| | Severity | | |
	Rating	Descriptor	Health & Safety
Potential Consequences	1	Insignificant	No injury/breach of guidance/procedures
	2	Minor	Minor injury/ill health (first aid or self treatment) reach of legal requirement.
	3	Moderate	Moderate injury/ ill health statutory obligations, improvement notice issued.
	4	Major	Major/significant injury or long-term incapacity/disablement. Prohibition notice issued.
	5	Catastrophic	Fatality and /or permanent incapacity/ disability. Prosecution.

As the grid descriptors were pre-arranged in rank order by the NHS system, the design of the hierarchal structure was effectively pre-designed, as shown below in figure 11.

Figure 11   The design of the Analytic Hierarchy Process structure
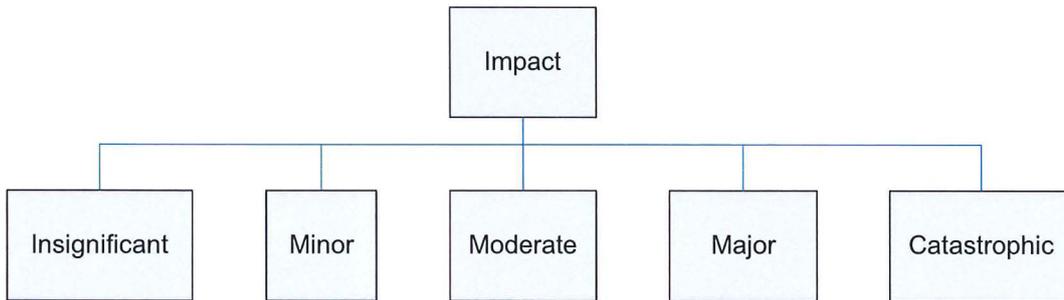
During the development of this part of the methodology four people separately used the Analytic Hierarchy Process system to create the ratio differences between the five impact severities via commercial decision support software Logical Decisions ®. The results are shown within figure 12 below.

Figure 12   Results from the trial run to determine the impact ratios using the Analytic Hierarchy Process method
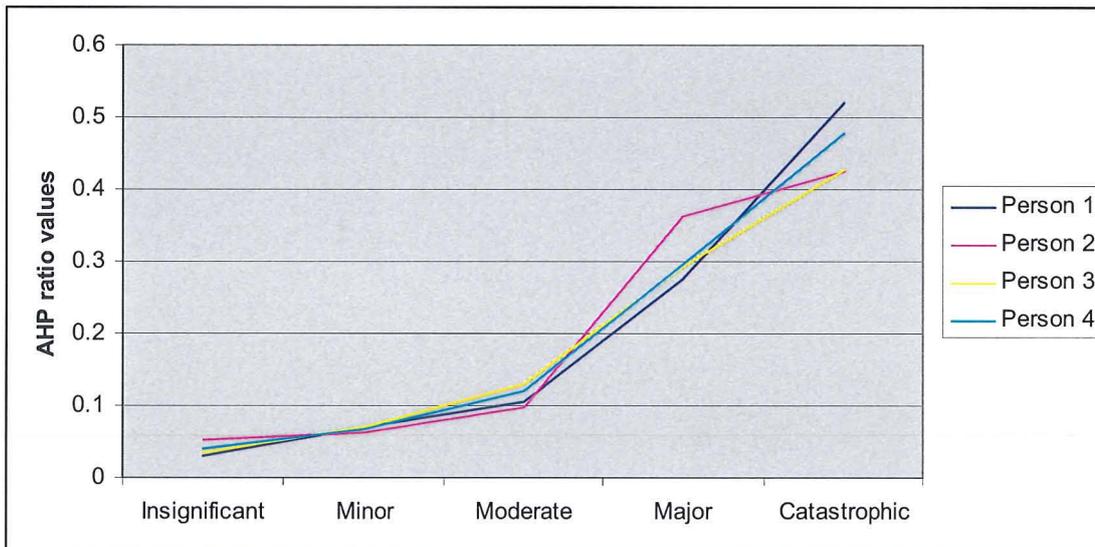
The closeness between the results can be easily observed and is verified from table 11 below.

Table 11   ***Ratios relating to the five descriptions of impact***

	Person 1	Person 2	Person 3	Person 4
Insignificant	0.031	0.052	0.036	0.039
Minor	0.069	0.062	0.069	0.068
Moderate	0.104	0.097	0.13	0.119
Major	0.274	0.363	0.293	0.296
Catastrophic	0.521	0.426	0.427	0.478
CR value	0.087	0.056	0.04	0.017

The consistency ratio (CR) within table 11 forms an integral part of the Analytic Hierarchy Process. It assesses the uniformity of the decisions made when conducting the pair-wise comparisons within the hierarchal structure. It is recommended that for this number of descriptors of impact that a consistency ratio of less than 0.1 is desirable to give acceptable results (Saaty, 2001).

The researcher suspects that the closeness-of-fit of the observed and calculated values within figure 12 and table 11 were formed due to the restrictions of the descriptor values being pre-defined and pre-ranked in order; thus leaving little room for excessive individual interpretation from the participants.

Although the Analytic Hierarchy Process ratios can be developed within a focus group setting (Saaty, 2001) the researcher decided not to follow this form for the assessment of these figures based on the results shown within figure 12 and table 11 above. The reason given for rejecting the group approach was due to the lack of detailed prior knowledge of the complex Analytic Hierarchy Process by the group members coupled with the already multifaceted and lengthy process of the methodology employed to develop the simulation and interpretation of the outputs.

f   Construct software model and test

The software was designed to simulate the effects of various resource availabilities against the Trust's actual historic data, for the directly employed Trust employees (as opposed to contract labour) manpower availability and the maintenance tasks that were needed to be completed on a week-by-week basis. These maintenance tasks were composed of two maintenance types, reactive and the planned preventative maintenance system. Figure 13 below shows the salient parts of the software design.

Although the diagram is relatively self-evident eight essential points and assumptions to note are given below, and were agreed with the focus groups

i Reactive maintenance is always completed before any planned preventative maintenance task, therefore the man-hours required for the weeks reactive maintenance is subtracted from the week's available man-hours. Any remaining man-hours for the week are then directed towards completing the planned preventative maintenance tasks that are generating the greatest risk, thus a perfect work scheduling programme is always assumed.

ii There was always sufficient manpower per week to complete that week's reactive maintenance demand.

iii Available manpower, reactive and planned preventative maintenance figures never assume negative values.

iv If any available man-hours were surplus for a given week after all of the weeks reactive and planned preventative maintenance tasks were completed, then the planned preventative maintenance scheduled tasks for the following week are brought forward. Thus the program is simulating a fully occupied workforce always having their resources directed towards where they would be doing the most good, reducing risk to its lowest possible level.

v All of the algorithms used within the software to model the variation in resource and calculate the resulting value of risk were void of 'add-factoring'. This is the practice of using a fudge factor to the output of a model so that it reflects the modeller's intuition.

**vi** No multi-year frequencies were simulated within the model. These frequencies relate to planned preventative maintenance inspections that are undertaken say every 2 years, 5 years, etc. The justification for this limitation was to greatly simplify the design of the software, and from the knowledge that these multi-year frequencies only account for a very small percentage of the total workload requirement for a Trust.

vii It is natural for a small number of maintenance tasks (either reactive and/or planned preventative maintenance) not to be totally completed at the end of each week but to be 'carried over' to be completed the following week. Therefore the software was designed to mimic this action.

The software was designed and written by the researcher to ensure that on a week-by-week basis all available resources were targeted to completing the reactive maintenance tasks first; any remaining manpower was then used to complete the scheduled and outstanding planned preventative maintenance work. These planned preventative maintenance tasks were then completed in priority order i.e. those tasks that were generating the greatest risks were completed first. Any planned preventative maintenance tasks not completed for that week were moved forward into the next week's stock of work, recalculated in terms of their risk generation and prioritised alongside the other planned preventative maintenance tasks scheduled for the week.

Where the resource availability was in excess of the total reactive and planned preventative maintenance needs for the week, the software brought forward the next weeks planned preventative maintenance tasks, prioritised them and assigned resources for their completion. In this way the software on a week-by-week basis was constantly assigning resources to where they would do most good i.e. completing all tasks in priority order, thus keeping the generated risk to an absolute minimum with the manpower available.

Figure 13   The salient parts of the software design.

127
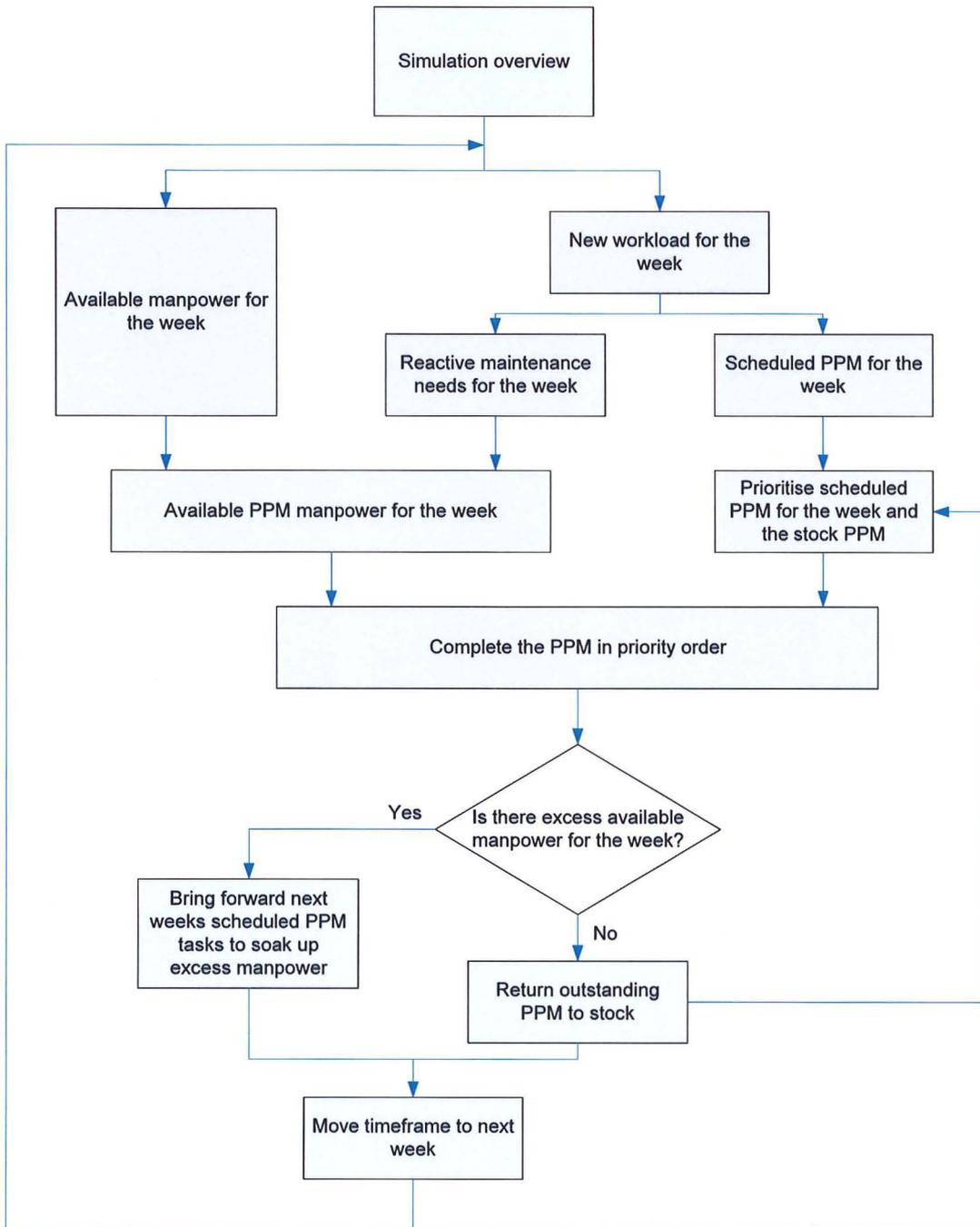
3.4.1 Simulations using actual Trust data

At this part of the project confidence and experience had been gained, from the simple estates model and the academic rigor relating to the estimated cumulative probability distribution profile and impact assessment figures. Thus the decision was taken by the researcher to load the simulation model with 'actual data' from a representative Trust. The simulation using 'actual data' was programmed to replicate 20 years of operation to allow the large data mass to equilibrate. Also at the start of each simulation the program assumed that previously all reactive and scheduled planned preventative maintenance tasks had been completed, thus a common starting point was established at the residual risk level.

3.4.2 Methodology Overview

For each of the Trusts involved, the exercise to simulate the artisan workload was split into distinct parts, as shown within Figure 14 below.

Figure 14   *The steps of the simulation process*

128
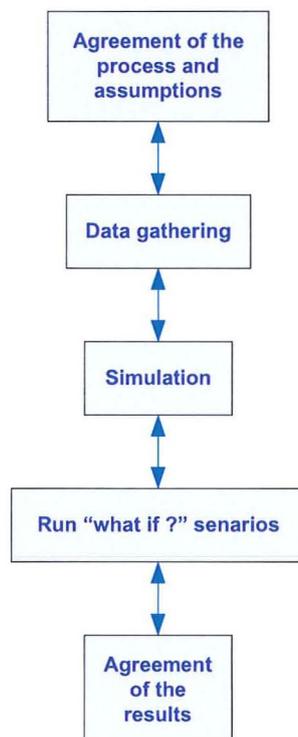
The initial arrangement started with a number of group presentations to the estate departments officer's to describe the process, timeframe, and probable output; together with the need for involvement of other Trust specialist managers for professional and clinical advice. During these presentations it was agreed that a series of initial simulations would be completed first, before any senior managers external to the department were actively involved within the process. This was in recognition that the simulation would involve a great deal of complex and detailed technical data, which would require much discussion before any agreement could be reached between the estates officers. And until this was completed it would not be effective to involve others. Following these initial discussions, the Trust estates operational manager gave an overview of the exercise to the senior managers external to the estates department with an indication of their expected involvement, timeframes, etc.

Following the presentations, and agreement about the process and timescale by all concerned, the task of gathering the base data began. This could only be undertaken by the local estates officers as it required access to and interrogation of the Trust's computer systems. The data gathered and agreed from this process formed the input into the simulation model. A natural development following the initial simulations was the creation of 'what-if' scenarios. These were formed from the focus groups questioning the simulation outputs, and were a direct reflection of their increased understanding of the maintenance processes.

Following agreement by the various focus groups on the interpretation of the simulation outputs, the results and recommendations generated were presented to the respective most senior managers of the Trusts. This gave them the understanding necessary to enable them to make policy decisions, knowing the probable effects in terms of the risk that would most likely be created.

Prior to analysing the data it was necessary to test the software to ensure that the software program would accurately perform to its design intent. This had been done initially on the simple estates model used to access the applicability of the Rudolph and Repenning work when applied to an estates setting. However since then the model had been significantly enhanced. This final testing procedure consisted of running the

software under the following five extreme conditions (reality checks) and observing the output against known values as shown within table 12 below:

Following the successful completion of these tests the 'actual data' was simulated and the resulting variation in risk was recorded and is discussed within the results chapter of this thesis. The output data was subjected to three separate tests to determine when a 'steady state' had been reached; these are fully explained within the Results Chapter. Following these tests the interpretation of the results was restricted to the final 500 weeks of the simulation, as it was considered that by this time the program would have settled and that this part of the output was a reasonable reflection of the situation that would occur in practice.

Table 12   The software extreme test conditions against the anticipated and observed outputs

Extreme Condition	Anticipated and observed output
Zero manpower input.	The total risk for each week rapidly rose and levelled out for each successive year.
Zero manpower input with an extremely large manpower step function spaced at irregular intervals.	As above, but within the week of the step function all tasks were completed and the risk reduced to the total resultant value. Following each successive step function the risk began to rapidly rise until the next step function encounter.
Extremely large manpower input.	The total risk for each week year-on-year was reduced to the residual value.
Extremely large manpower input with zero manpower level step function spaced at irregular intervals.	As above, but within the week of each step function the risk rose, followed the next week with a drop of risk to the residual level.
Confirmation of the software programmed algorithms against hand calculations for the rise in risk for a given asset, should the available manpower be unavailable and the planned preventative maintenance tasks were not completed for successive periods.	The software programmed algorithm was confirmed by the hand calculations and recorded the rise of risk until the risk levelled at its maximum value when the planned preventative maintenance task had not been completed after the fifth time.

There have been eight Trusts participating within this third research objective. Each focus group employed between five and seven engineering officers with support from their specialist advisors, totalling between nine and eleven experts per Trust. Thus

130

between 72 and 88 experts in total were involved. The trade groups simulated varied depending upon the size of a Trust estate's workforce. In total 47 trade groups have been simulated.

The next section describes the fourth (and last) research objective which is to:

> Develop a robust, transparent and repeatable methodology to assess the directly employed artisan revenue resources needed to ensure that the maintenance activities conform to the National Health Service standards.

3.5 Development of the fourth research instrument

A major part of any managerial system is the need to regularly monitor the organisation to give assurance that the local managerial systems that have been specifically designed to conform to the various national standards were actually being practiced (as discussed and referenced within chapter 2). Therefore an assurance framework must be implemented as part of the day-to-day managerial system process as opposed to the reliance on a periodic (normally 12 month) audit. This next section describes the design of such a system.

The managerial procedures operating within an NHS estates department are primarily safety systems, of which the monitoring routine is a major component. Ericson states that the: "goal of system safety is to find out what can go wrong (before it does) and establish controls to prevent it or reduce the probability of occurrence" (Ericson 2005). In the ideal world the monitoring technique should be designed in conjunction with the physical infrastructure in order that they can both be influenced by the other, rather than trying to retrofit or enforce design changes in either system afterwards.

Unlike the majority of engineering solutions that once designed, installed and switched on, over effectively long periods of time, managerial issues cannot be simply left to look after themselves; they must be watched, fussed over, tweaked, and adjusted on a regular basis. Thus the monitoring system must be regarded as a dynamic structure, constantly under review and subject to change. As external forces alter the procedures must respond and adapt. For this to be achieved the managers designing the monitoring regime must understand hazard theory, the techniques for hazard identification and risk

131

elimination and/or reduction. This understanding together with knowledge of how their engineering systems are designed, acceptable tolerance limits of performance, detailed operation of their engineering infrastructure under normal and fault conditions and how each sub-system will interact and the effects of failure on the Trust, staff, patients, stakeholders, etc. are needed to design a robust and effective monitoring scheme.

3.5.1 Hazard Analysis Techniques

There are a number of methods used to undertake a hazard analysis, and no single technique will necessarily identify all of the hazards within an organisation's engineering systems. Therefore different techniques may need to be employed throughout the varying depth of an analysis. Ericson states that there are over 100 such techniques in use, obviously some more general in nature than others and most minor variations of others The majority of these employ both qualitative and quantitative data within their methodology. Within a previous chapter it was stated that the fully quantitative data was not available, but in a risk analysis (although desirable) precise numerical accuracy is not always necessary (Ericson, 2005).

3.5.2 What to monitor

The basis for this section can be traced back to Taylor who coined the phrase 'Scientific Management' (Taylor, 1911). He improved productivity by breaking a process down into small elements, analyzed how each step contributed to the overall task (goal), then by redesigning the process steps reduced the time (cost). His work spawned much research by others over the years; one of these areas became to be known as Operational Performance Management. There has been a resurgence of interest in this field lately within the NHS through their Payment by Results initiative. In this the payment mechanisms have completely changed in relation to how a Trust receives funding. In the past the monies allocated to Trusts were based on historical formulas based on population served, geographic location, etc. This has now altered to a system where the funding for a particular patient follows the patient, irrespective of where the patient receives their treatment, or what treatment they are actually given. Therefore, it is imperative that effective means are found for Trusts to pay close attention to current and projected income cash flow movements (Dr Foster Research Team, 2006). Frameworks need to be developed so that Trust boards can identify

132

their information requirements. These include the need to ensure that they spend more time on strategic matters by using their information more efficiently and effectively. This in turn means that they must take a balanced focus when scrutinizing current, recent and projected operational performance. This requires that the information is presented in a format that enables Boards to focus on a particular area of service provision, understand the patterns of need, the current service response to that need, variations in performance and the risk factors.

3.5.3 The operational performance reporting system

Five criteria were outlined within the Literature Preview Chapter that created a framework around which the monitoring system had to be designed. In summary the criteria required that the system must conform to international and national standards, reference mainstream academic theory, employ existing hierarchal staffing structures, job descriptions and software, be applicable to a multi-professional team, and be revenue neutral against existing budgets.

The methodology presented for this final part conforms to all of the five criteria outlined above. This is achieved by the design of a 'driver' to push a common 'generic framework', which employs standard operational performance management techniques. The output of the process is a quantitative assessment of the department's efficiency, quality and effectiveness towards supporting the Trust in achieving its objective goals. This is presented as a series of performance index profiles against a base period, showing the actual achievement against a target value.

As such, reference was taken of the work by Johnston (2002) and his research into such performance systems. He criticises organisations for creating in his words, "a cottage industry" from the production of performance measurement data, to feed the plethora of frameworks and tools that have been developed over the years, citing a number of the most dominant systems. His reasoning arises from the effort and cost of producing this data compared to the usefulness of the outputs, and goes on to discuss the worthy work of academics who over the years have attempted to understand and quantify the relationships between the performance measures, by creating casual models and computing the correlations between elements, but seriously questions the results (Johnston, Brignall & Fitzgerald; 2002). It was from this background that the monitoring

133

system was developed and designed. At all times simplicity was paramount, as it was obvious that it was only when the most basic of monitoring information was available in a relevant, robust and transparent manner, could the more sophisticated and complex performance models be considered in future years. The researcher thus created a series of generic key steps to follow when designing an outline project plan and shown in Figure15. As each Trust and estates department differs in their organisational structure, reporting arrangements and managerial systems, only a generic approach can be presented within this methodology chapter as each project must be tailored to cater for each situation.

Figure 15   *The generic key steps to be followed shown as an outline project plan*

Development and design

- ❖ Initial survey
- ❖ Detailed identification of the process steps
- ❖ Define the process
- ❖ Analysis
- ❖ Design of the efficiency measures
- ❖ Measuring the efficiency of the process
  - o Aggregation of multiple operations
  - o Non-productive time
  - o Overheads
- ❖ Input specification
  - o Units of labour
  - o Salary cost
  - o Operating maintenance costs
  - o Capital costs
- ❖ Measuring the quality of the process
  - o Reactive maintenance
  - o Planned Preventative Maintenance
  - o The management function
- ❖ Measuring the effectiveness of a process
- ❖ The reporting system
- ❖ Structure of the monitoring process

o The driver

o The assurance framework Operational Risk

The implementation plan

❖ The implementation team

Maintenance

The outline plan above is split into three distinct headings. These are the development and design, implementation and maintenance phases. But due to the extreme importance attached to the monitoring system for the safe delivery of patient care, it is essential that the introduction of the system is supported from the Trust board throughout the three phases. This commitment should be communicated to all concerned through briefing notes, announcements and updates regarding progress. As the monitoring system is a managerial tool it must also be supported throughout the managerial structure. This can only be achieved if the respective managers are satisfied that the measures selected provide them with relevant and timely information, in a format that they can digest and use to assist in achieving the maximum performance of their respective departments.

The following paragraphs expand and explain each step shown within the outline plan, starting with the development and design phase which splits into various sub-sections.

3.5.3   Development of the research instrument

As the output from this fourth research objective is to develop a robust, transparent and repeatable methodology to monitor the day-to-day activities of the maintenance department in line with the requirements of the National Health Service standards, then only a 'generic' methodology can be created as a research instrument. This is due to each organisation, department and processes differing in their structure and procedures.

The development and design of the monitoring system requires the interaction and co-ordination of a diverse range of groups in a way that cannot normally be accommodated by existing organisational structures. Therefore it would be necessary to form a 'steering committee' chaired by a Trust board member. The task of the steering committee would be to direct, co-ordinate and evaluate the activities of the project team, commit resources and make definitive decisions. However a prime task for the steering

committee is to ensure the relevance and technical integrity of the monitoring system and its integration as a supporting mechanism for the whole of the Trust's information systems and decision making process. The committee in turn must be supported by a project team, who under the terms of reference dictated by the steering committee, must develop and design the monitoring system.

The first steps that the project team would take would be to instigate an initial survey to determine the likely scope that the monitoring systems would cover within the particular department under investigation. This can effectively be achieved by forming a focus group containing a cross section of the department's managers and operatives (i.e. experts in the department's processes, as discussed at the beginning of this chapter and successfully employed within each of first three research instruments. However this is presented as only one probable methodology, when in the 'field' the research instrument must be designed to match the prevailing environment and constraints.

3.5.4 Initial survey

This would consist of a review of the existing organisation charts, operational plans, existing estimates and forecasts, management reports, and selected interviews. From this the project team would create an overview of the existing inputs and outputs of the department, with a comparison against the national standards and local needs of the Trust. Put simply this work would begin to answer the 'where are we now?' question. Following this an initial project plan should be expanded to include the first area to be selected for monitoring. This extended plan should include milestones, target dates and resource estimates, together with an overview of the next steps. Codling (1998) states: "Experience has proved repeatedly however, that the quality of the result achieved is directly related to the effort invested at the outset". In other words she is recommending that by putting a great deal of effort into this the planning, the overall process would be easier (Codling, 1998).

Following approval from the steering committee to the now extended plan, it would be necessary to determine the critical success factors of the organisation. This would necessitate the involvement and support from the most senior managers, to ensure that efforts were directed to areas of strategic importance rather than to those that were 'easy'. Two questions must then be asked of the key operational processes. These are,

136

'"What is it?" and "How do they do it?" These answers would give a picture of how an organisation's systems 'actually' worked, as opposed to the actions outlined within a department's policies and procedures. Only through a detailed understanding of the in-house procedure would it be possible to understand how it functioned, and then to form an appreciation of the key steps that would be needed to create its monitoring process. This understanding would be developed from a structured approach analysing the process steps as opposed to the department's outputs. During this part it would be extremely important to comprehend the interdependence with other departments/processes within the organisation, due to the complex and tightly coupled interactions that exist.

From this, the project should begin to develop a very detailed picture of the department's links to other departments, work procedures, origins of work, work flows, factors affecting workload and the end use of the outputs produced. The quality, quantity and availability of the input data, the components and actors that form the outputs and the quality of the outputs, would have to be identified and quantified. This process would create an appreciation of the key steps that would need to be monitored.

The tolerance limits for each of these initial 'key steps' would need to be established from reference to the relevant national and local standards, and the methodology for data collection designed. The gap between the acceptable levels of output and that actually achieved, must then be formed, followed by the setting of future targets. The completion of this process would not be a set of results; rather it would be a measurement of a set of quantitative or qualitative tasks.

From this an outline action plan for the design of the operational performance system as a whole would be formed. This would contain estimates relating to milestones, target dates and resource needs. At this part the action plan would be very basic but would form a template that in time would be expanded in detail as the project developed. Following the approval of the steering committee, the project team then would begin to identify in detail the department's process steps.

3.5.5 Detailed identification of the process steps

The chosen areas to investigate would be stripped down to their base components so that each step in the processes could be analysed in turn. This would identify the critical inputs and outputs from each step, and the position of the weak points in the system. It would be this attention to detail that would be essential to successfully identifying the critical points to monitor. This in turn would guide the design of the parallel defensive systems to give assurance that the step in question would be robust and relatively free from failure. One of the outputs from this work would be a diagram that resembled an organisational chart, which would show the basic steps in sequence that formed the whole process. The development of this diagram would be crucial to the whole exercise, as it would assist the team (and others) in learning how the procedures work, by identifying their component parts, interrelationships and actors involved in the total system. But in order to analyse such a process it would be necessary to accurately define it, identify its boundaries, and discuss each step in turn.

3.5.6 Define the process

Different people and/or professions sometimes describe, name and define things in different ways. As such, confusion and misinterpretation can cause problems unless they are correctly defined to everyone's agreement at the beginning of the analysis. To do this requires statements that answer questions such as: what is the actual and desired output of the process?, who are the actors within each step?, who are the clients for each output and what are their requirements?, does the process deliver what is required?, what are the start and end points?, and who is responsible for the process? These 'boundary' questions must be addressed as comprehensively as possible as these could identify the root cause of problems or dysfunctions. The more steps and interactions there are within a process, the greater the potential for problems to be created. Where the process networks with other departments they must be consulted to fully agree the boundaries. This in turn would facilitate in gaining ownership and responsibility for the whole process.

Following the definition of the process and its boundaries it would be essential to re-list and re-name each single step in sequence. It is at this point that the 'detailed' flow diagram would be developed to create a pictorial view of what was actually happening. This could then be converted to a process flow diagram. This would be the final phase in

defining the overall process that was being analysed. The map would give a pictorial overview that would show the interfaces and interconnections with other departments (either internal or external) to the Trust.

The discipline involved in producing the map creates numerous spin-offs. For example, it facilitates the analysts to point to the potential steps that could lead to failure, and gives guidance towards those steps that would require close monitoring, to detect any deviation in output from acceptable levels.

As Codling (1998) remarks; "Time and again the process map proves how few people really understand the process". A bonus that this methodology naturally gives is the ability to critically examine the system in terms of streamlining the process by: reviewing the skills of the operatives necessary to complete each step, removing duplicate steps, reducing any inherent delays and exploring the introduction of technology to reform and reduce the resource needs to deliver the same output. The completion of this section will go a long way to answering the question, 'where are they?' in terms of what people are doing and how they are doing it. From this base information the actual design of the main body of the operational performance measurement system would begin by an analysis of a comparison of actions 'at-the-coal-face', against the organisations policies and procedures.

3.5.7 Analysis

This whole process involves interviews with selected line managers as a major part of the data gathering exercise. This must include discussions on such topics as: the organisation and structure; the staff roles and responsibilities; the work procedures from the origin of the workload, workflows and factors effecting it; the end-use of the outputs; the department's objectives and personal objectives, and the links into the Trust's objectives; the line managers views relating to the outputs and the individual components making up the outputs; the availability and quality of the data sources, and the ability to quantify the data; the applicability and use of any weighting system; and finally the use of the existing information systems.

The previous phase in this planning section identified the data sources and the most appropriate methods of collection, as attention was directed at the areas most likely to

create suitable (useful) information. In order not to create 'information overload', it would now be necessary to craft another framework, but this time one that considers the answers to the following questions:

- ❖ What is the objective of collecting this particular type of information?
- ❖ What do staff need to look for and why?
- ❖ How accurate must the data be?
- ❖ How much information do they need?
- ❖ How much resource can they allocate to the collection of this data in terms of time, staff, and budget?

The data collection plan above must also show the source of the data to be collected, be it internal or external to the Trust, who is to undertake the collection, and how. The data considered in this part would consist of both input and output data. It would be during this phase that suggestions would be made relating to possible alternative outputs, and consideration of the pros and cons of each suggestion argued. It would be important to document the output of these discussions for future reference. It must not be the case that the subject to be monitored was someone's 'pet-subject' or this month's 'flavour-of-the-moment', but should be one of strategic importance to the Trust; where failure in this area would significantly affect the Trust in achieving its objectives in delivering patient care. As the department is a part of a highly complex interactive and tightly coupled organisation, it would be necessary to engage with other departments in order to form an opinion of which subject areas to choose form an output and to monitor. At the operational level the outputs should be large packages of work performed, e.g. the completion of reactive maintenance tasks/week. These outputs should be representative of the work performed, quantifiable, repetitive and reasonably uniform over time, mutually exclusive and affect the Trust's objectives. Also the outputs must be created within a given time period so that problems of measuring work-in-progress are avoided, and the outputs must also be reasonably predictable in terms of forecast trends.

Outputs could be combined to form a single measure. Where there was a constant output mix ratio such as the output/m^2 for a number of sites within a Trust, then the volumes of each output could be added to give a single output figure. Where there was a variable output mix, then a series of output weights could be derived, where the various

output volumes could be multiplied by their respective weights to arrive at an overall final output figure.

Irrelevant and 'nice-to-know' information must be discarded at the earliest opportunity. The precise date of collection would be important as vague timelines such as mid-month would be imprecise and open to misinterpretation and abuse.

Fundamental to any gap analysis would be the knowledge of what should be done by the department when compared against the national and local standards. Managerial systems that did not conform to the recognised codes of practice could readily be identified by the use of a very much simplified 'V' diagram. This could be used as a framework to aid the auditing and checking of the system under scrutiny. This type of diagram is used within the software industry, by creating a structure to describe the methods to be used, procedures to be followed and the functional requirements applied when developing complex software (Hirschberg, 2000). Figure 16 below gives a very much simplified example of its possible use within the NHS to highlight gaps within a department's managerial systems.

The inputs into the management system cascade down the left hand side of the V, and range from the government and Strategic Health Authority strategies, through the nationally recognised management standards, and on to the technical briefs and specifications; i.e. increasing in levels of detail. At the base of the 'V' are the actual individual tasks that are being undertaken. It is these tasks within the management system that are assessed against the standards that rise up the right hand side of the 'V', i.e. decreasing in level of detail.

The first of these standards on the right of the 'V' are grouped under the heading Technical Specifications, and in this case make reference to NHS Health Technical Memoranda, Approved Codes of Practice and manufacture's recommendations. These in turn cross reference to the left of the 'V'.

The second node that is rising on the right of the 'V' relates to the NHS Healthcare core and developmental standards, Statutory Acts and the NHS Corporate Governance Risk Management Standards. These follow on to the Corporate Governance responsibilities

of the organisation, and the ISO 9000 guidelines relating to good quality managerial systems. Lastly the activities are compared to the Trusts Clinical and Estates strategies which should match the government's and Strategic Health Authority's strategies.

Each reference document is made up of a series of discrete tasks that must be undertaken to fulfil the requirement of the management system under investigation. Assuming that the task under examination conforms to all of the standards included within the left and right legs of the 'V', then the task is deemed to pose no untoward perceived latent hazard to the patient, organisation or society as a whole.

Figure 16   *A generic much simplified 'V' model structure for a department's output requirements*
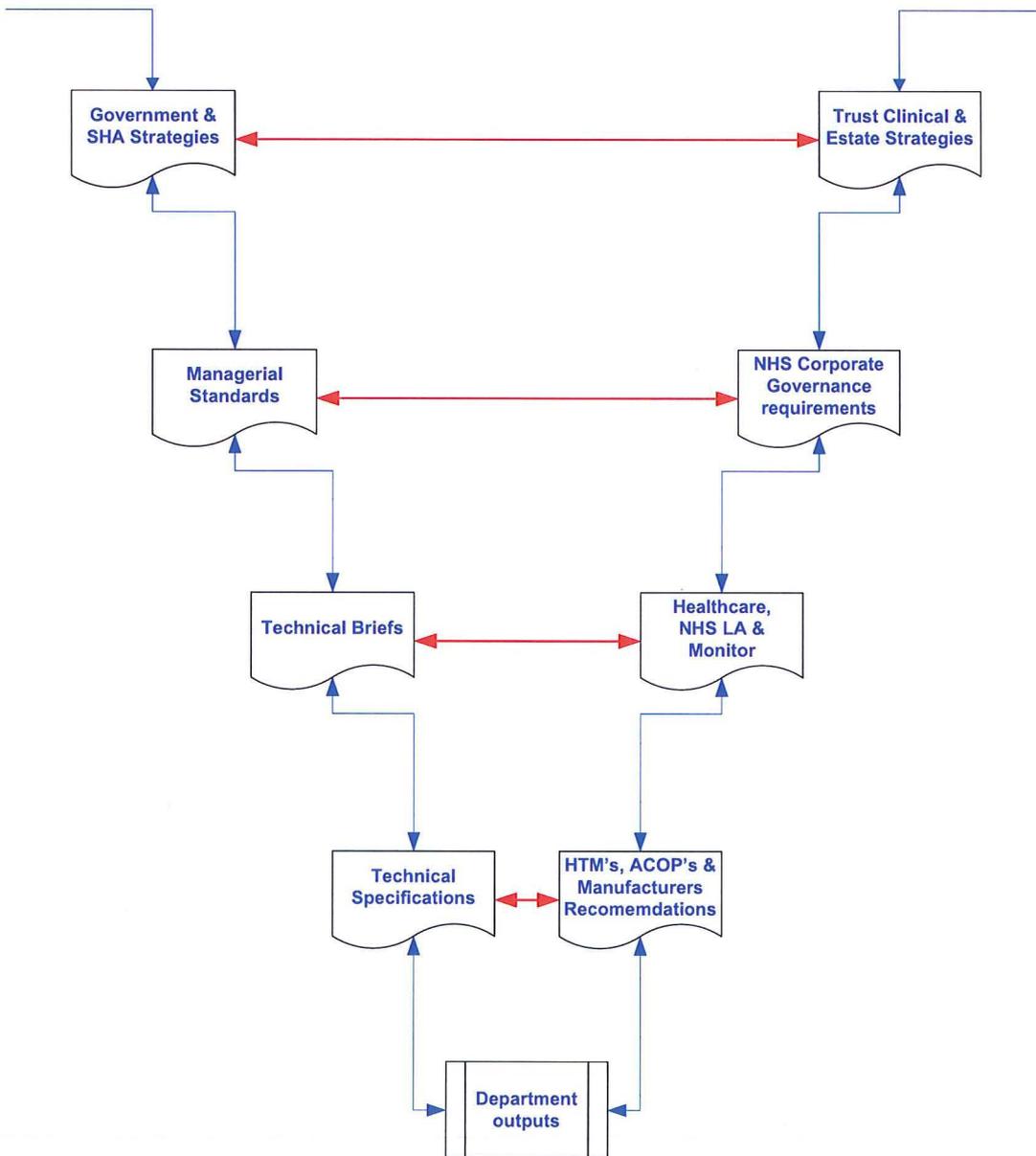
If however, there is a gap between the standards and the actual task(s) undertaken, then a perceived latent condition is assumed to exist that is generating a perceived risk. It is at this point that the level of perceived risk must be analysed by a risk characterisation

process. Following the analysis the relevant managers must decide to act (or not). This naturally moves the monitoring process into the next phase, what to monitor.

The final data plan must be agreed by all of the actors within the process and those receiving the outputs of the monitoring information. On collecting and delivering the monitoring information a 'negative feedback' loop would be created, which could inform those departments controlling their systems inputs of the likely effects of their actions through an analysis (simulation) of their processes. One important consideration must be the format in which to present the data, as each department would probably have different needs, methodologies of interpretation and use of the records. Here the views of the line managers throughout all levels must be sought, whether they are creating the initial work load stream, managing the work process or using and relying on the work load output. Throughout and beyond this development phase of the project the line managers at the operational level must be in agreement that the information produced will assist them to assess their performance, and will provide them with the information that is of use to them. On completion of this phase the inputs and outputs of the department would be known, therefore the design of the efficiency measures could begin.

3.5.8 The design of the efficiency measures

The techniques created within this methodology chapter are aimed at producing operational performance management information, that would assist in the analysis and forecasting of both resource budgeting and management control. The performance measurement would begin from an observation of the gaps between the accepted and recorded values. The cause and seriousness of the deviation would be assessed and suitable action taken. Sometimes the comparison between the actual values could lie within the acceptable tolerance limits, but could show a trend suggesting that the process is heading towards a breach of these limits, thus giving a warning of things to come. The gap analysis would contain as an example:

* Descriptive and numeric data comparing the actual against the acceptable figures
* Calculations of the size of the gap
* Future trends should the gap be ignored

- ❖ Explanations of the possible cause
- ❖ Resources needed to close the gap
- ❖ Actions in priority order to close the gap (reverse the trend)
- ❖ Proposed implementation programme

The design of the efficiency measuring methods would rely on the correct identification and measurement of the outputs, quantification of input/output ratios and the delineation of the dimensions of quality. From this performance could be assessed in terms of its efficiency and effectiveness through the indicators proposed.

The efficiency of a process would be a function of its input and output ratio. The effectiveness of a process would be the assessment of its contribution to the achievement of the Trust's overall strategic objectives/targets. Both the indicators of efficiency, quality and effectiveness must be developed over a number of time periods for the analysis to be meaningful, when forecasting projections in relation to resource and risk.

It was important to separate the terms of efficiency and effectiveness into two separate indicators, as it is possible for a single process to be highly efficient but have little effect when assisting the Trust to achieve its strategic objectives. The reverse is also true, where a low efficiency process could be vital to the Trust in their strategic planning. By separating these two indicators it allows for the identification of the greatest importance to the organisation, so that they could target their limited resources to where they would do most good. This also assists in removing the belief that: 'by increasing a department's resource, the more effective it would become'.

Fundamental to the development of the indicators is the acceptance of the 'means-end' principle, that begins with inputs such as labour, materials and capital, which in turn lead to the process ends, the operational outputs, such as the completion of reactive maintenance tasks. In a similar way these operational outputs serve to form the 'means' for producing the programme 'ends' (outputs), such as meeting the NHS Healthcare Commission's core standard C21 which would be one of a Trust's strategic objectives in relation to the patients 'well being'.

145

The values of labour, capital, materials, etc reflect the expenditures on a particular process, and cannot be considered to be useful proxies for changes in well being. The process outputs are the services produced as part of an objective/target. These are, for example the completion of the scheduled planned preventative maintenance programme for a given time period. This creates the unit costs for producing these outputs, which in turn reflect the operational efficiency of the process. These figures can only represent partial proxies in relation to the overall effectiveness of the process as they relate to efficiency only. It is an objective's output that reflects the effectiveness of the process, rather than how the process was undertaken and completed. It is these indicators that are the real proxies as it is these high level outputs that effect a patient's well being.

In the above, various levels of outputs have been discussed in terms of their use as proxies for the objective of well being. Even if well being could be perfectively measured there would still be a need to measure the process efficiency and effectiveness to exercise managerial control over the process. In the private sector there is a raft of operational performance indicators managed below the final and ultimate profit and loss statement. The control of the eight engineering topics for their day-to-day operational maintenance activities within an NHS Trust is no different.

3.5.9 Measuring the efficiency of a process

The efficiency measurements for these eight engineering topics would be associated with the tasks undertaken by the artisan staff of an NHS estates department. This basic measure relates to the ratio of the processes' operational outputs to the inputs used within the production, e.g. if 200 reactive maintenance dockets were completed by expending 60 man-hours, then this would equate to 3.03 completed dockets/man-hour. The inverse of the efficiency ratio represents the unit cost of the output, in this case being 0.3 man-hours per reactive maintenance docket. To compare the efficiency across a number of time periods an efficiency index would need to be created. Figure 17 below gives an example of such an efficiency index graph. The efficiency index would be computed from the following formula:

Efficiency index (yr n) = (Output/input ratio (yr n) x 100)/(Output/input ratio (base year))

Figure 17   An example of an efficiency index graph.
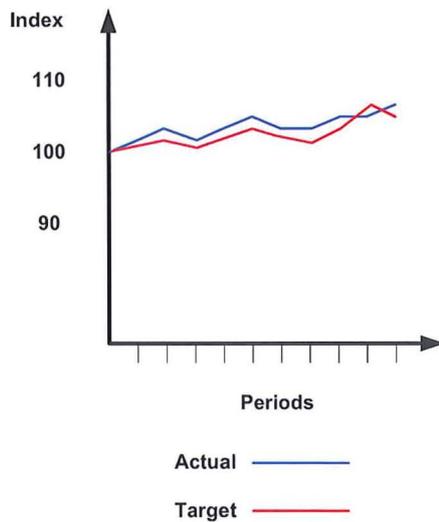
A base period would be awarded a value of 100 and any subsequent period compared against this base value. The more efficient periods would attract a ratio greater than 100, and those generating a lower efficiency a ratio less than 100 when compared to the base period. Where there were a number of like efficiency measures, these could be aggregated to summarise the group's or department's performance. In these cases the operational performance of each measurement ratio would be monitored through the efficiency ratios, e.g. separate efficiency ratios for electrical, mechanical and semi-skilled sections for each site within a Trust. But the efficiency measures for all 'like' operations within the department would be aggregated into one efficiency level, e.g. artisan efficiency ratio for the department. The use of a weighted figure could be employed to differentiate between trade groups to arrive at a weighted efficiency index for the department as a whole, but the researcher questions the added value of this increased computation and the complication that this would give to the analysis.

An important consideration when measuring efficiencies would be to account for non-productive time and overhead costs. The workforce naturally produces unproductive time through holidays, sickness, training, etc. This again could be measured using an efficiency index, where the base figure was a set target agreed by the senior managers as a reasonable value for their Trust. Overheads include such topics as buildings, supervision, administration services, etc. Within a Trust these could be split into costs and man-hours, but whatever figures would be produced they would be a constant,

147

therefore the creation of these on a periodic basis for day-to-day monitoring would be meaningless, unless they were used as part of a benchmarking exercise. But as previously stated such analysis is beyond the scope of this thesis.

The operational performance measurements above would not indicate if the organisation was operating to the maximum possible efficiency, but it would allow a comparison between periods and indicate if it was doing better or worse than other periods. For such a detailed assessment against the maximum possible efficiency, the Trust would have to implement a work measurement programme such as a Labour Management Scheme, but this is beyond the scope of this research.

The cost per man-hour could have been used in place of the hours expended to create the efficiency ratio, but this would have to have been subjected to price inflation, wage increases, etc, before any comparison across time bands could have been made. The input expended to produce an output includes all resources used to create the output, e.g. labour cost, materials, capital, buildings, land, time, etc. Defining the input components to produce an output is fundamental in any operational performance measurement and requires an input specification to be created.

3.5.10 Input specification

In practice the measured performance is only a partial measurement as it is often impractical to include all factors which go to create the final output. Those readily quantifiable are normally labour and consumable costs over a stated period of time. Thus only partial efficiencies are normally calculated and 'trade-offs' in accuracy must be accepted. Inputs could be specified as: units of labour, salary cost, operating and maintenance costs or capital costs. These four types of input descriptors are briefly described below.

Units of labour

These are readily available, such as man-hours actually expended completing all reactive maintenance tasks per week. The number of trades people employed in a given time period would not give a consistent measure of input to the output, as sickness, holidays etc would vary, as time input is not a function of price changes, it is a measure of 'real' input. Therefore the use of worked time would

148

give a more consistent measure of input over time. But this takes no account of an individual's level of efficiency or skill; again a trade-off must be accepted if units of labour are employed.

Salary cost

Use of the salary costs to produce a given output would be hampered from a like-by-like analysis with respect to time, due to the differing wage rates paid to the varying skilled staff employed, possible overtime rates were paid for certain duties, and the change in inflation rates from one year to the next. These problems could be overcome by assigning various price adjusters within the calculations, but this would make the maintenance of such a measurement system complex.

Operating and maintenance costs

This unit of input measure would combine all labour and non-labour costs together. The identification of the contributing factors and the accuracy of the estimations could be simplified to exclude capital, land, buildings, etc. The problem for an NHS estates department would be separating the non-labour revenue costs from those relating to the maintenance function, particularly if a small directly employed labour force was employed. Again, as with the previous section relating to the subject of inflation, etc, the complexity would be increased when maintaining such a system of measurement.

Capital costs

Including these costs for the proposes of a day-to-day monitoring system would create a major problem, as buildings, plant and equipment depreciate in value, at different rates and at varying levels with respect to time. Also there would be time lags in relation to determining the opening and closing values of these various assets.

As described above, an NHS estates department produces a range of outputs in relation to the eight engineering topics studied as part of this research. These must be combined to form a small number of performance measures. The major determinant influencing the preferred method includes factors such as accuracy required, measurement system

complexity and the availability of the relevant data. In all cases within this research the recommended method by the researcher was to adopt the units of labour measurement. The accuracy of the computations was acceptable, the data was readily available and no alteration to the existing managerial systems was needed to generate the information. The administration of the measurement system was simple to operate and understand, particularly by non-technical peoples.

With the methodology outlined to identify the relevant inputs and desired outputs of the various processes, and the calculation of the relative efficiencies defined, it would be necessary to develop and design a method for measuring the quality of the process. The measurement analysis should report on the quality of the outputs. This would give an indication of the quality with respect to pre-agreed standards. An example of such quality measures would be the timeliness of the output. In terms of the artisan activities these quality measures would be split into two topics, reactive maintenance and planned preventative maintenance tasks.

It is common for Trusts to set response times for the completion of reactive maintenance tasks relative to the urgency that the user places on the failure being reported. For example the urgency could be defined in one of three ways, as shown within Table 13 below (NHS Estates 2002).

Table 13   Typical reactive maintenance fault classification and response time needs

Classification	Definition	Response time
Emergency	Fault creating immediate danger and/or disrupting a key business operation	Immediate
Urgent	Fault could develop into an emergency	Within 24 hours
Routine	Fault is a non-emergency and is non-urgent	Within seven days

The measure of quality would be a measure of the department's ability to match (or exceed) these pre-set response times. Again the measurement could be presented as an indexed graph for each response need. The conventional planned preventative maintenance system uses a series of regular inspections against a set pre-agreed

150

schedule. The frequencies of these tasks are commonly: weekly, three weekly, six weekly etc. In this case the measure of quality would be a measure of the department's ability to match these pre-set frequencies. As with the artisan functions, measurements of the management function could be formed from the timeliness of the outputs from the line managers against a pre-agreed schedule. Again the measurement could be presented as an indexed graph for each response need, probably aggregated into an overall departmental measure of quality.

3.5.11 Measuring the effectiveness of a process

Prior discussions created the case for the need to measure outputs in terms of efficiency, quality and the effectiveness of the output in achieving the Trust's overall objectives. Of all three measures it is this final evaluation that will cause the most problem to design and determine to an acceptable and useful level of accuracy.

The first step would be to determine the programme output; this could be the socio-economic effect of creating a patient well being. This is quite different to the operational inputs which would be task orientated and relatively straight forward to measure. Probably the way forward would be to measure the effectiveness of the operational outputs against an intermediate operation, which itself would form an input into a higher-level operational output. This higher-level output may itself be another intermediate output or the final programme output.

The ratio of programme output to operational output would form a proxy measure of effectiveness. But this approach has its drawbacks: it only measures a relationship not its significance, e.g. should the ratio be greater and if so how much greater? It ignores the influence of external factors, and it disregards the effect of any time lags between the two outputs, and whether relationship is linear or not. If the effectiveness could be measured then it should be possible to produce meaningful targets for the effectiveness to achieve. This measure is not only determined by the values of quantity and quality of the operational outputs, but is also influenced by the operation's interrelationships with other operations both within and external to the department. Ideally these other factors influencing performance should be delineated, leading to quantitative relationships specified to link all significant variables.

151

What this short discussion has shown is that there is no simple answer to the evaluation of operational effectiveness. As the researcher dug deeper into this topic it became obvious that the development of a robust measurement of effectiveness was beyond the scope and timeframe of this thesis, and therefore would be included within the discussions relating to further research topics.

From the development and design of monitors the efficiency and effectiveness of the processes defined the project should naturally move on to the next phase, which would be the creation of a reporting structure.

3.5.12 The reporting system

The content and presentation of each report must be matched to the needs of the individual line managers. Some would prefer graphical as opposed to tabular presentations, or a combination of both. In either case a narrative should accompany the data to give explanations, interpretations and elaborations as necessary. As a general rule the level of detail from the lower ranks would be summarised when reports are presented to the next higher level of authority. In all departments the need for this vertical dissemination of reports would be without question. But in such a highly complex interactive and tightly coupled organisation, it must be recognised that in addition it would be necessary to design and disseminate operational performance reports, also on a horizontal level, to other departments on a similar systematic basis.

The frequency of the reports would vary dependant on where the recipient sits in the hierarchal structure. Volume and efficiency data would be dependent on the variability of output demand and input. Therefore where these figures were subjected to high fluctuation, then these types of reports would be required more frequently to ensure that resources could be matched against output needs in a timely manner, e.g. weekly or monthly.

In contrast, the levels of performance relating to the effectiveness of the operation would probably only be reported say every quarter. This would be in recognition that to view progress in its proper perspective then it would be necessary to report against a greater time period, as month to month variations would normally be insignificant and/or immeasurable. Also the frequency must be matched against the time it would take to

create and disseminate the report. There would be little point in reporting on a weekly basis if it took three weeks to produce and circulate the document.

The reporting system would be split into five parts which must be integrated into the overall managerial system for the department and organisation. These five parts would be: setting the targets, initial planning to achieve the targets, measuring the results from the pre-determined methodology, comparing the results achieved against the set target, and finally taking some form of corrective action to bring the situation back on target. Therefore the reporting system for the operational output of a department must contain measures of efficiency and effectiveness in parallel to the more traditional personnel and budgetary information, etc. Outputs should be presented on an aggregate and individual basis in both an absolute and index format. The internal and external factors which would have an effect on the output should be reported, if they could be identified and/or quantified. The data relating to the measurements of the quality of the output against the pre-agreed standards should also be recorded.

The performance report should contain corresponding input measures. These should be split into:

a. Input utilised in measurable operations, to include apportioned overhead.
b. Input utilised in immeasurable operations, to include apportioned overhead.
c. Input utilised in overhead operations, indicating allocations to measured and unmeasured operations.

The inclusion of this data would assist in the analysis of any efficiency changes. The measures would also facilitate the development of the unmeasured to total input ratio, and the overhead to total input ratio. The operational efficiency indices could be presented as unit volume or unit cost ratios. But as described earlier the use of a cost figure complicates the administrative burden of the report production as inflation, wage agreements, etc must be converted to a common base for any meaningful comparison and interpretation to be made with respect to time. Operational effectiveness measures must be included which would show target vs. actual programme output/operational output ratios. If it was not possible to directly measure the programme output then proxy measures for the output volume and quality measures must be employed.

153

The previous phases indentified the inputs and required outputs of a department. Once completed the project then would then progress on to creating a managerial structure for the monitoring process.

3.5.13 Structure of the monitoring process

As previously discussed reference was made to the work of Perrow (1999) and others when researching into why systems fail, and the comparisons between these failed organisations with those regarded as High Reliability Organisations A common theme developed from their research was that High Reliability Organisations' managerial systems were designed with inbuilt parallel defensive systems. From research into errors in maintenance it has been proposed that managers should develop their monitoring systems to mainly deal with error-provoking tasks and error-inducing situations, as opposed to error-prone people (Reason & Hobbs, 2006); and that based on the evidence organisations should: 'be treating errors as an expected and foreseeable part of maintenance work'. Reason and Hobbs research has shown that no one type of defence within the managerial system is sufficient to capture all errors before they cause undue harm, but that an array of such barriers needs to be constructed. Ideally these safe guards should have both redundancy (multiple backups) and diversity (a variety of safeguards), and be designed to limit and contain the effects of the errors. They go on to state that: 'removing error-promoting situations and improving defences form the two most important parts of an effective error management programme'. Checks and inspections are examples of defences that are designed to identify these latent conditions (errors).

Following this research parallel defences have been formed as an integral part of the monitoring system. A typical way of presenting these types of defences pictorially is shown below in figure 18. The diagram in figure 18 shows a series of boxes representing the parallel managerial defensive systems, surrounding a red dot signifying an incident. The inner most box (first defence system) would be designed to identify the incident (the red dot) and to surround and contain it. This then would alert the organisations manager's of the incident so that it could be immediately contained and corrected. The box is shown not to be a perfect square. This is in recognition of the work of Perrow and the Normal Accident theory, which in its simplest form states that all management systems will fail sometime.

Figure 18   Pictorial representations of managerial systems in parallel
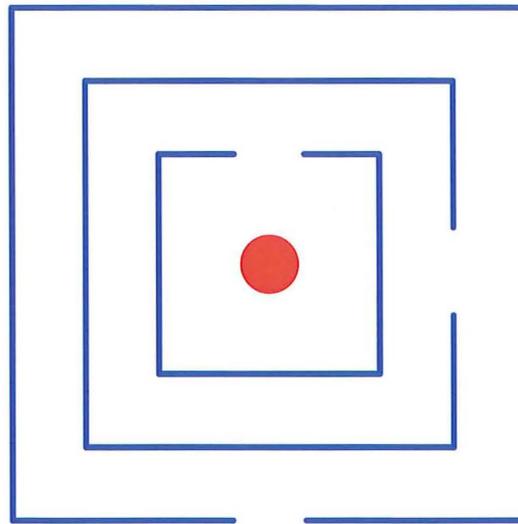

Reason and Hobbs, 2006 Page 13

Assuming that the inner most box (the first defence) fails to recognise the incident for whatever reason; the middle box (the second defence) should identify the incident and similarly alert the managers. In a similar way if both the first and second defences fail to spot the event, then the outer box would be designed as the third and last line of defence. When this theory is applied to a practical application with an NHS estates environment, the managerial structure would resemble that in figure 19 below. The figure shows a typical chain of command within a NHS estates department, from the trade's staff through the supervisor, engineering officer, estates manager and up to the board director. One form of parallel defence is shown as an external auditor reporting on the actions of the estates operatives and managers. The weakness with this reporting structure arrangement is that the external auditor would be reporting into the department being audited. Therefore any adverse comments could stay within the very department that could be causing the problem.

A more robust defensive system is shown that includes the risk manager, control of infection officer and internal auditor as an example. In this arrangement the three managers are totally divorced from the estates department in terms of budget or managerial control, but all have a personal responsibility to advise the Trust of any deviation away from national and/or locally agreed standards/working practices. By

involving these three managers within the monitoring system, they could report any departure from the agreed standards through their own department's managerial structures, free from the influence of the estates department

Figure 19   Basic parallel defence structures.
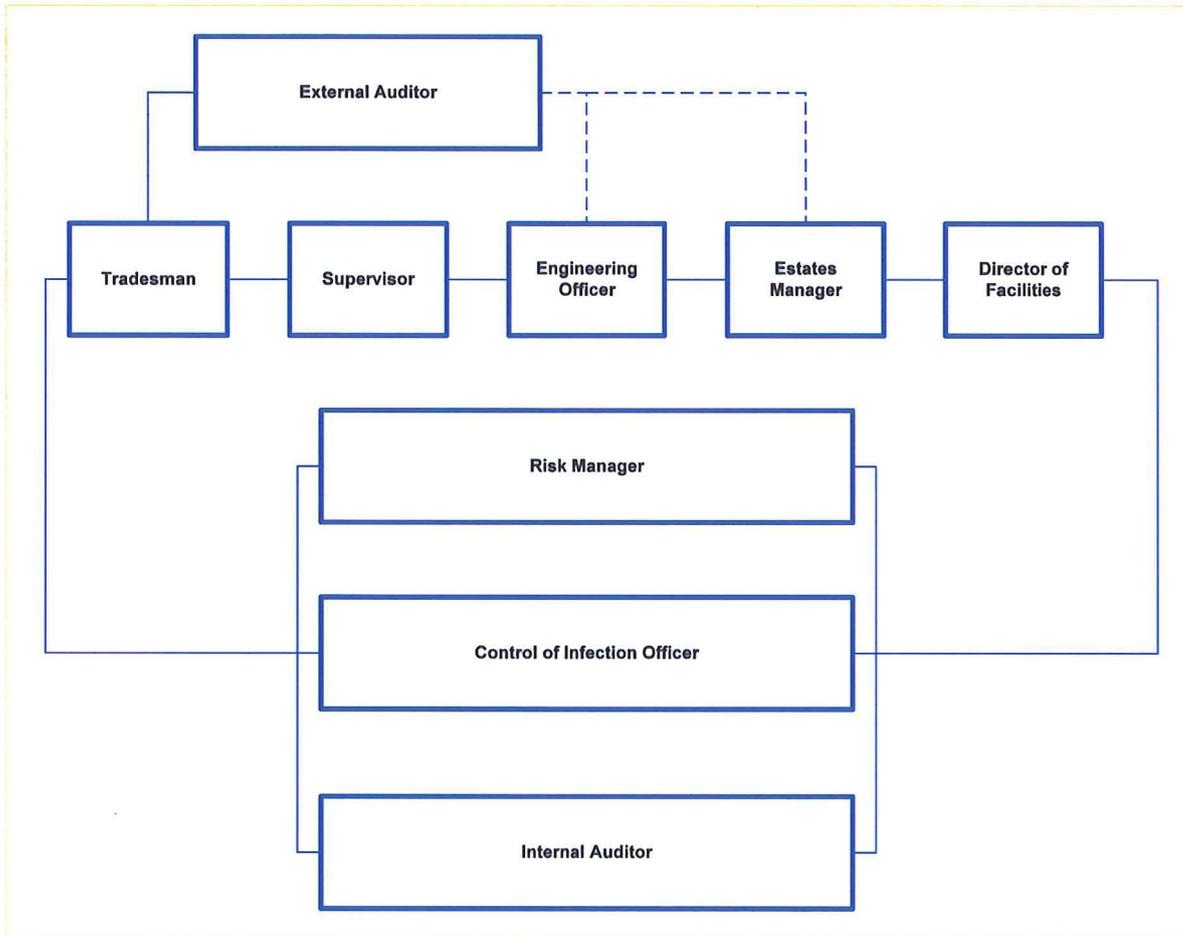
In the experience of the researcher, one of the fundamental flaws within some monitoring systems is their reliance on a single person (or small group) to operate and maintain the processes. The reliability and robustness of the whole monitoring process would be greatly increased by removing this immediate human dependence, and creating an independent 'driver' to force the monitoring procedures to be completed.

3.5.14 The driver

Within the structure developed and designed within this research, the creation of this independent 'driver' is achieved by utilising the existing estates department's planned preventative maintenance software. This currently produces work dockets for use by the trades' staff from a prescheduled list of periodic tasks to be completed at set frequencies throughout a calendar year, e.g. weekly, three weekly, six weekly quarterly, half yearly and yearly. The software records the issuing and completion of these planned preventative maintenance tasks for analysis and auditing purposes. This research has adopted this software to produce dockets addressed to the managerial staff of the Trust. The instructions contained within these dockets would be for the respective managers to complete their monitoring tasks, all against a pre-agreed schedule. This schedule would form an assurance framework as shown in figure 20.

The figure shows an example of the various personnel within a monitoring regime together with their frequency of monitoring. The system operates without relying on a manager's memory to undertake a certain monitoring task, as the dockets are produced automatically by the computer, dispatched and returned via the internal mail system. As a typical Trust estates department produces approximately 800 such dockets a month, the additional docket production per month to operate the assurance framework system is negligible. The next section explains how in practical terms the assurance framework would operate.

3.5.15 Assurance framework operation

The work content contained within the dockets would instruct the recipient to undertake a specific monitoring task, with the results of the task being recorded in the computer for analysis and auditing at a later date. The instructions, technical detail and knowledge for each person within the framework would differ depending on their place within the organisational structure.

As an example, the engineering officer would be trained in the requirements of the national codes of practice, and have intimate local site knowledge of the plant and equipment fitted. His command of the technical and practical aspects of the topic would be far superior to say the Trust internal auditor. It would be the engineering office's role to manage and oversee the safe operation of the engineering topic in question, but it

would be the internal auditor's function to ensure that the locally agreed monitoring procedures were been followed, and that the staff have been involved at the correct parts of the process and have undertaken their respective monitoring duties to the agreed timescale. Any deviation from either the national or locally agreed standards would be uncovered and reported to the Trust senior managers, as the robustness and reliability of the process would be complemented by the three (sometimes four) parallel defence systems inbuilt into the monitoring assurance framework.

Figure 20 An example assurance framework.
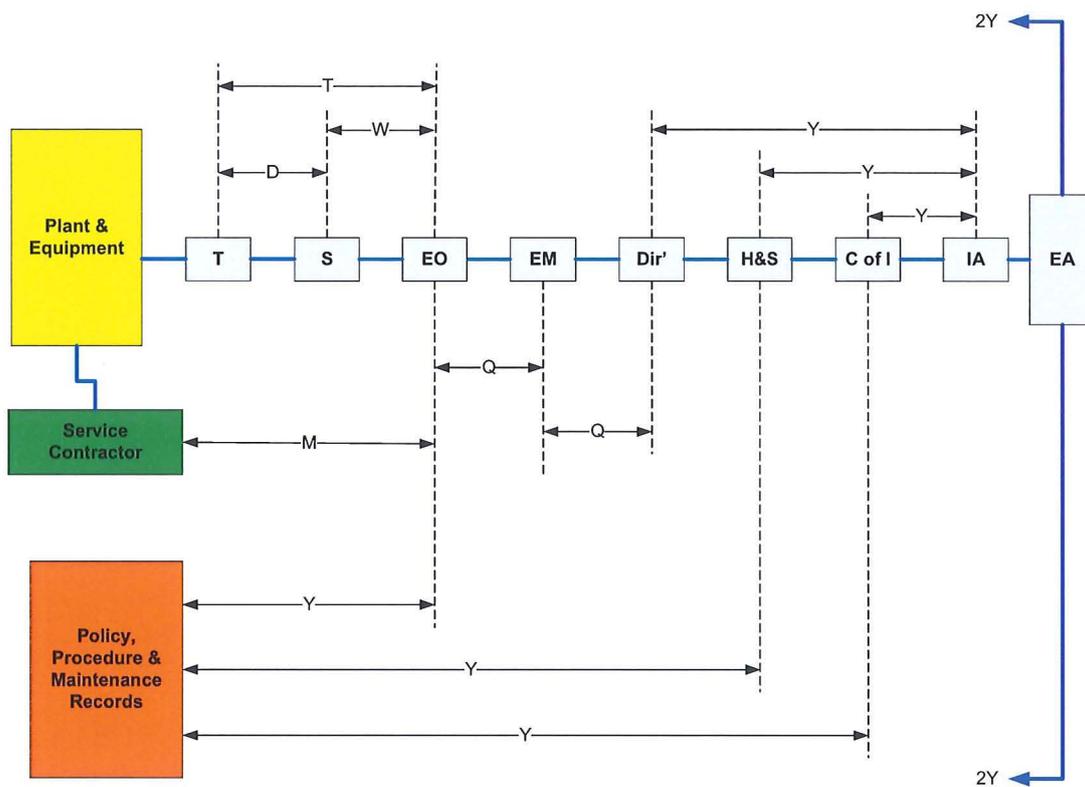
Key

T	Tradesman	H&S	Health and Safety (Risk) manager
S	Supervisor	C of I	Control of Infection Officer
EO	Engineering Officer	IA	Internal Auditor
EM	Estates Manager	EA	External Auditor
Dir'	Director		

Figure 20 shows the involvement of the Trust's health and safety manager, control of infection officer and internal auditor, forming three of the four parallel defences. Each of

these personnel would operate through a totally separate managerial structure from those within the estates department, yet they would all have a role to play in monitoring various activities and advising the Trust of any risk that may be generated. The health and safety manager would report to the Trust health and safety (risk) committee, the control of infection officer would report to the control of infection committee, and the internal auditor in turn would report to the Trust audit committee. Depending on the governance arrangements within the Trust, all of these committees eventually would report to the Trust board. The fourth defence system that could be employed would be created through the commissioning of an external auditor (advisor) reporting to a committee or management stream outside of the estates managerial chain of command. His/her role would be to provide a managerial and technical evaluation of the monitoring process and content as a whole to create a culture of 'continuous improvement'.

The previous sections discussed the methodology used to create a driver to ensure that the locally agreed actions were completed to an acceptable standard and within a tolerable timeframe. Once this basic operational assurance framework has been agreed, then the detailed actions for each operative within the monitoring structure could be created.

This phase concludes the development of the whole of the operational performance measurement system, as the inputs and outputs of the department would have been matched to the national and local needs of the Trust. The efficiency and effectiveness indexes would have been created from which the department could be measured. And the design of the generic mechanism to drive the performance management process completed, with inbuilt managerial parallel defence systems to create the high reliability demanded.

This would conclude the work of the project design team, but before they were disbanded they would have to record all aspects of the operational performance measurement system developed; report on the methods of approach, the outputs considered and those rejected with reasons given; report on the availability of the information systems and report on the reporting format. Then they must develop a plan for the implementation phase of the project, this would involve setting up an implementation team.

3.5.16 The implementation plan

Following the disbandment of the project design team the implementation team must be formed. Their role would be to prepare the briefing and instruction manuals, deliver the orientation training to the operatives, commence the data collection and begin its processing.

The introduction of the monitoring system could be either simultaneously across the whole department or phased. Whilst introduction across the department as a whole has the advantage of a reduced implementation programme, in large complex highly interactive and tightly coupled departments this approach could be problematic; therefore a sequential approach would be favoured as the practical alternative, particularly if the resources of the implementation team were small. The time frame for the introduction must be specified within the project plan, starting from the bottom and moving to the top of the hierarchal structure, but including time for the horizontal process that must take place

The time horizon for this implementation phase must take account of the available information. It may be that a first part would be concerned with the data immediately at hand, the second part utilising data accessible only in the short term, through to the final and third part when all data would become accessible.

The methods for collecting and processing the data into the pre-agreed report format and frequencies should also be included within the implementation plan. It should state the manual and automatic steps of the process, the personnel responsible for each of the steps, the analysis methods, the training required for each operative/line manager, and the suggested implementation team structure.

The implementation team reporting to the Steering committee should include representatives from the department that the operational management system had been designed for. In this way the team would consist of a diverse range of skills, knowledge and experience to implement the design of the monitoring system in the sprit and content of the original intent. An essential part of the team's tasks would be to create and deliver familiarisation training to all staff in the monitoring process loop. This would include an explanation of its need and links to the overall Trust's corporate governance

requirements, objectives, risk assessments and to legal compliance issues. Particular stress must be made to all concerned that the measurement system is only relative and not absolute; and that it is a representation of the group as a whole, and thus could not be used to assess an individual's performance in terms of efficiency and/or effectiveness. Also the staff should be given a clear indication of how the information would be interpreted, used, and by whom, for planning and managing operations. Without this initial training and support then the monitoring system would quickly disintegrate and become unreliable. Once implemented the operational performance measurement system would enter its final phase of maintenance. In this state the whole performance management process should be continuously reviewed to ensure that its composition is still delivering the information applicable and relevant for the needs of the Trust. This is in recognition that conditions change both within and outside the organisation, and that organisational structures, reporting arrangements and departmental functions naturally evolve over time. Therefore the measurement system must be constantly appraised to ensure that it is not only reflecting reality, but that it is also measuring the relevant topics. This would involve the critical re-examination and alteration as necessary, of the outputs and their structure, the validity of any weighting applied, and a reconsideration of the continued validity of the base period integral to the efficiency and effectiveness indexes. In other words a culture of 'continuous improvement' would be created.

3.6   Chapter Three Summary

The literature search demonstrated to the researcher that to adequately answer the research question the methodology employed must be split into four disparate research objectives, in order to examine the managerial system as a whole.

First research objective

The methodology created for each Trust taking part within this research, a multi-professional focus group comprised of a Trust's own senior managers, for each Trust analysed. Their role was to compare their managerial systems to the respective national guidance, identify the gaps (latent conditions) and assesses the level of risk generated. Within the methodology developed as part of this research to assess these risk levels was the rejection of the NHS national standard '5x5 risk criticality grid' in common use throughout the Service, in favour of a series of specific 'Utility Functions'. It was shown

that these give a transparent, quick, repeatable and auditable assessment of risk that is far superior to the NHS risk assessment system. A critical review of the main factors that could influence the risk score developed by the participants using this preferred method is presented, together with a description of the techniques used to minimise their effects. To the researcher's knowledge, this is the first time that the use of 'Utility Theory' has been applied to such scenarios.

Second research objective

The methodology for this objective designed and used the simple project management techniques of Gantt Charts and Monte Carlo simulations, to determine the resource needs in terms of man-power, cost and time frame to close the identified managerial gaps.

Third research objective

Whilst the outputs from the first two research objectives of the process identified the latent conditions creating the risks, and estimated the resource needs to close the managerial gaps to reduce the risks to an acceptable level, they did not give any indication of the revenue consequences needed to keep the gaps closed in terms of the directly employed artisan shop-floor staffs. Thus the crux of this objective was to assess the effects of reducing the available artisan man-power resource below an acceptable level, and to predict the resulting increase in risk. The technique used to explore this previously unsolved problem within NHS estates departments created a simulation program that modelled the changing work load patens, resource availability and assessed risk with respect to time. The methodology used a Trust's own historical data and risk profiles derived from utility functions created via a Trust's specific multi-professional focus group; therefore the resulting outputs were particular to each individual Trust.

Fourth research objective

The methodology developed and designed a generic day-to-day monitoring assurance framework that integrates audit and review, and which can serve multiple functions, e.g. compliance checking, policy enforcement, embedding performance management into governance systems, and creating the structure for information transfer across organisational interfaces.

Chapter Four   Results

4.1   Introduction

The previous chapter gave a detailed description of the methodology that was employed to answer the research question:

> Does the managerial system of internal control within the NHS give adequate assurance that Trust engineering day-to-day maintenance strategies conform to national guidance?

The literature search demonstrated to the researcher that to adequately answer the research question the methodology employed must be split into four disparate research objectives, in order to examine the managerial system as a whole: This chapter presents the results of the research findings with a brief discussion.

4.2   First research objective

The aim of the analysis was to develop an appreciation of the risks that NHS Trust estates departments were generating through latent conditions within their engineering core managerial systems. This was undertaken by referencing their current systems against eight main national codes of practice that were specifically written for the NHS estates departments. The gaps between the national standards and a Trust's actual compliance and associated risks were assessed via a multi-professional focus group exclusive to each individual Trust. The result from this objective was the creation of a Risk Profile explicit to each Trust.

In order to analyse these profiles and form an opinion about the general level of compliance of NHS estates departments relative to the national standards, a comparison of like-for-like Trusts was formed. Trusts within the NHS are classified into five groups to reflect the differing types of healthcare that they provide, see below:

a. **Acute Trusts.** These employ a large part of the NHS workforce. They provide both outpatient and inpatient services which include such departments as accident and emergency, theatres, intensive care, special care baby units, etc.

b. Ambulance Trusts. As the name suggests these organisations control the ambulance service. But they are also responsible for providing transport to get patients to hospital for treatment.

c. Care Trusts. These are organisations that work in both health and social care. They carry out a range of services, including social care, mental health services and primary care services.

d. Mental Health Trusts. Again as the name suggests these Trusts provide services for people with mental health problems.

e. Primary Care Trusts. These Trusts command approximately 75% of the NHS budget. They provide the type of service that people normally need when they first have a health problem, such as NHS Walk-in centres, the phone line service NHS Direct, etc.

Each of the five types of Trust has differing needs from their respective estates departments. This is a reflection of the type of service that the Trusts provide to the patients they serve. For example; Acute Trusts normally operate from large sites which house highly technical building services. In contrast Primary Care Trusts normally provide their function from a large number of relatively small sites with technically simple engineering services. The detailed analyses of the risk profiles were restricted to 10 Trusts in total. All of the sites were classed as Acute, with one exception. In this case the organisation belonged to the Primary Care group. Its inclusion within this detailed analysis was due to the very unusual situation that its engineering building services within its property portfolio closely resembled those of an Acute site, therefore the researcher considered it justifiable to include it for analysis. The combined data for all hospitals for each critical element were grouped into one of the eight topics. The maximum, minimum and average values were computed for each critical element for the three values of weighting, percentage compliance and risk score. From these data three graphs were developed for each of the eight engineering topics, creating 24 graphs in total. The next section (Data Analysis) shows all 24 graphs. A detailed description of the first three graphs is given to aid interpretation. A discussion of the results follows the data analysis section.

4.2.1 Data Analysis

Figure 21 below shows the maximum, minimum and average risk score for the topic of Medical Gas Pipeline Systems (MGPS). It should be noted that late in the year 2006 the name for this topic which was previously Piped Medical Gas (PMG) was changed to MGPS when a revised set of national guidance standards was issued (Department of Health B 2006). The maximum and minimum values for each critical element are shown as a vertical black straight line, with the average value depicted as a blue diamond. The data is arranged so that the critical elements with the highest average risk score are to the left of the graph. The NHS traffic light system of red, amber and green has been superimposed onto the graph, to indicate those critical elements that were generating unacceptable levels of assessed risk. The colours were acting as a guide 'only'. They were part of the final presentation of the risk profile but played no part in the actual assessment of risk.

Figure 21   The Medical Gas Pipeline Systems average risk score for each critical element.
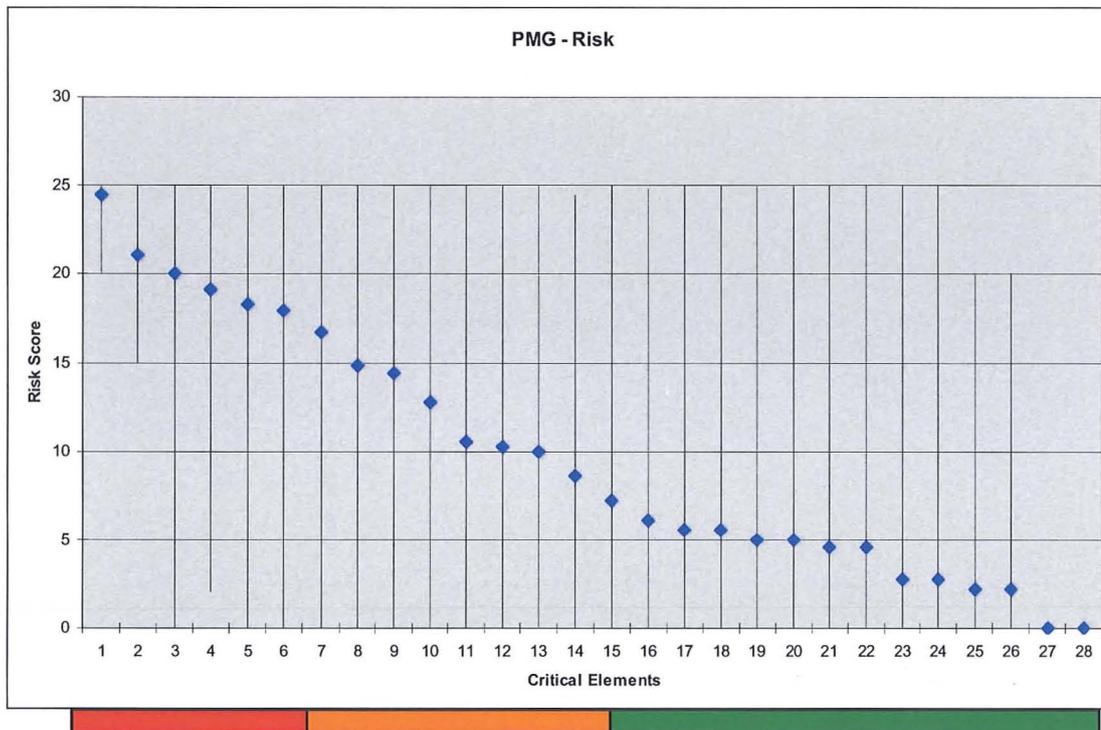
The critical element number one (at the left of the graph) shows a maximum risk score value of 25, and a minimum value of 20, with an average close to the maximum of 25.

Therefore the Trusts have all regarded this particular critical element to be generating unacceptable levels of assessed risk. In contrast, critical element number 26 (at the near right of the graph) shows a maximum risk score of 20, and a minimum value of near zero, with an average value of approximately two. This demonstrates that although the majority of hospitals regard the risk that they are generating from this critical element to be very low (well within the green zone), there are outliers where this is not considered to be the case. From a visual examination of the figure it is obvious that there is a wide spread of assessed risk for the topic of Medical Gas Pipeline Systems. The methodology used to develop the assessed risk scores employed a series of six utility functions that reflected the weighting and percentage compliance for each critical element. Therefore, in order to analyse the data further, it was necessary to break down the assessed risk score for each critical element into the individual components of weighting and percentage compliance, as in the following two graphs. To allow cross references to be made between all three figures for this topic, the critical element numbers are common. Thus, critical element numbers one, two, etc in figure 21 are the same as critical element numbers one, two, etc in figure 22, and so on. Figure 22 below shows the weighting applied to each critical element for the topic of Piped Medical Gas.

Figure 22   The Medical Gas Pipeline Systems average weighting for each critical element.
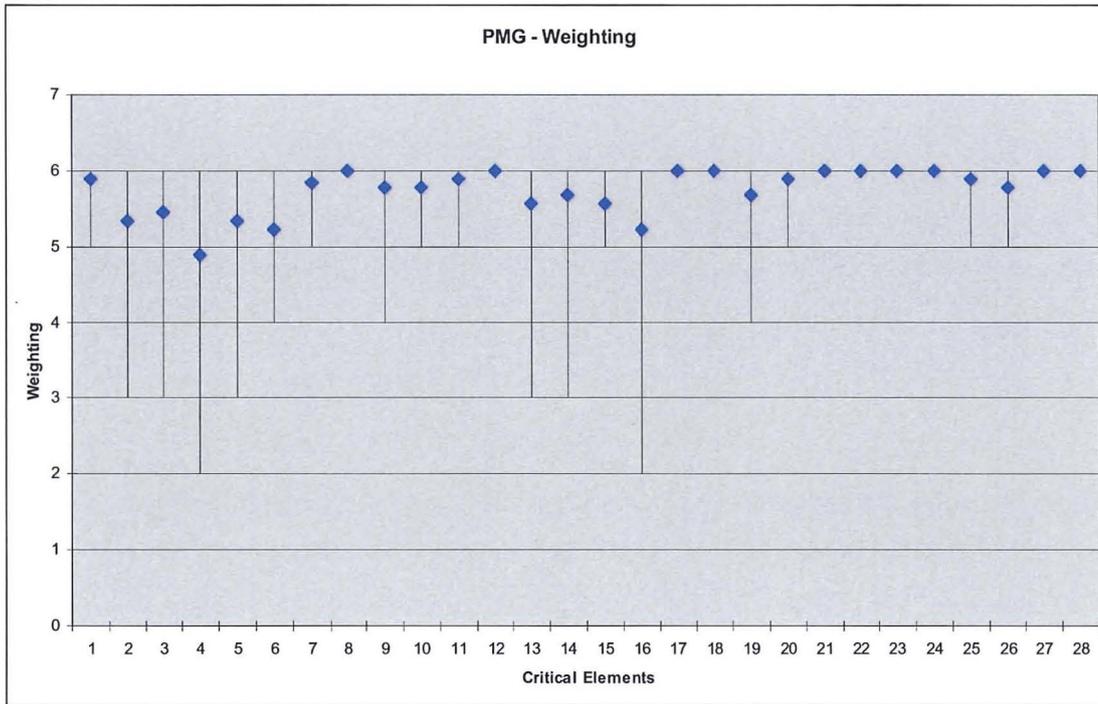

Figure 22 produced a small spread around the average values for the majority of data points, indicating a general consensus between Trust focus groups regarding the importance of these critical elements. To analyse the data further, figure 23 (below) was developed to examine the compliance of the managerial systems for each critical element against the nationally agreed standards.

Figure 23   The Medical Gas Pipeline Systems average compliance for each critical element
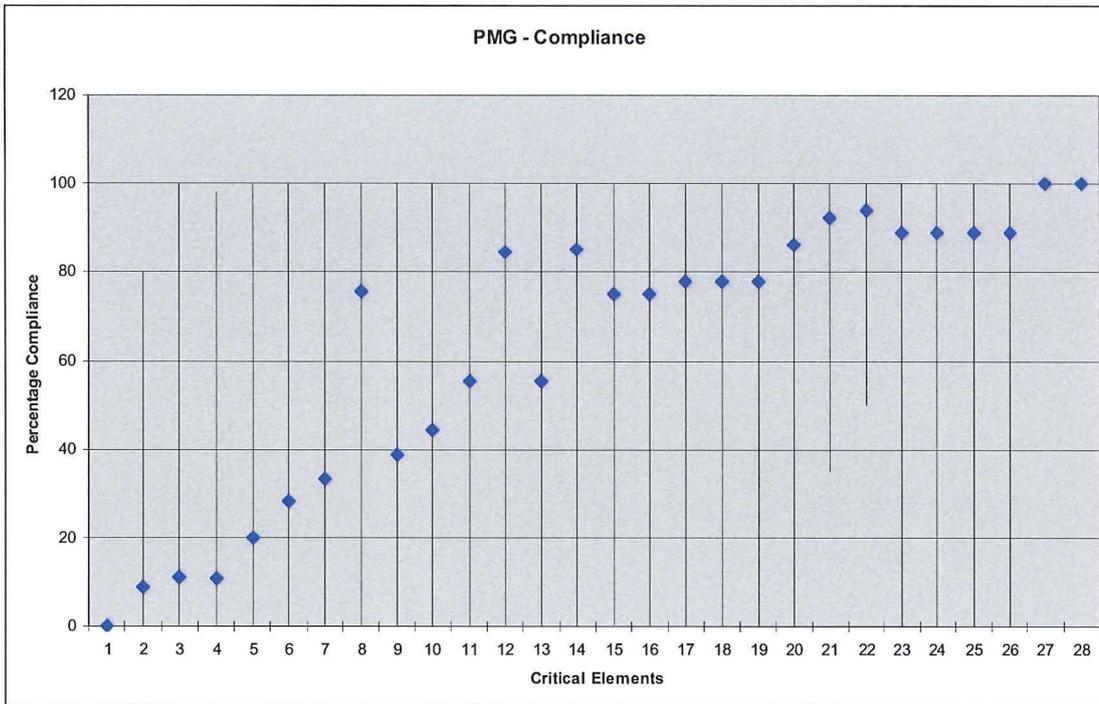
In figure 23 the same convention applies to the numbering of the critical elements as in figure 22, e.g. figure 22 critical elements one, two, etc, are the same as figure 23 critical elements one, two, etc. Critical element one shows a near zero percentage compliance value when assessed against national standards. When this is computed with the average corresponding weighting assigned to critical element number one in figure 22, then figure 21 gives an average risk score of a near maximum value of 25. The interpretation that can be placed upon this result is that even though all Trusts considered that the Medical Gas Pipeline Systems critical element number one is extremely important, their managerial systems fall far short of the national standards. In contrast, critical element number 26 shows a near average of 90% compliance. When this is again computed with the corresponding average weighting in figure 22, then figure 21 records an average low risk being generated from this critical element. The interpretation that can be drawn from this is that all of the individual focus groups considered that the Medical Gas Pipeline Systems critical element number 26 was extremely important, and that on average their managerial systems almost matched the national standard requirements. Figure 21 showed the average level of risk being

generated from the Medical Gas Pipeline Systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red and amber zones:-

❖ The Trust must have an operational policy that clearly states that the executive manager is responsible for overseeing the policy to ensure that it is be properly implemented. This should be carried out a regular basis, and the procedure for such monitoring should be set out in the policy. The document must also outline the procedures to be followed in the event of an emergency.

❖ The designated medical or nursing officer is the person in each department with whom the authorised person liaises on any matters affecting the system, and who would give permission for the planned interruption to the supply. The policy should list these designated persons and their training regime relevant to their area of responsibility, together with the arrangements for cover due to their absences.

❖ The authorised person should be appointed in writing by the chief executive on the recommendation of a chartered engineer who has specialist knowledge of medical gas pipeline systems. The recommendation for appointment or reappointment as an authorised person should be made by a Chartered Engineer who has specialist knowledge of medical gas systems, via an individual assessment.

❖ The authorised person should be consulted prior to the purchase of any medical equipment, which will be connected to the system. And the operational policy should state the procedures to be followed.

❖ The estates department should have accurate and up to date drawings of the system showing main sections and branches, departments served, control valves, terminal unit and alarm systems for each gas service. And each isolating valve should be individually identified by the unique reference number.

❖ Evidence of current maintenance contractor's registration should be by sight of the current and up-to-date certificate of registration. And all maintenance work should be carried out accordance with the relevant codes of practice as a minimum.

An identical arrangement and analysis has been undertaken for the remaining seven topics. This has produced three figures per topic showing risk, weighting and percentage compliance. These 21 figures are presented below. Under each of the seven figures that show the average risk score for each topic is a summary listing of those critical elements falling in the red and amber zones. Following the detailed analysis of the 21 figures is a discussion of the overall results from which general conclusions are drawn.

*Figure 24   The Water average risk score for each critical element.*
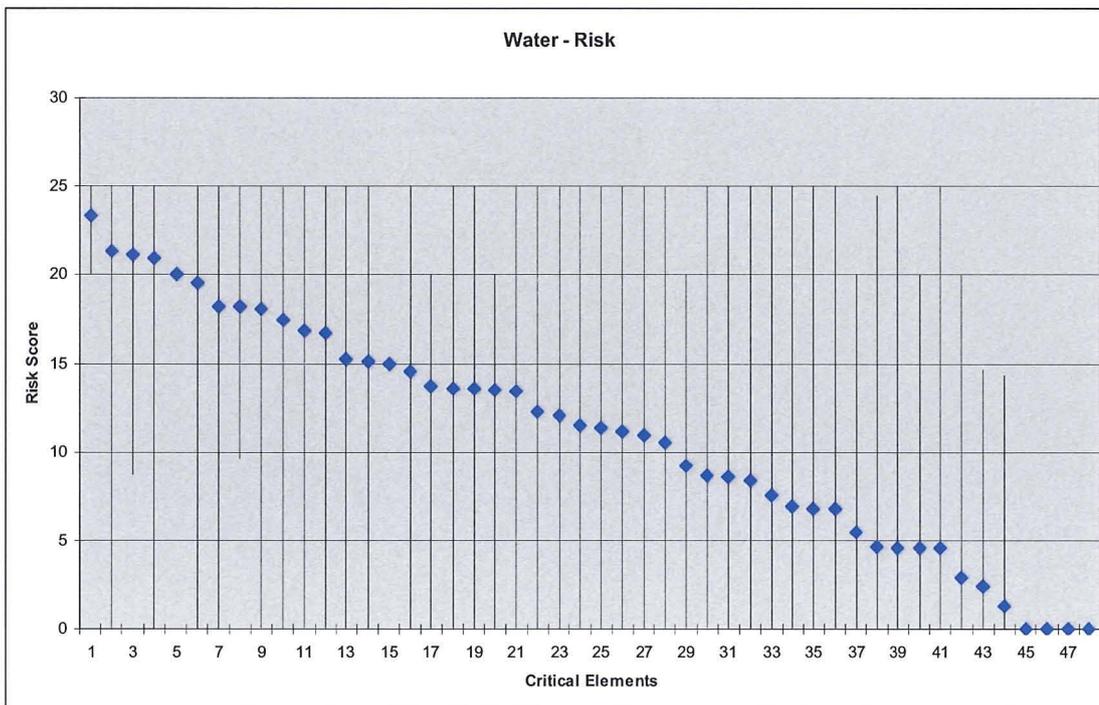
Figure 24 showed the average risk being generated from the water systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red and amber zones:

❖ Emergency deluge showers and eye baths are to be flushed weekly to prevent stagnation of water.

170

❖ A written operational plan, policy and procedures should be formed from a detailed risk assessment, and should be in place for each site. This should comprise of: up-to-date drawings and descriptions of all the supply, storage and distribution systems within those premises, and step by step instructions relating to the operation, maintenance and shut down for each branch of the system. These documents must identify the nominated person (and his deputy) responsible for overall accountability together with details of their training needs. All details contained within these documents must be acceptable to the control of infection team.

❖ These risk assessments should be undertaken at least annually. The risk assessment should be conducted not just for the routine operation or use of the system, but also in unusual circumstances, breakdown, abnormal operation and commissioning.

❖ A contract supervising officer having specialist knowledge, training and experience should be appointed to witness tests and checks under the terms of the contract. Pre-commissioning, commissioning and testing must be carried out upon systems before handover. And all new and/or altered installations must at the time of handover be supplied with installed record drawings and operating/maintenance instructions, and test certificates. Also additional monitoring checks/analysis should be carried out after any repairs or modifications to the system.

❖ All isolating valves should periodically be worked through their full range of travel.

❖ The Trust should ensure that an accurate record of all assets relating to the hot and cold water distribution systems is set up and regularly maintained.

❖ During a period of low ambient temperature, a check should be made relating to the temperatures of the outflow from the hot water supply calorifier to establish that the temperature is above 60 degrees centigrade and that the temperature at the return connection is 55 to 50 degrees centigrade. Also the

most distant draw-off point on the system should be checked to ensure that the temperature reaches a steady state value between 60 and 50 degrees centigrade within one minute of running the water at full flow.

❖ Whenever a calorifier is taken out of service or its flow temperature falls below 45 degrees centigrade for any reason, then it should be pasteurised.

❖ It is essential to check the temperature settings and operation of all water mixing devices routinely, at least half yearly.

❖ Where two or more pressurisation/supply pumps are installed for pressurising systems, each pump must regularly be brought into service at least daily as the main duty or lead pump. Hot water circulating pumps should be switched daily to ensure that any standby or back-up pump is regularly brought into service as the main duty or lead pump.

❖ The maintenance of ice making machines, water coolers and drinks vending machines should be carried out in accordance with the manufacturer's recommendations.

❖ During temporary closure of wards or departments, a procedure for flushing the hot and cold water service systems should be instituted.

❖ In multiple cistern installations a check must be carried out for stagnant water.

Figure 25   The Water average weighting for each critical element.


Water - Weighting
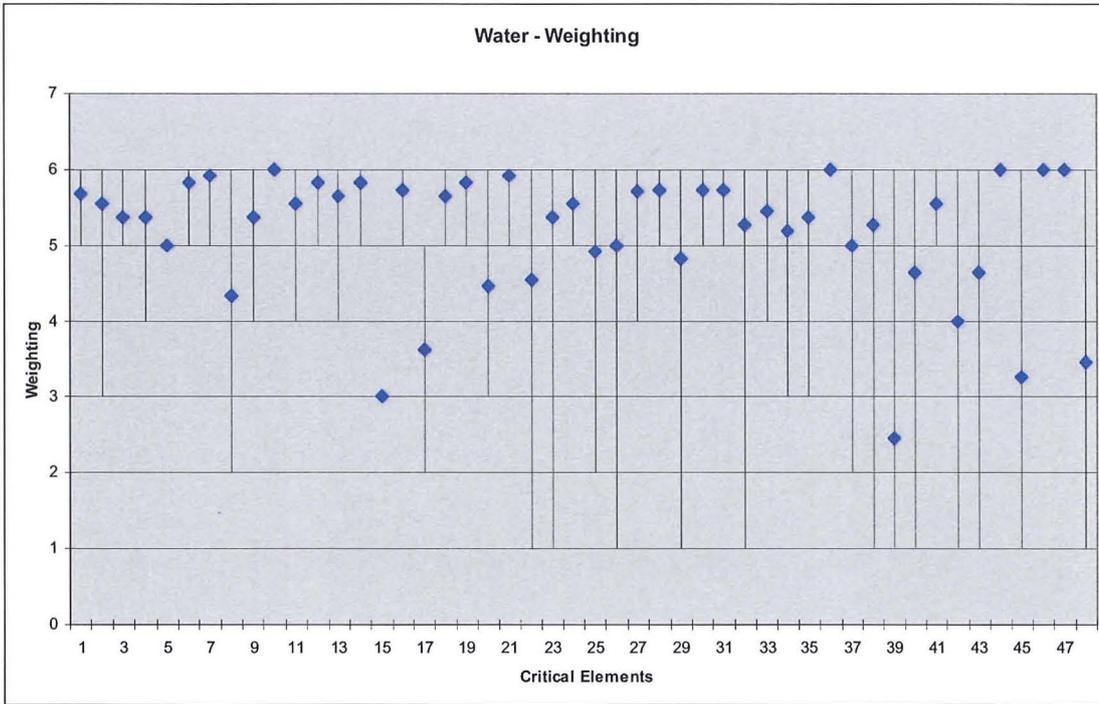
Figure 26   The Water average compliance level for each critical element

Water - Compliance
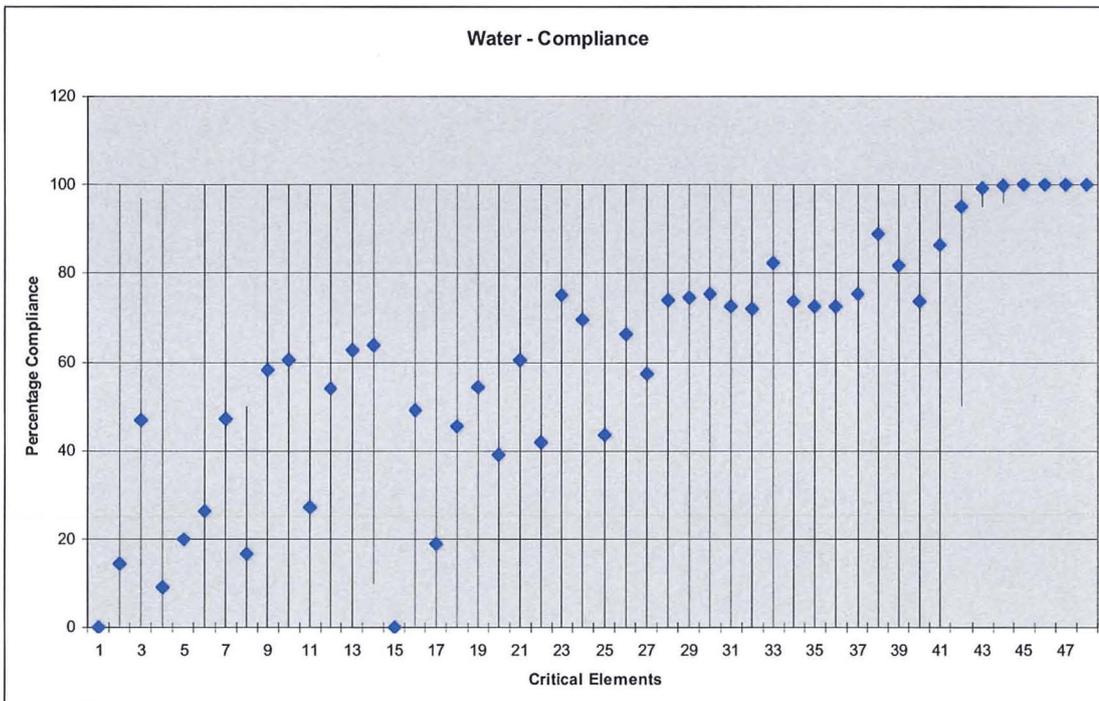
Figure 27   The lifts average risk score for each critical element.

Lifts - Risk
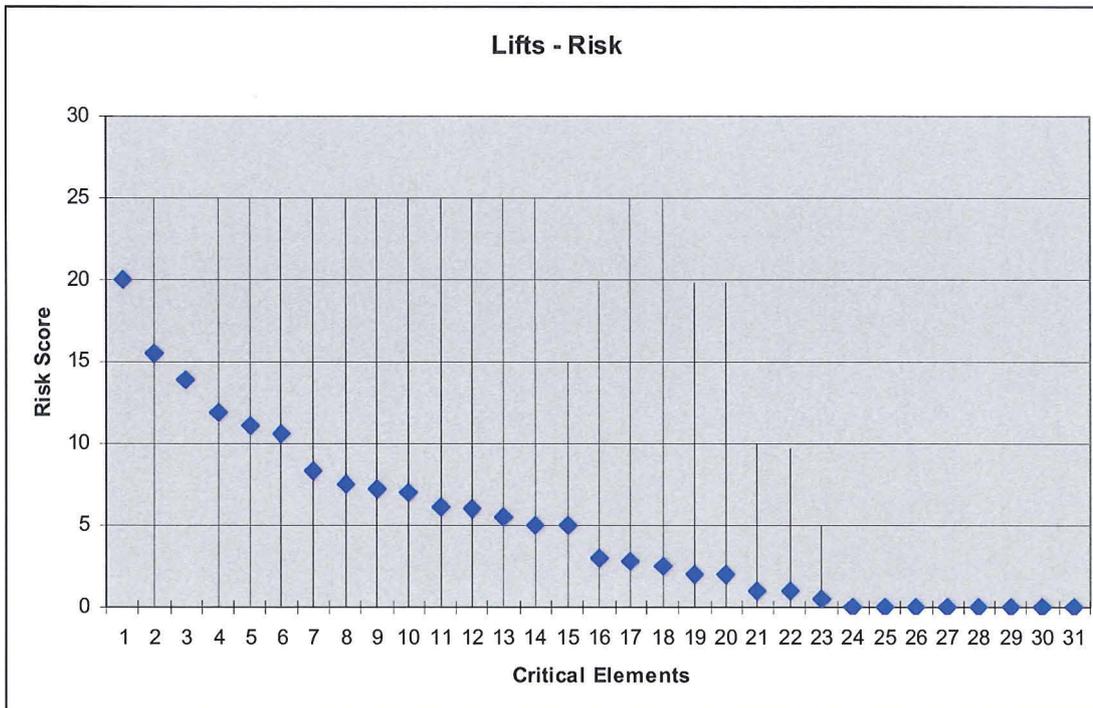
Figure 27 shows the average risk being generated from the lift systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red or amber zones:-

❖ All passenger and bed/passenger lifts should be fitted with an emergency intercommunication point (two way).

❖ Record of inspection and servicing procedures should be maintained.

❖ Trust staff members should be trained in methods of emergency hand winding of traction lifts and emergency hand lowering of hydraulic fittings. Records of training and refresher dates should be maintained.

❖ A member of the hospital maintenance staff should regularly check the emergency lighting and intercoms systems.

Figure 28   The Lifts average weighting for each critical element
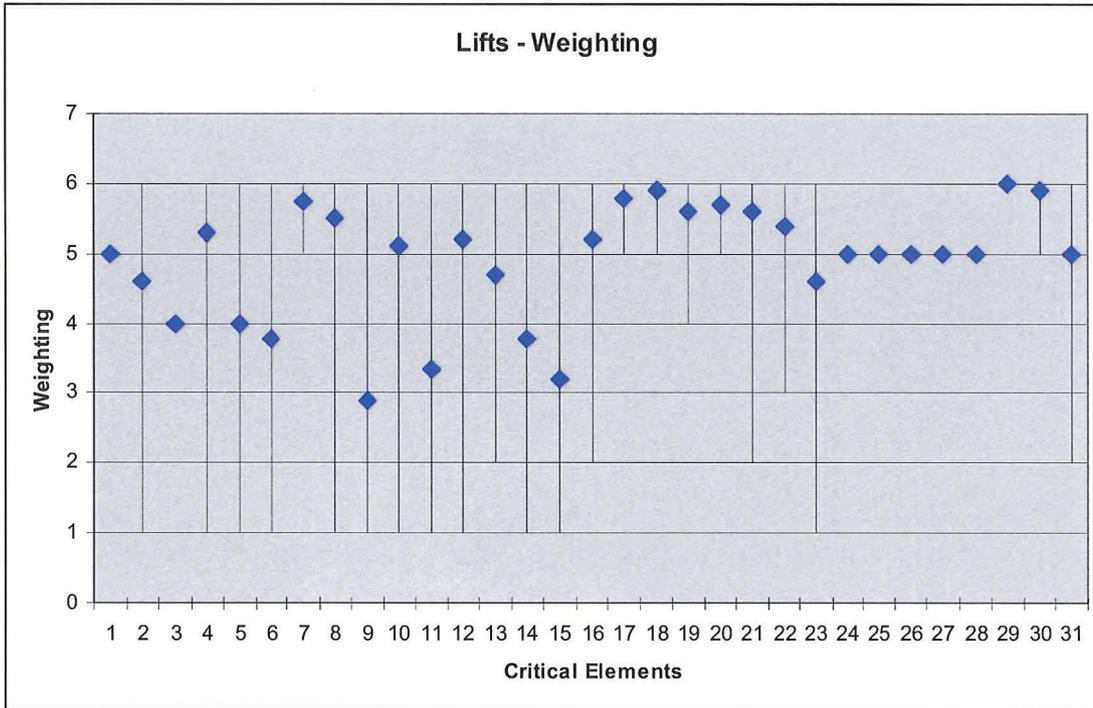
Figure 29   The lifts average compliance level for each critical element

175
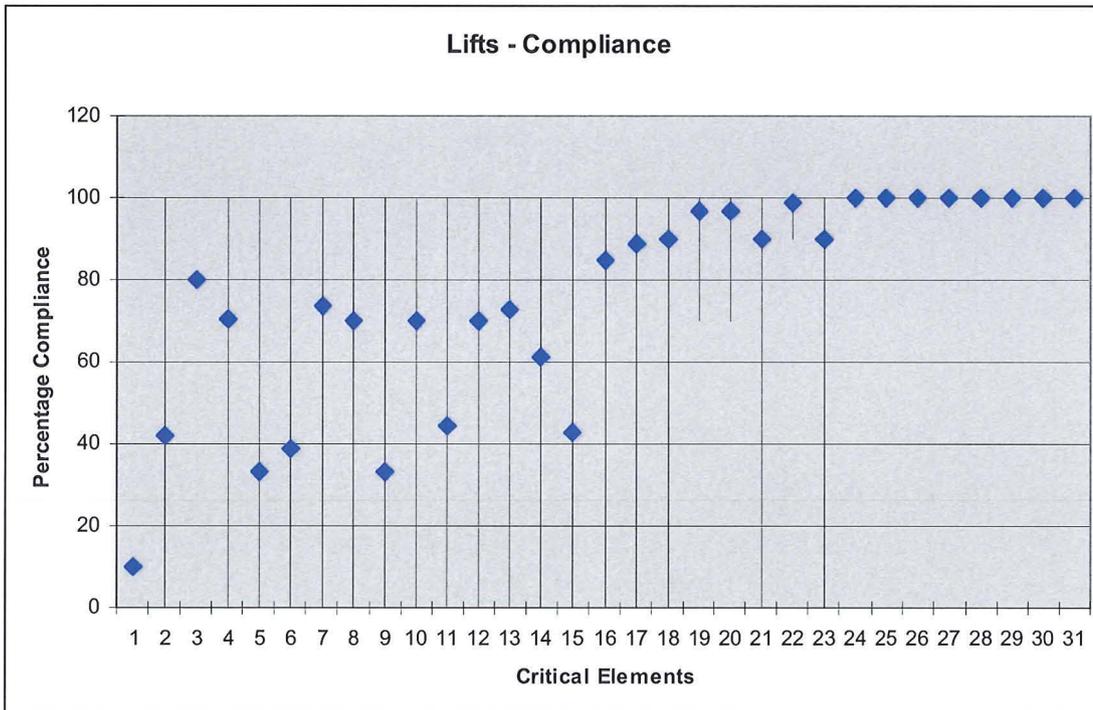
**Figure 30   The Ventilation Plant average risk score for each critical element.**
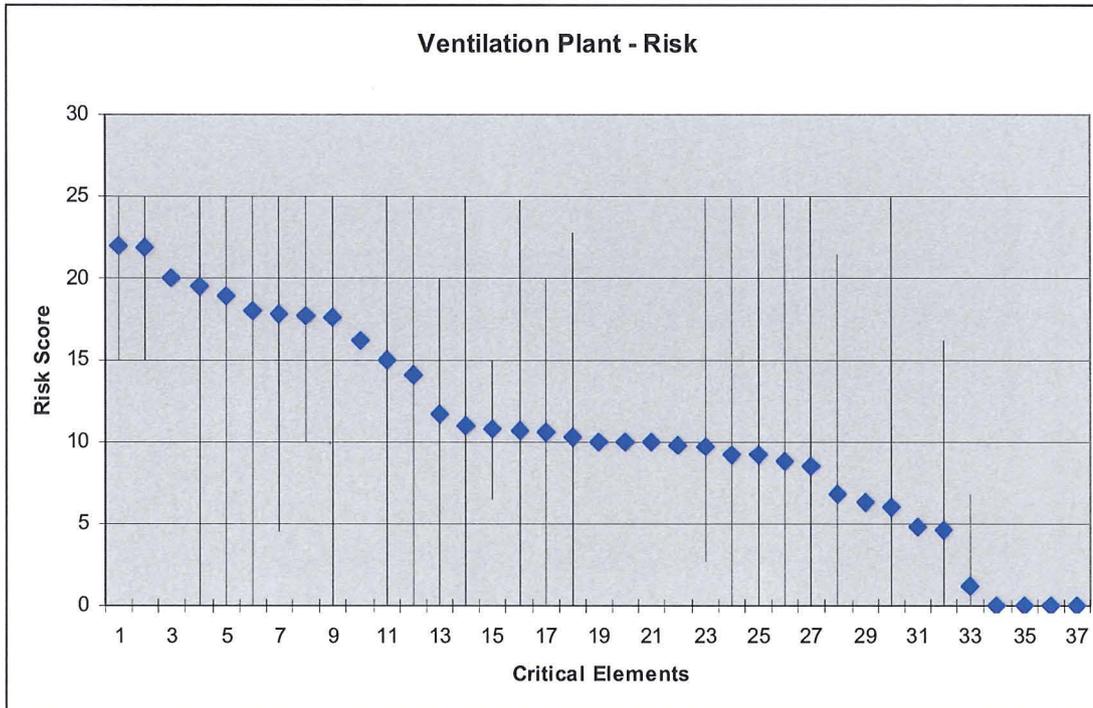
Figure 30 shows the average risk being generated from the ventilation systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red or amber zones:-

❖ An asset register should exist that lists all ventilation and air conditioning plants, together with the relative areas been served from each plant. The maintenance schemes should consist of sufficient measurements of the plant's performance to determine its continuing ability to perform to its designated function. Information for each plant should be available in relation to the filter grades and the dirty and clean pressure drops.

❖ Regular tests, at intervals agreed with the local fire prevention officer, will need to be carried out in order to demonstrate the continuing efficiency of the fire detection and containment systems. These tests include the testing of fire dampers. The location of these fire dampers should be marked on the ductwork, and a visible point on the fabric if the ductwork is concealed.

176

- ❖ All ductwork should be marked in order to identify its purpose and direction of airflow, (fume cupboard and safety cabinet extract ducting).

- ❖ The maintenance schemes should disinfect all sections that are known to become damp in normal use, with a record being kept of who carried out the disinfection procedure, when it was completed and the method used.

- ❖ Records of dynamic tests measuring the volumetric airflow at randomly selected supply and extract terminal devices: together with records comparing these results with original commissioning information.

- ❖ All ductwork should be marked in order to identify its purpose and direction of airflow, fume cupboards and safety cabinet extract ducting.

- ❖ Plant located outside roofs should be provided with designated walkways to protect the roof surface.

Figure 31   The Ventilation Plant average weighting for each critical element.
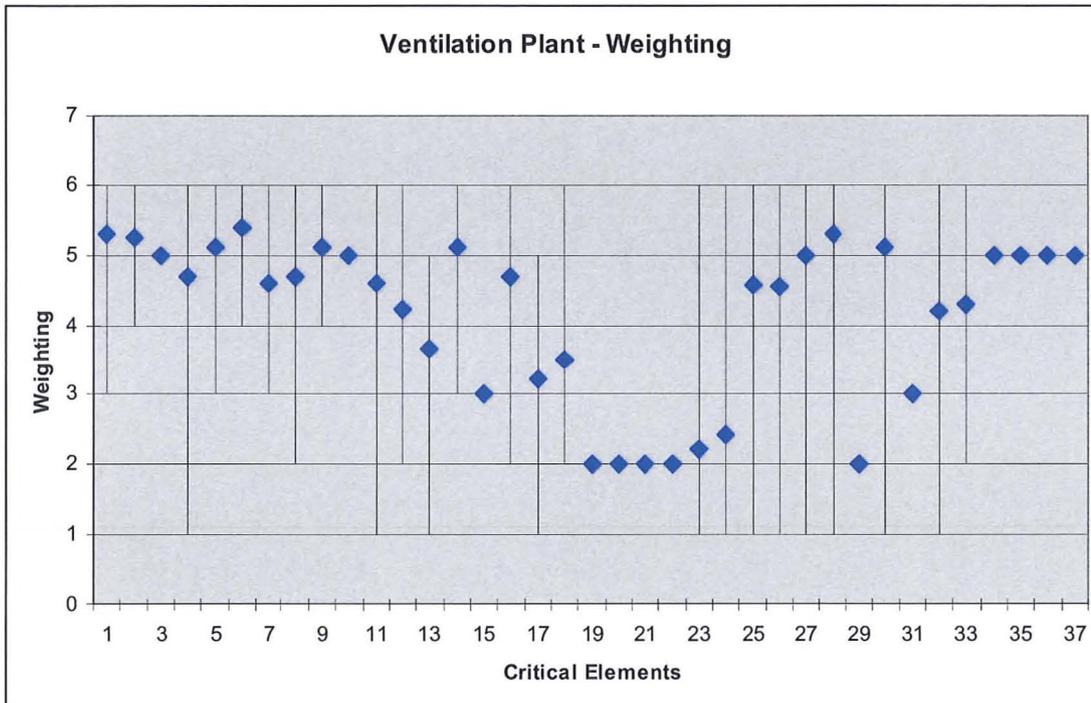
Figure 32   *The Ventilation Plant average compliance level for each critical*
element
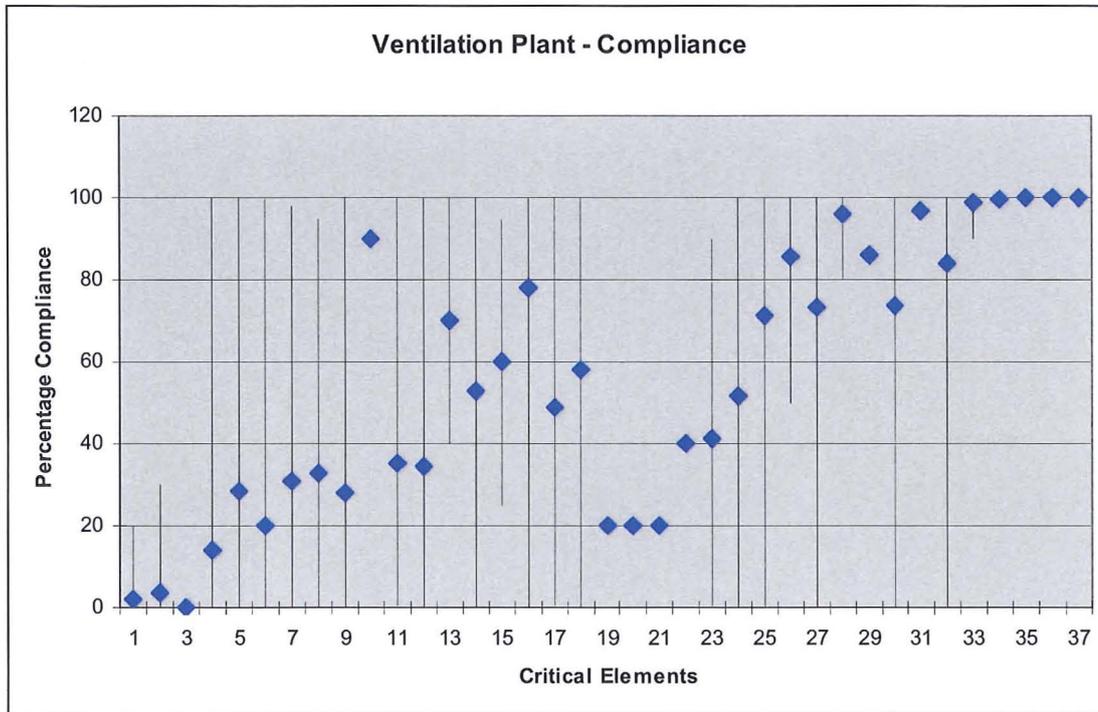
*Figure 33*   *The BMS average risk score for each critical element.*
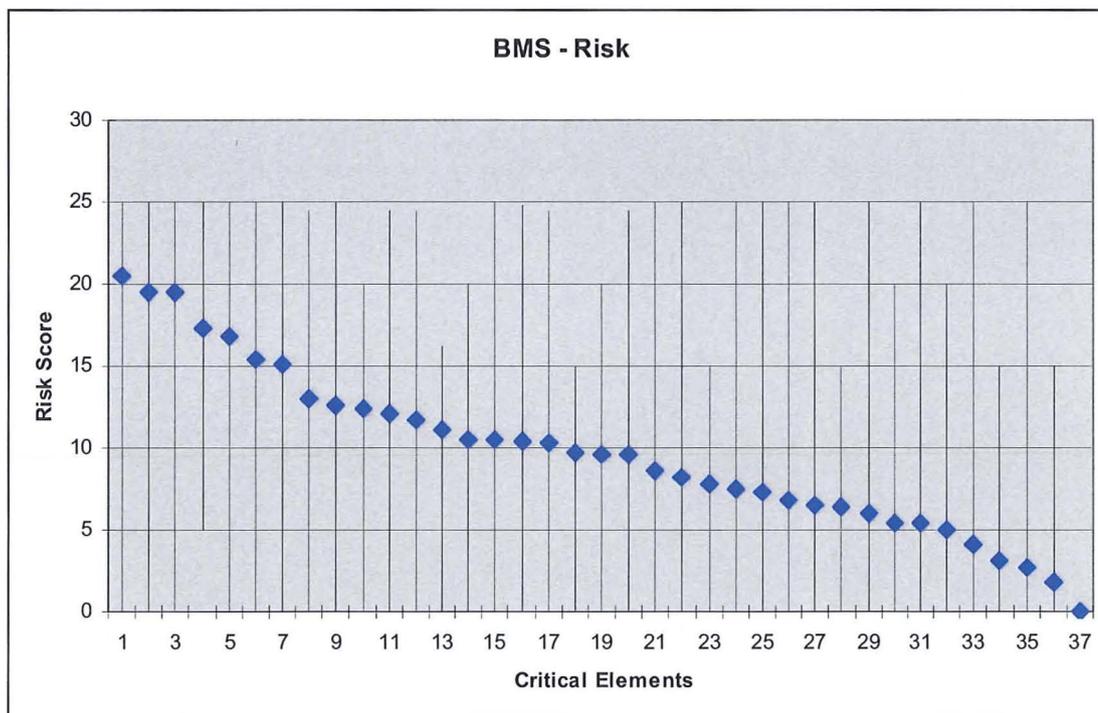
Figure 33 shows the average level of risk being generated from the building management systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red and amber zones:-

* ❖ The operational procedures for the BMS should detail the main and foreseeable potential problems and hazards that can arise from failing to follow the agreed operating, monitoring and maintenance procedures, as well as the possible legal consequences

* ❖ Arrangements for dealing with BMS alarms should be graded according to severity, and agreed with a Trust multi-disciplinary group e.g. cross infection officer, microbiologist etc.

* ❖ Performance monitoring, risk assessments, performance reports and auditing of the system should be undertaken on a periodic basis.

* ❖ Written operational procedures should contain: -
  * ▪ Periodic training regimes as the BMS system and staff change.
  * ▪ Training records for each BMS operative that clearly shows that they have been specifically trained.
  * ▪ A user brief comprising a description of a BMS, plant under BMS control and the intended mode of operation.

* ❖ The following records should be available: -
  * ▪ Plant diagrams showing locations of field devices
  * ▪ Precise design requirements with regard to physical parameters measured and controlled by the BMS
  * ▪ Commissioning manuals listing the results of commissioning tests.
  * ▪ Record drawings
  * ▪ Schematics

Figure 34   *The BMS average weighting for each critical element.*

BMS - Weighting
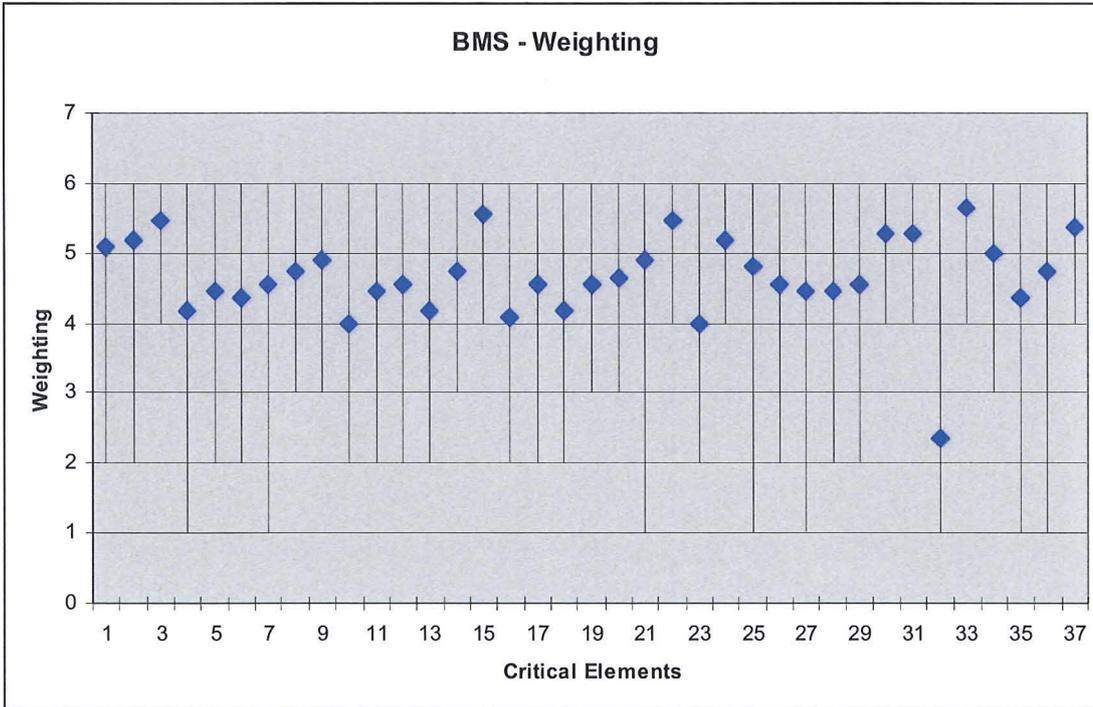
Figure 35   *The BMS average compliance level for each critical element*

BMS - Compliance

180
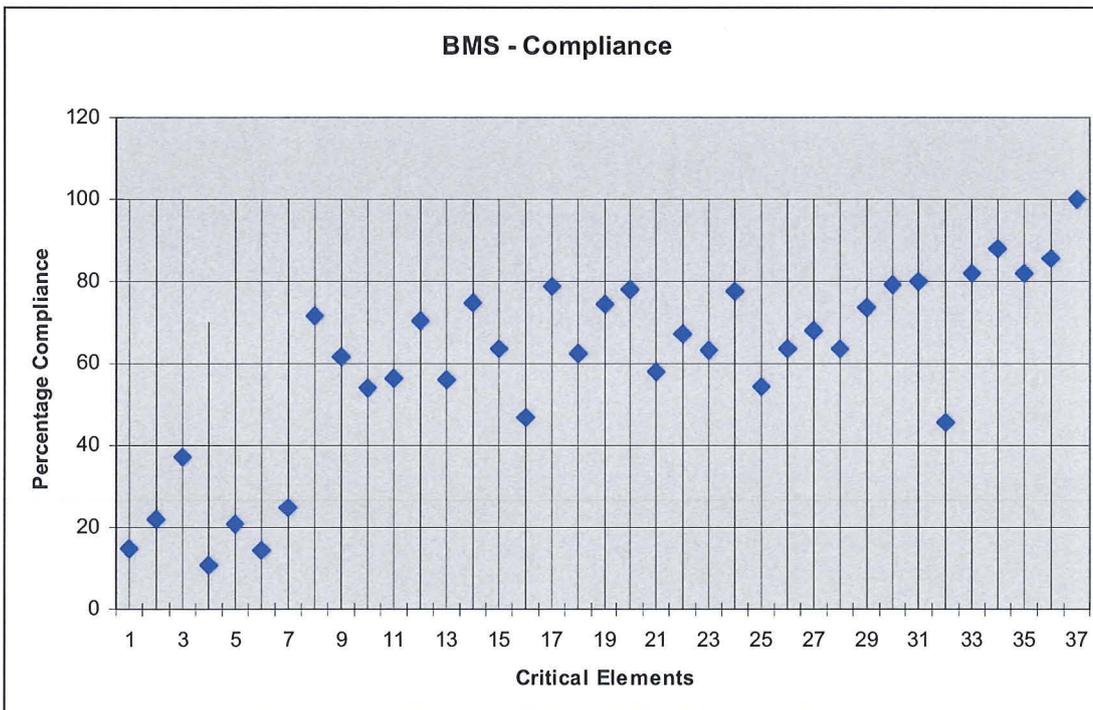
**Figure 36   The Pressure Systems average risk score for each critical element.**

Pressure Systems - Risk
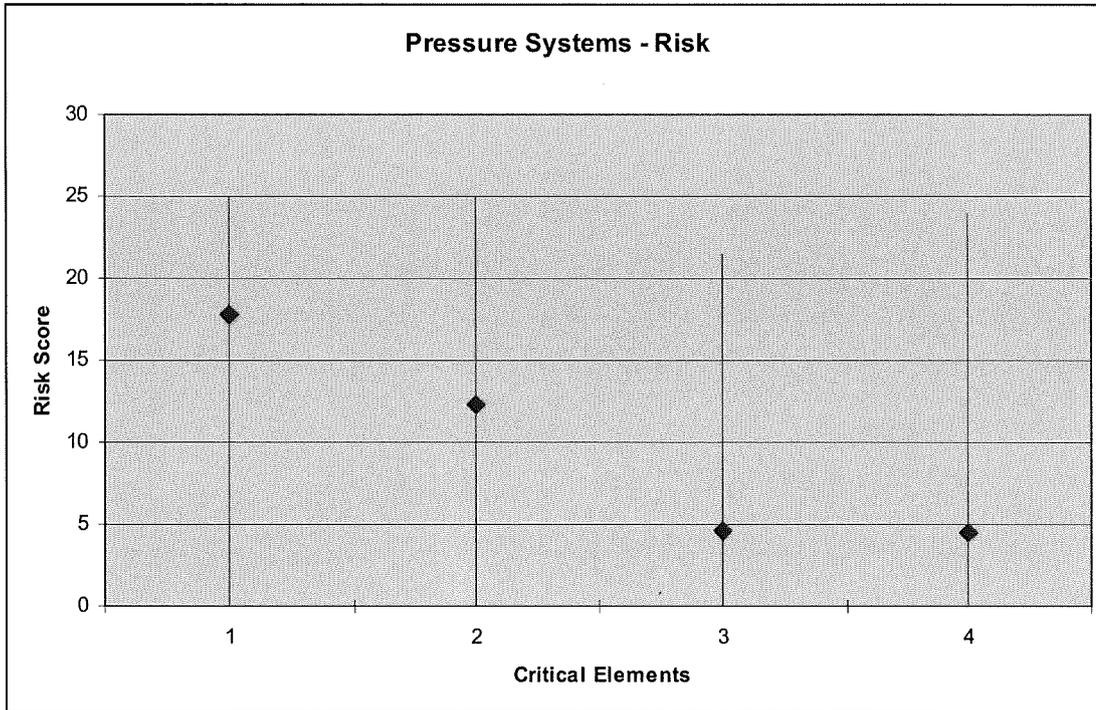
Figure 36 shows the average risk being generated from the pressure systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red or amber zones:-

❖ The Trust shall provide instructions for the operation of the plant and equipment, and shall include for the safe operation of the system and the action to be taken in the event of any emergency. The Trust shall also be responsible for ensuring that the system is properly maintained in good repair, so as to prevent danger.

Figure 37   *The Pressure Systems average weighting for each critical element.*

Pressure Systems - Weighting
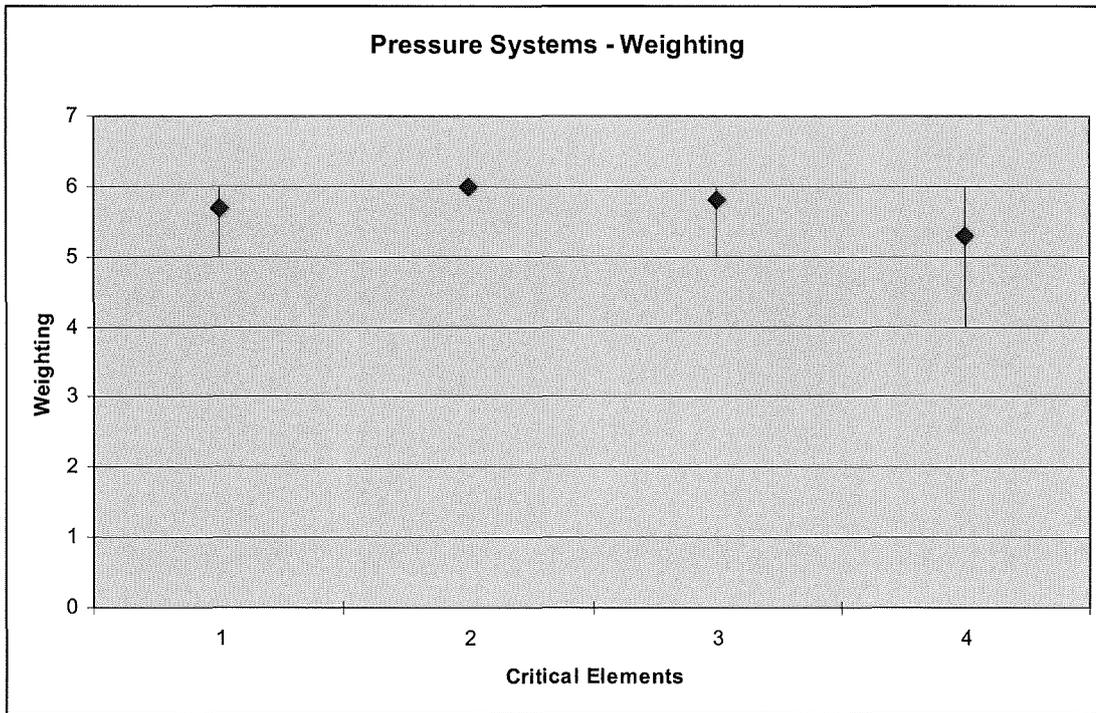
Figure 38   *The Pressure Systems average compliance level for each critical*
element

Pressure Systems - Compliance

182
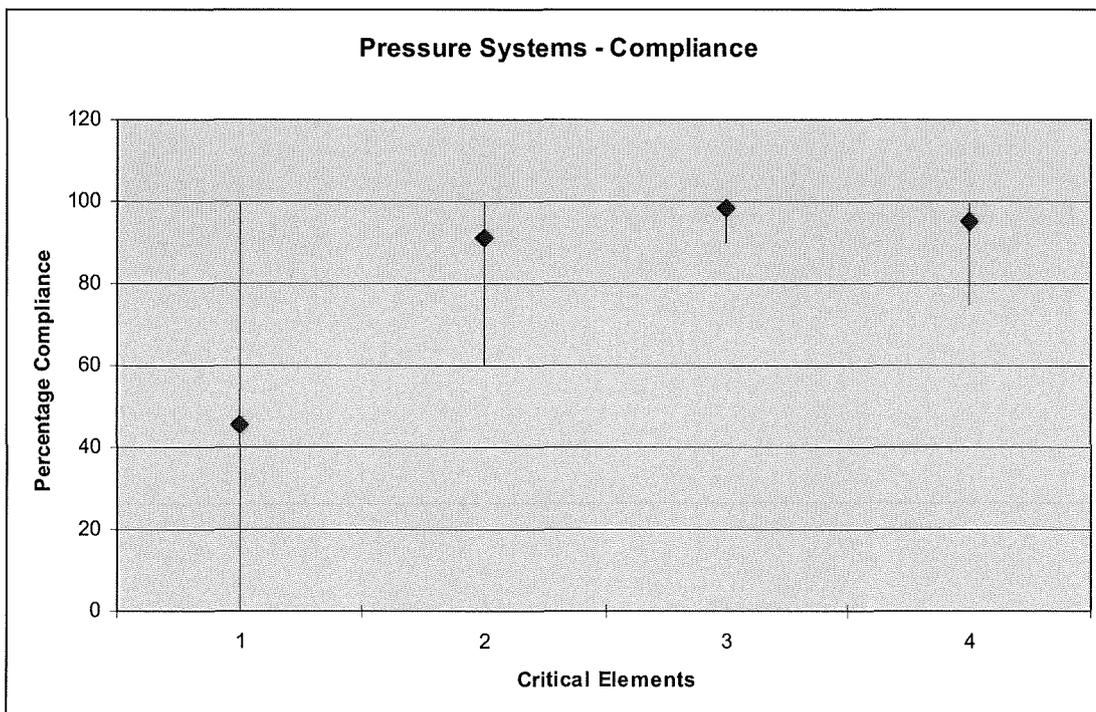
Figure 39   The Fire average risk score for each critical element.
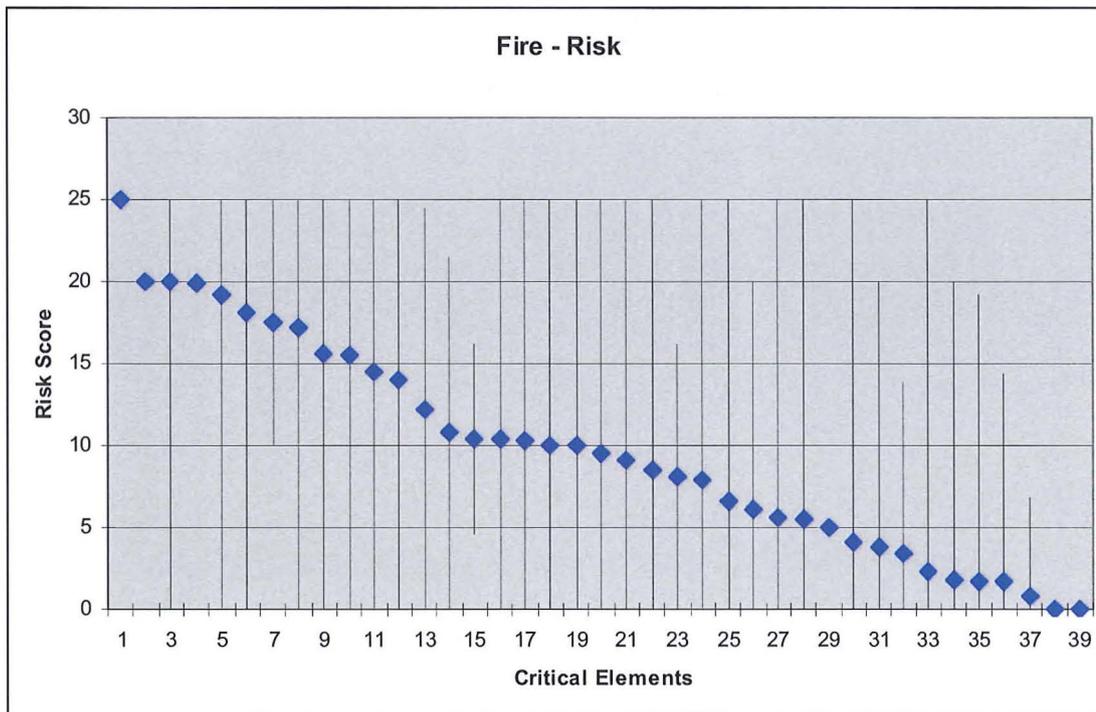
Fire - Risk

Figure 39 shows the average risk being generated from the fire systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red or amber zones:-

 ❖ A risk assessment must be available that assesses the risk of maintaining these hoses against the risk associated with Legionella. This risk assessment must be undertaken by a multi-disciplinary group.

 ❖ A risk assessment should be undertaken to determine the frequency of planned preventative maintenance for each set of fire doors.

 ❖ The minimum acceptable levels of fire precautions should be determined from a full fire risk assessment undertaken using the nationalised guidance relating to fire. These risk assessments should be used to develop an emergency evacuation plan.

❖ The emergency lighting system must be periodically tested. The testing programme must make reference to the manufacture's requirements, in particular, the shelf life of the battery backup cells.

❖ The Trust must maintain an up-to-date set of drawings, which indicate: alarm and detection systems, means of escape, emergency lighting, containment, first aid fire fighting equipment and fire brigade access.

❖ If an automatically started generator is used as part of the standby supply, then it should be started up once each month by the simulation of a failure of the normal power supply, and a check should be conducted to ensure that it energises the fire alarm system.

❖

A visual inspection should be made to check whether structural or occupancy changes have affected the requirements for manual call points, detectors and soundness.

❖ All dry and wet fire mains should be checked for correct operation on a periodic basis.

❖ Each and every fire detector should be periodically checked for correct operation in accordance with the manufacture's recommendations.

❖ Where lightning protection systems are provided they shall be subject to a yearly planned preventive maintenance inspection and report.

Figure 40   The Fire average weighting for each critical element.

Fire - Weighting
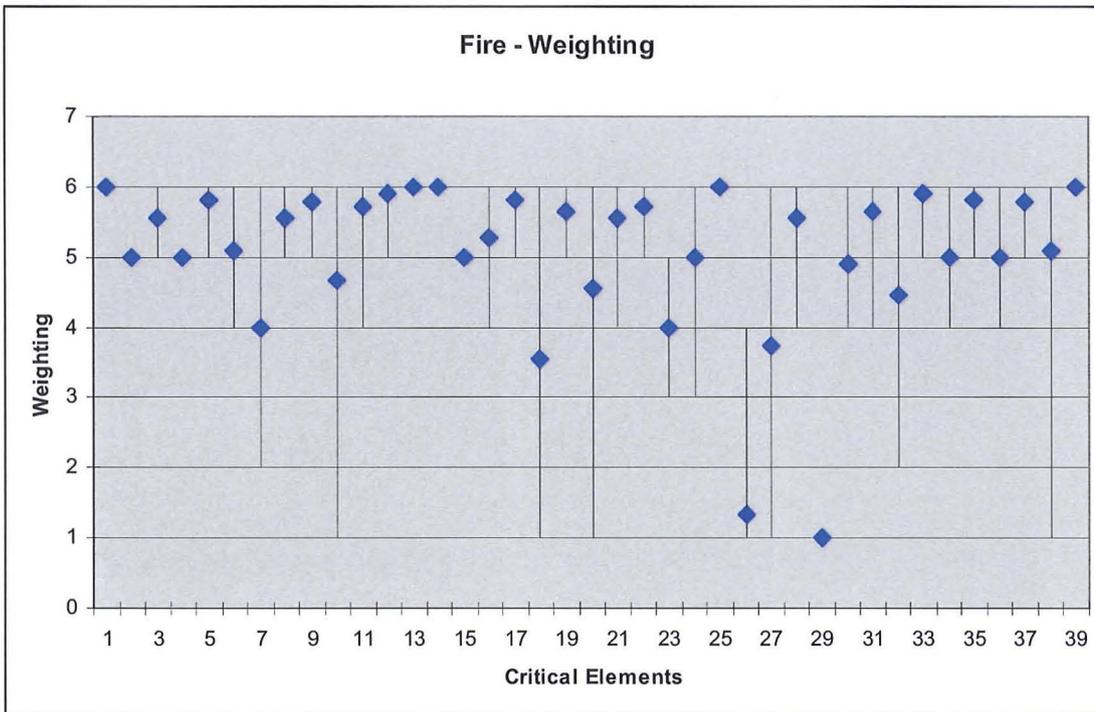
Figure 41   The Fire average compliance level for each critical element

Fire - Compliance

185
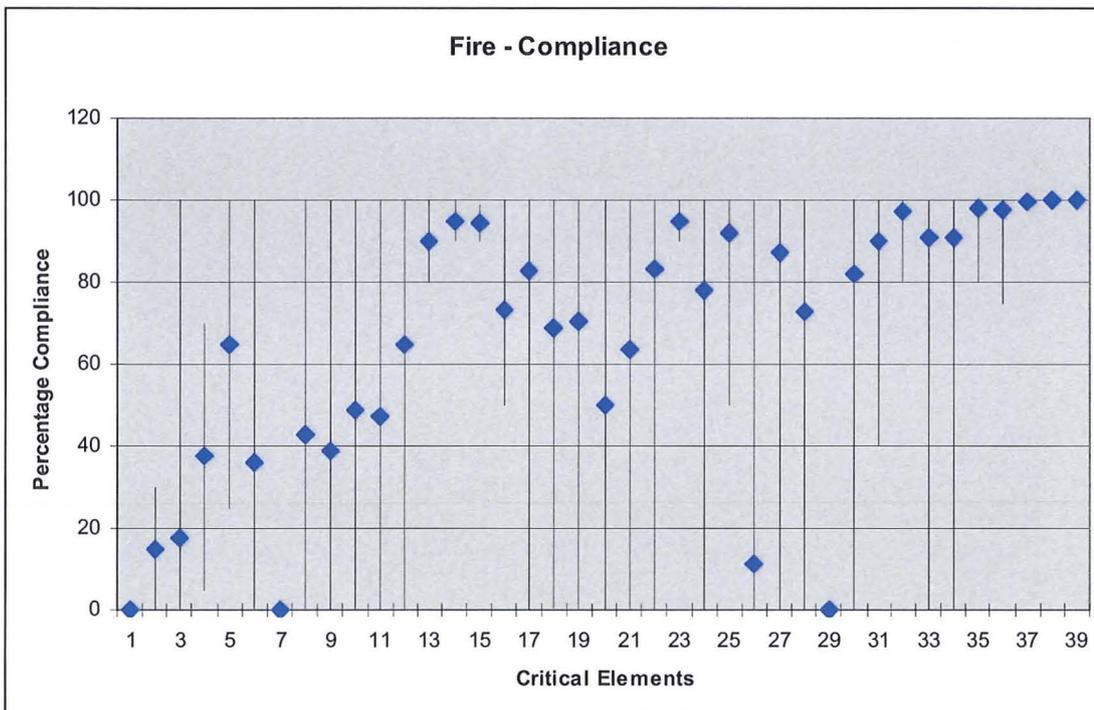
Figure 42   *The Electricity (LV) average risk score for each critical element.*
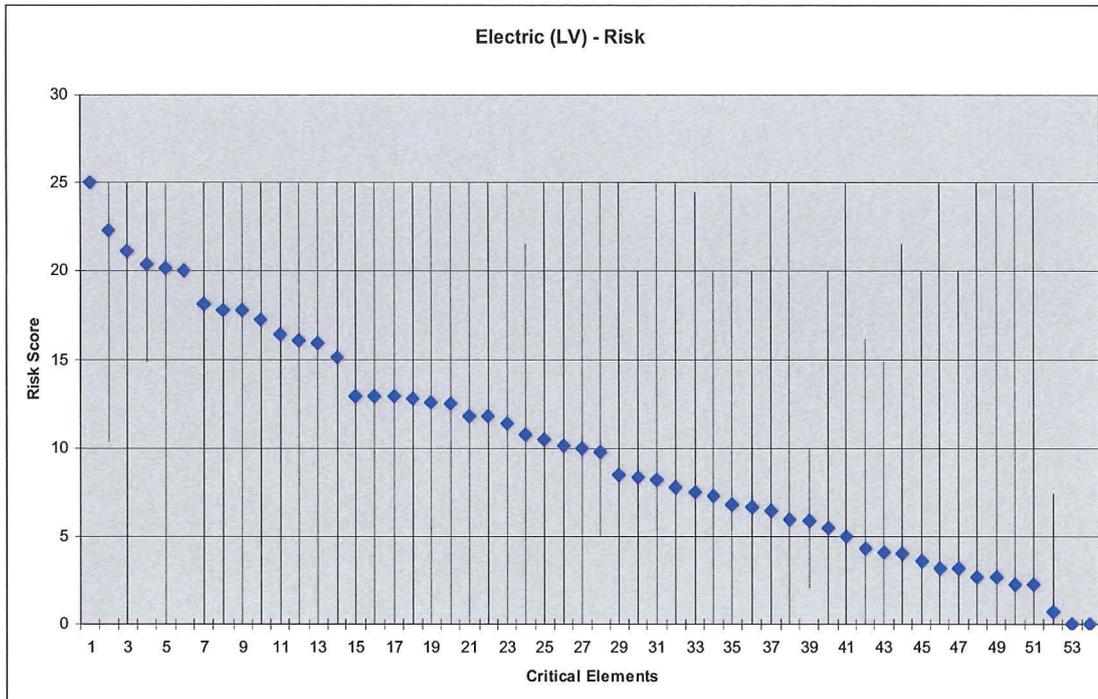

Figure 42 shows the average risk being generated from the electrical systems within the Trusts analysed. Listed below (in no order of importance) is a summary of the failures that have attracted an average risk score within the red or amber zones:-

❖ Management must have a clearly defined electrical safety policy and programme for the operation and servicing of their low voltage systems and equipment. All authorised or competent persons concerned with operating maintenance or work on the low voltage systems of the managed premises shall be provided with a personal copy of management electrical safety rules.

❖ The operational policy should contain a schedule of possible emergency incidents, with remedial operational procedures, which may cause a loss of normal electrical supplies.

❖ Essential power supply facilities for existing health care premises should be periodically reassessed. Upgrading of emergency power should be considered to maintain essential clinical and surgical life support facilities.

186

❖ Simulation of loss of mains supply should be initiated once per month. It is recommended that an emergency exercise be undertaken and independently adjudicated to test the contingency plans for alternative emergency generating plant.

❖ All fixed low voltage electrical systems shall be periodically inspected and tested. Therefore records shall be presented for every system, together with a pre-agreed programme of inspection routines. The frequency of inspections and testing shall not exceed 5 years for normal installations and 3 years for industrial areas e.g. kitchens and workshops.

❖ All fixed or portable items of electrical equipment which are connected to the electrical supply should be subject to registration for approved use by the authorised person responsible for the electrical supply and distribution system.

❖ Distribution boards controlling final outlet circuits shall be accurately identified by labelling of each board and circuit.

❖ Risk treatment plans should be developed that show how the Trust is correcting the deficiencies that are required to be addressed.

❖ A system or installation accepted from a contractor must be accompanied with handover documentation including test certificates.

❖ The operating theatre emergency lights must be tested on full load for a period of not less than 3 hours.

❖ All portable electrical equipment shall be subject to inspection and testing procedures.

Figure 43   The Electricity (LV) average weighting for each critical element.
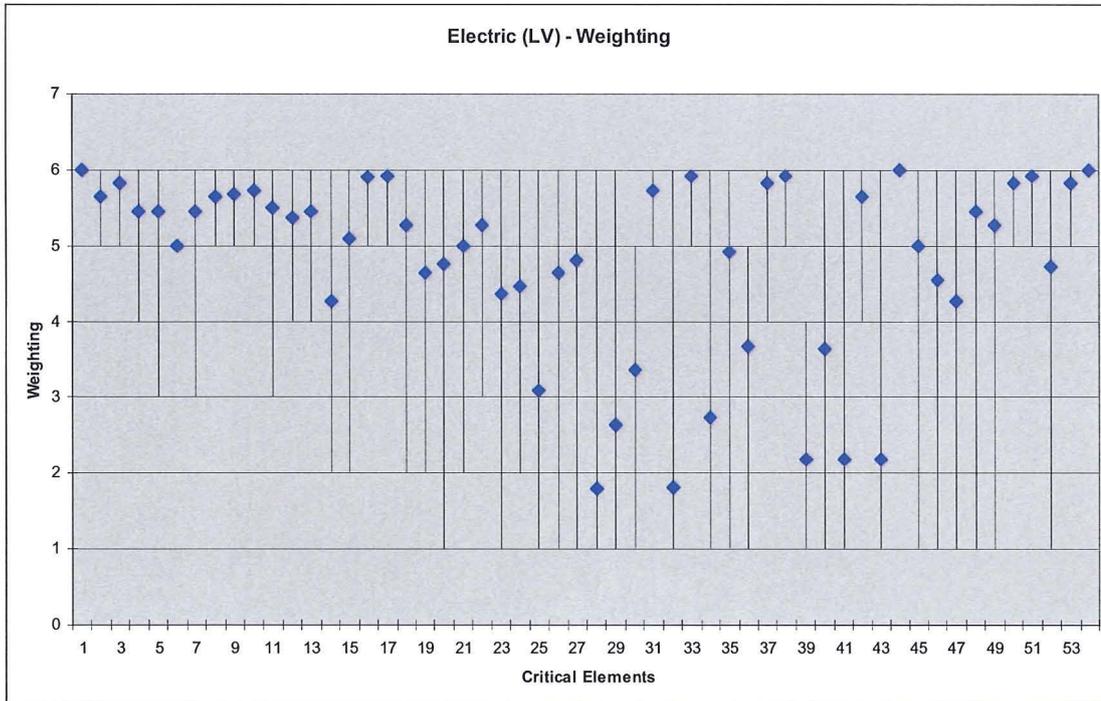
Figure 44   The Electricity (LV) average compliance level for each critical element

188
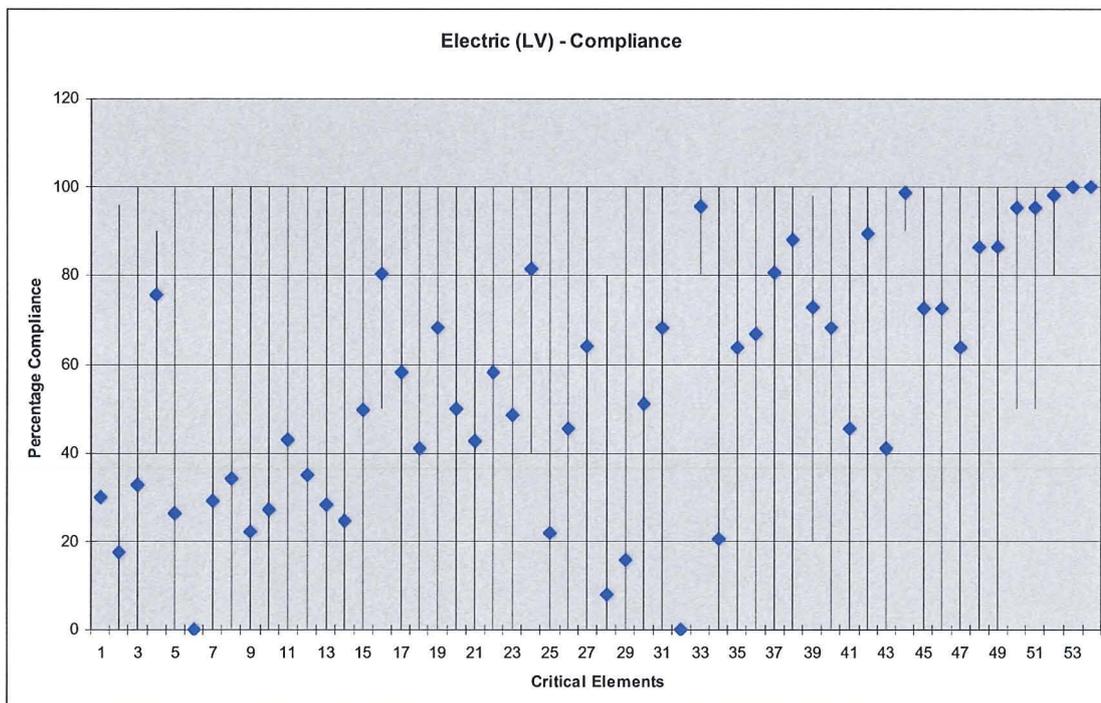
4.2.2 Brief discussion of the results from the first research objective

The approach employed within this first research objective was very similar to the work of Ramanujam and Goodman (2003) which asked similar questions when analysing organisations against pre-specified routines and generally accepted norms which uncovered latent gaps within the managerial systems. From the results of this research it has been shown that all of the ten Trusts analysed have been at risk from unacceptable and intolerable unknown latent conditions within their management systems. The results showed that all of the Trusts analysed generated risks in the red and amber zones. From an examination of the critical elements within these red and amber zones it is apparent that these risks were caused by long term failures of the managerial systems, and not by momentary 'errors of judgement' or 'lapses of memory'. Following the development of each profile large resources were redeployed by each of the participating Trusts, to reduce their specific identified risks. Thus, it is argued that the Trusts have accepted the use of the overall methodology of the research instrument developed to identify latent conditions within their managerial systems and to assess a risk score. Also a trait common to all profiles developed was that the departments were not ISO 9000:2000 compliant. Thus the main conclusions drawn were that:

- ❖ **Methodologies** used by the process have been accepted by the Trusts involved.

- ❖ **Current** monitoring systems employed by the NHS are woefully inadequate. The model has uncovered numerous previously unknown latent high risk conditions.

- ❖ **Without** the alignment of a Trust's managerial systems with recognised quality standards, then the problems and risks of latent high risk conditions will persist, and the NHS corporate governance initiative will be compromised.

4.3   Second research objective

This objective sought to assess the likely resource needs (manpower, outsource costs and timeframes) to close the identified gaps within the managerial systems uncovered.

189

The resource needs required for each Trust were not consistent between Trusts. This was to be expected as each organisation has operated independently. What was common was that each Trust reallocated their existing limited resources to where they would 'do most good', based on the findings. From this it can be concluded that they collectively accepted the methodology and assessments.

4.4   Third research objective

This was by far the most complex in terms of methodology, project management, data collection/analysis and interpretation. It sought to answer the question, "What directly employed artisan revenue resources do we need to keep the gaps closed?" In the researcher's personal experience it has been a question that has been asked throughout the past three decades, without an answer coming fourth – until now.

During the course of this information gathering period it became obvious to the researcher that, each Trust controlled and monitored this basic managerial information differently. This variation ranged from virtually no apparent control through to practicing close monitoring procedures. In all cases the researcher had to question the accuracy of some of the information and subsequently had to require the organisation to justify some figures. These historical base figures needed for the simulation consisted of the following three data sets, all containing week-by-week totals over the last one to three years for the mechanical, electrical and semi-skilled trade groups/site. The time frame of between one and three years of the historical data used was dependent on the availability and accuracy of the records.

a. Available man hours

b. Defect hours

c. planned preventative maintenance schedule

a.   Available man hours

The available man hours were the actual hours available to undertake the defect and planned preventative maintenance tasks. These were the hours remaining after their diverted time was removed. The diverted time was a phrase used within the (now) old NHS Labour Management System, and consisted of a number of codes that were agreed locally, and used to record unproductive time such as holidays, training, trade

union meetings, etc. In all cases the researcher had to discuss and agree the appropriate diverted time for each trade group/site.

b. Defect hours

This data was readily available from each Trust's computerised work information system. The data recorded all of the directly employed labour artisan man hours that had been expended undertaking the day-to-day maintenance duties, which were generated from the various wards and departments.

In all cases it was agreed that the artisan staff workload consisted (in simple terms) of: reactive and planned preventative maintenance tasks, with a small amount of low cost 'new works', where the term 'new works' related to the adaptation and/or alteration of the existing infrastructure and contained a portion of 'Backlog' maintenance activities. It was agreed that that the man hours expended undertaking these 'new works' would be removed from the simulation process, as these were not classified as day-to-day maintenance. Thus only the true day-to-day tasks of the reactive maintenance needs and the planned preventative maintenance schedule were to be subjected to the simulation process.

c. planned preventative maintenance schedule

In all cases the planned preventative maintenance regime, associated frequencies and estimates of man hours to complete each planned preventative maintenance task, had to be either partially or completely rewritten before the simulation could commence. As it was found that the systems had not been updated to keep pace with the changes to the national guidance, and/or local plant and equipment and/or advances in technology. This was undertaken initially by the estates officers by identifying each task required, and referencing the relevant national guidance and local specific requirements. This resulted in an 'assessment of need', together with descriptors relating to the assessed impact should an item fail whilst in service; and with an assessed probability of failure with respect to time should a planned preventative maintenance task fail to be completed for whatever reason. It was this revised planned preventative maintenance schedule that was used within the simulation and not any historical data relating to periodic maintenance. Following the design of the planned preventative maintenance regime, the

191

relevant tasks were balanced throughout the year to form an even workload on a week-by-week basis within each trade group.

In order to initiate the simulation the impact and probability profile assessments were initially assessed by the engineering officers within the planned preventative maintenance workshops. These preliminary judgments allowed the simulation to undergo a series of iterations, fine tuning of the data and output format, in preparation for presentation and discussion with the clinical and medical professions. In all but a few cases the original values assigned by the engineering officers were accepted. Thus the final maintenance simulation outputs were formed from the active involvement of a multi-professional team, as recommended by the NHS guidance Estatecode. But before the simulations were run, the researcher considered it necessary to rewrite the original simulation software, for the reasons discussed below.

4.4.1   The simulation programs re-design

The software program for this third research objective was initially designed using Microsoft Visual Basic for Applications language. The program took approximately 1½ hours to complete one simulation run for a single trade group for one Trust site, when operated on a laptop computer with average processing power. The speed of the program run time was increased by employing the use of a reasonably high powered desk top machine, but the researcher considered that it was still too slow, therefore an alternative software compiler was sought.

The chosen software was FORTRAN 2003, one of the oldest high level programming languages. It was introduced in 1957 and used in the scientific, numerical and engineering fields. Through the years the program has been the subject of a number of revisions to keep pace with computer technology and to be competitive with other programming languages. Due to the longevity of the language there is a wealth of information and experience readily available, along with a demonstrable track record of its use. This with the programs syntax being similar to the Microsoft Visual Basic for Applications language made it this researcher's obvious choice.

Thus the original Microsoft Visual Basic for Applications program was rewritten into FORTRAN 2003. The resulting increase in the speed of a single simulation moved from

1½ hours to just a few seconds, when operated on the same laptop machine. This increase in speed allowed the original program to be enhanced to give a greater detail of analysis. This in turn promoted additional understanding of the operational maintenance activity and the associated generation of risk, and allowed the simulation to be expanded from a 10 year to a 20 year period.

The base data for the simulation was provided by the estates departments in Microsoft Excel format. This was imported into FORTRAN as a series of arrays. Prior to any simulation run each base data array was viewed on the screen for the researcher to confirm (or otherwise) that the transfer of the data had been completed successfully. Once the researcher was confident that the transferred data was correct the program was allowed to create the following arrays:

Tests 1 & 2

These two arrays produced a series of step functions for the available man hours which were used to ensure that the program responded correctly to pre-determined extreme levels.

Available Hours

This array calculated on a week-by-week basis the time that was available for the completion of the planned preventative maintenance tasks by subtracting the defect hours from the available man hours.

Sickness/vacancy

The effects of long term sickness were explored from these three arrays by assuming that the available hours were reduced by one WTE for three, six and nine months respectfully, starting at week 420 of the 1040 week (20 yr) simulation in each case. The term WTE is commonly used within the NHS to describe the available man-hours of one operative to undertake one week's work. Currently an operative is employed for 37.5 hours per week, but this figure is reduced by holiday entitlement, sickness, and other diverted time.

The same effect was judged by the researcher to be experienced when a Trust deliberately holds open a vacancy for a period of time in an attempt to balance the revenue budgets.

Impact assessments

As previously discussed the figures within the NHS criticality grid are unsuitable for use within computations as they are ordinate scales. Therefore this array converted the Trust's response to the descriptors into a dimensionless utility function via the Analytical Hierarchal Process (Saaty 2001).

Probability assessments

Similarly this array was formed from a conversion of the probability assessments into a dimensionless utility function.

Rate of Change

A multi-dimensional array was formed that calculated the rate of change of the probability of failure between the probability utility function set points using an interpolation technique.

Following these initial calculations the program simulation was activated. Initially two test simulations were completed to confirm the correct program responses to extreme values. Although these tests were originally used within the development part of the software, the researcher considered it prudent to repeat the tests for each operation of the program as a protection against any unknown error that could possibly develop within the source code.

To test that the model could perform to basic physical realities the model was exposed to extreme unrealistic step values and the outputs recorded. Again to aid the discussion (as in the previous chapter) examples of the outputs created for Trust X are shown The first test simulation presented the existing available man hours as a series of successive step functions. The simulated output for test 1 that was developed for the electrical trade group in site H, is given in Figure 45 below. The graph shows the total assessed risk generated within the green, amber and red zones (G, A & R), those risks lying within the amber and red zones (A & R) and finally the risks within the red zone (R).

Figure 45 below shows that for an excessive available man hour input between weeks 1 and 104 the resulting profile is equivalent to the residual risk, i.e. the minimum risk possible. Between the weeks 105 and 299 the available man hours were reduced to zero via a step function. This caused the resulting assessed risks to rise from their residual values towards the maximum. The rate of increase with respect to time differed for each planned preventative maintenance task, and was dependent on their respective assessments of impact and their individual probability profiles

Figure 45   Test 1 output for site H electrical trade group.
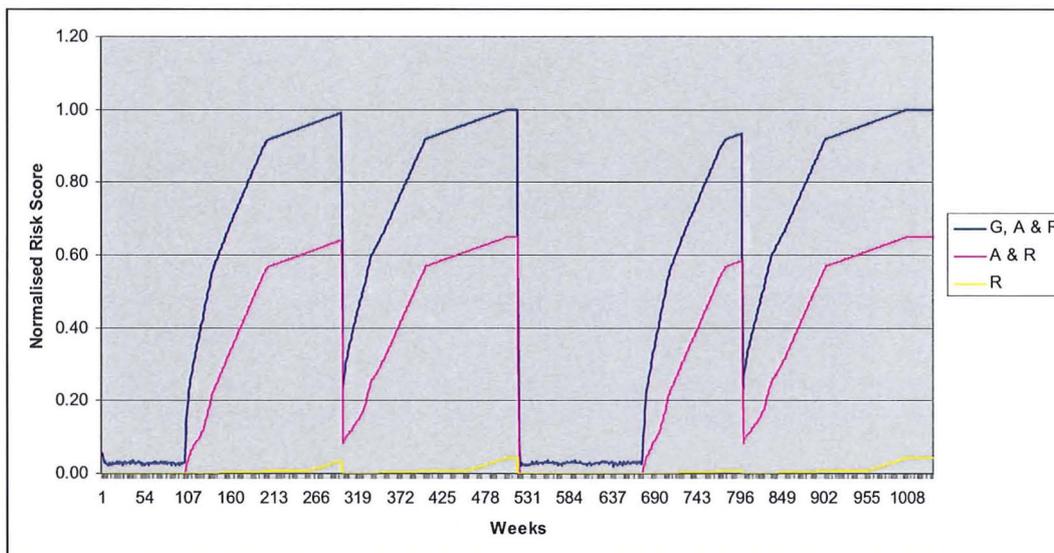
At week 300 a step function was introduced where the available man hours were increased for one week to excessive levels. This caused the model to complete all outstanding planned preventative maintenance tasks within the week, and their assessed risk fell to their residual levels by week 301. Between weeks 301 and 530 the available man hours remained at zero, and the assessed risk began to rise again. For week 530 through to week 690 the available man hours remained at excessive levels, and the assessed risks were kept at their residual values. These unrealistic step functions were repeated at weeks 531, 691 and 800. The program response to these extreme values gave confidence to all concerned that the model appeared to be operating correctly.

The second test was to run the simulation with the existing level of man-power, but to instantly reduce this level to zero at week 420, and observe the output over the full 1040

week (20 year) simulation period. Again using the values relating to site H and the electrical trade group within Trust X, the resulting graph that was produced is shown below in figure 46.

Figure 46   Test 2 output for site H electrical trade group.
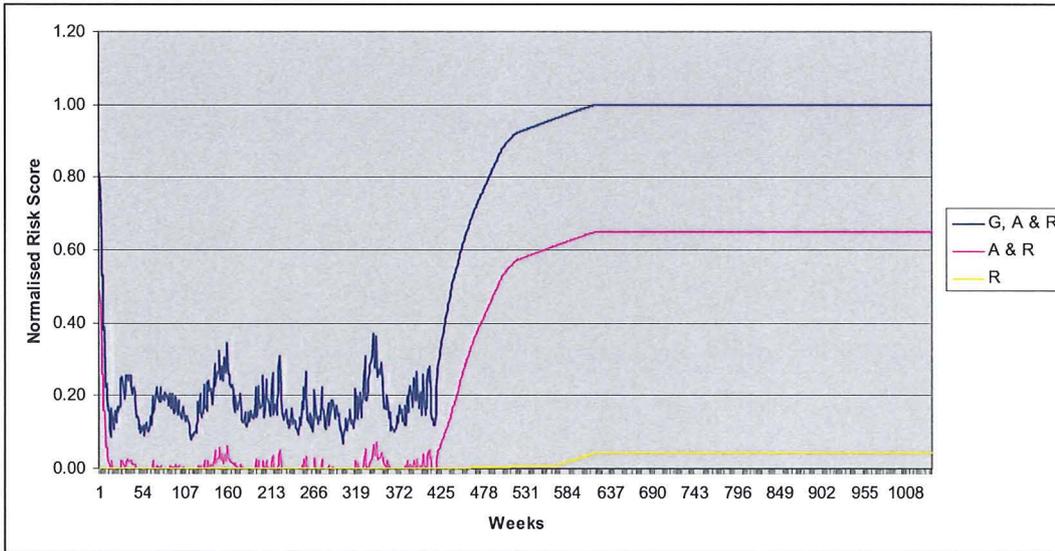
The graphs in figure 46 show that the simulation appeared to quickly settle after say week 20. Thereon after, the level of risk responded naturally to the variation in available man-power against the combined workload of the reactive and scheduled planned preventative maintenance need, until week 420. As expected at this point with the man-power reduced to zero, the assessed risk within all the three zones of green, amber and red began to rise to their maximum values. This sudden increase is in agreement with the work of Rudolph and Repenning (2002). They produced a very similar graph that showed an organisation's inability to handle quantities of tasks beyond their capability, where the input varies in an apparent random form. The remaining simulations were conducted with available man hour levels -3 Whole Time Equivalent (WTE), -2WTE, -1 WTE, existing WTE, +1 WTE, +2 WTE and +3 WTE.

From the computations, 36 additional arrays were created that summated the risk for each of the 1040 weeks (20 years) simulated. These arrays recorded the risk within the green, amber and red zones for each of the twelve levels of available man hours. The arrays were normalised to simplify the analysis and presentation, then arranged into 26 output files, as shown within the four tables 14, 15, 16 and 17.

196

Table 14

This table contained the title of the output files that recorded the simulation results from the two test available man hour levels. In both cases the test runs used the existing man hour figures as their base data, and presented their outputs as the assessed risks within three risk zones: green, amber and red zones (G, A & R), the amber and red (A & R) zones only and finally the assessed risks within the red zone (R).

Table 14   *Output files from the two test simulations.*

File No	Title
1	Test 1
2	Test 2

Table 15

The output files within this table recorded the total computed risks within a series of 100 week steps, which advanced one week at a time generating 940 steps throughout the 20 year simulation. It was from these files that the researcher (and others) could confirm when each simulation had settled down into a steady state and was ready for analysis. The files record the computed risk for the following seven levels of available man hours, -3 WTE, -2 WTE, -1 WTE, existing WTE, +1 WTE, +2 WTE and +3 WTE. Again the outputs were split into the computed risks that fell within the G, A & R, A & R and R zones. In order to assist in determining the settling down period, the standard deviation (SD) for each of the 940 periods was recorded.

Table 15   *The computed risk within each of the 940 sampling periods and their respective standard deviation (SD).*

File No	Title
3	G, A & R in sampling periods
4	A & R in sampling periods
5	R in sampling periods
6	SD of G, A & R sampling periods
7	SD of A & R sampling periods
8	SD of R sampling periods

Table 16

This table shows the largest of the output file groupings. In order to ease the interpretation the outputs were arranged into five sets to reflect the available man hours used within each simulation. These sets were: very low WTE, low WTE, existing WTE, high WTE, and finally very high WTE. Again all were subdivided into the various risk zones.

Table 16 *The simulation outputs split into five sets of available man hours.*

Set	File No	Title	WTE		
1	9	G, A & R very low WTE	-3	-2	Existing
	10	A & R very low WTE	-3	-2	Existing
	11	R very low WTE	-3	-2	Existing
2	12	G, A & R low WTE	-2	-1	Existing
	13	A & R low WTE	-2	-1	Existing
	14	R low WTE	-2	-1	Existing
3	15	G, A & R med WTE	-1	Existing	+1
	16	A & R med WTE	-1	Existing	+1
	17	R med WTE	-1	Existing	+1
4	18	G, A & R high WTE	Existing	+1	+2
	19	A & R high WTE	Existing	+1	+2
	20	R high WTE	Existing	+1	+2
5	21	G, A & R very high WTE	Existing	+2	+3
	22	A & R very high WTE	Existing	+2	+3
	23	R very high WTE	Existing	+2	+3

Table 17

This remaining table shows the files that recorded the assessed risk if there was a long period of sickness or a vacancy was deliberately held open. These sickness/vacancy periods were three, six and nine months. In this case the simulations only recorded these effects against the existing available man hours. Again the files were split to show risk within the three zones.

Table 17   The effects of sickness/vacancy

File No	Title
24	G, A & R sick/vac
25	A & R sick/vac
26	R sick/vac

The 26 output files were converted from the FORTRAN format into that suitable for opening within the excel environment, as this was in all cases the preferred data format for Trusts. From this the Trusts were able to access these files for their own future analysis and forward man power planning, etc. These files were converted to initially 26 output risk profiles for each trade/site. As the Trusts began to interrogate the information they created additional profiles from the data.

4.4.2   Simulation

In order to illustrate the typical results obtained from the various Trusts that were part of this research, reference is made throughout this section to one particular Trust. In order to protect its identity throughout this thesis it shall be named Trust X. This Trust operates from two large acute sites that are approximately seven miles apart. The two sites shall in turn be called site H and site C. Both sites H and C have electrical, mechanical and semi-skilled trade groups, i.e. six independent trade groups. Only the simulation outputs from Site H electrical are shown and briefly discussed below as an example of the inputs/outputs.

Due to the nature of the planned preventative maintenance schedule being pre-planned the waveform developed from this set of data created an almost straight line of the man-hours need throughout the three years. Figure 48 below shows the waveform developed from the available man hour computations, following the removal of the backlog maintenance, new works and diverted time allocations for the past three years, all on a week-by-week basis. Similar waveforms were created for the remaining five trade groups.

Figure47   Site H electrical available man hour waveform
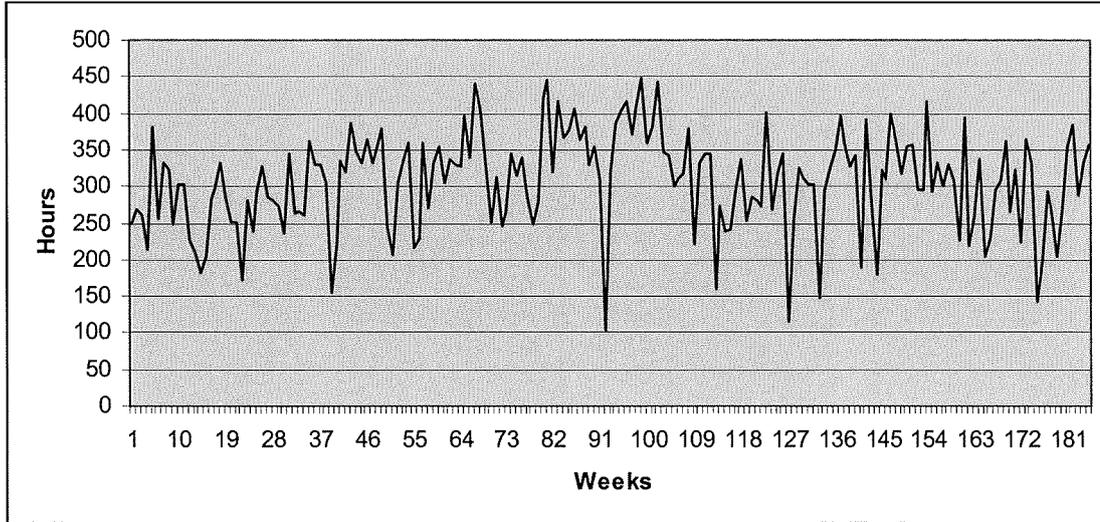
The reactive maintenance waveform shown in figure 48 produced a similar apparent random fluctuation of values as in figure 47 above.

Figure 48   Site H electrical defect hour waveform
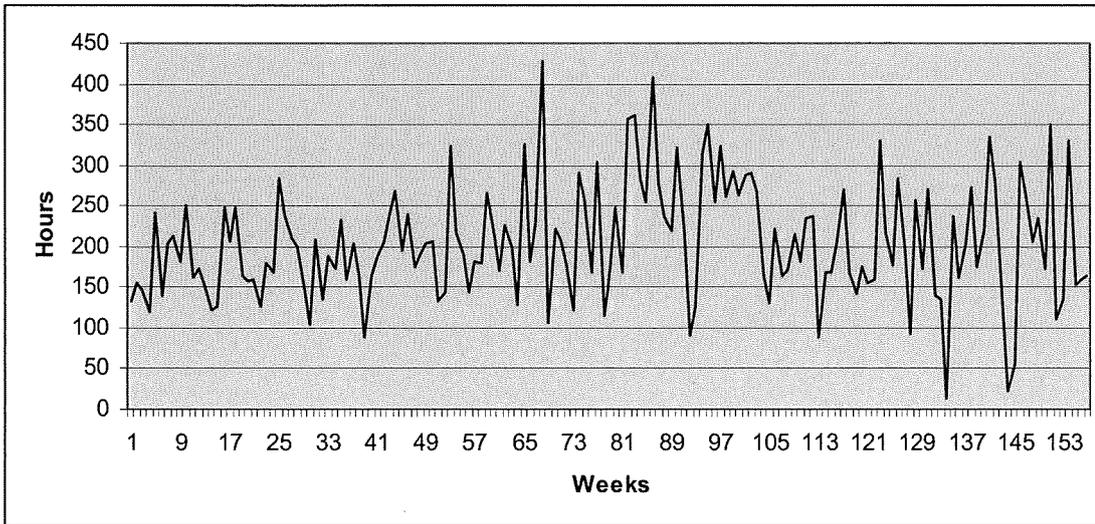
From the simulation of the three data sets of available man hours, reactive hours and planned preventative maintenance scheduled hours seven values of available man hours were simulated over a 20 year period. These included -3 WTE, -2 WTE, -1 WTE, existing WTE, +1 WTE, +2 WTE and +3 WTE. Also during the process the risk generated from each uncompleted planned preventative maintenance task was split into

their respective green, amber and red risk zones. Figure 49 below shows for the existing available man hours only the total normalised assessed risk computed for the last 10 years of from the 20 year simulation, as it was assumed that the simulation had settled into its steady state by year 10.

Figure 49 *The total normalised assessed risk computed from the last 10 years of the 20 year simulation for the existing man hours only for site H.*
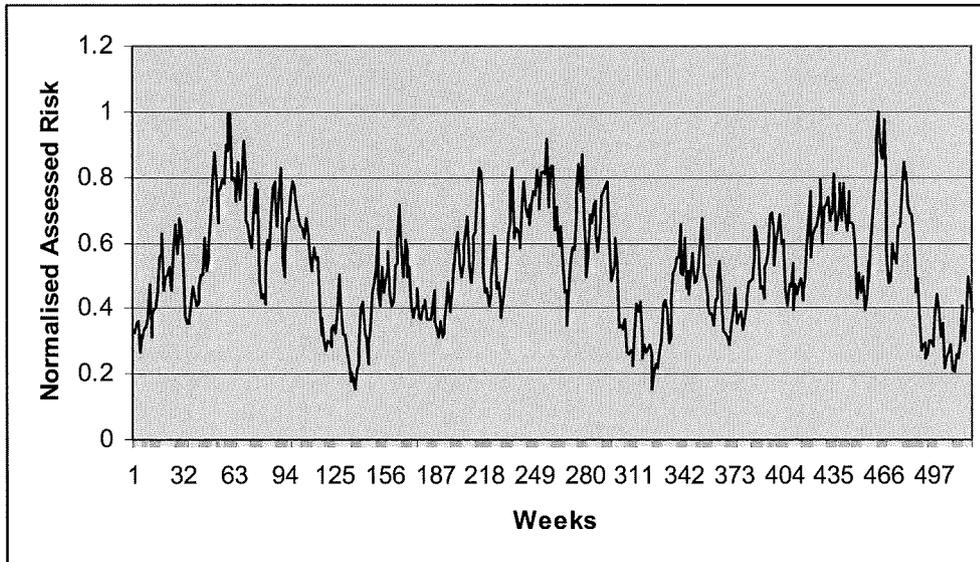
Figure 50 below shows a 100 week period taken from this steady state phase. All data relating to the computed assessed risks within the green zone have been removed leaving only those assessed risks within the amber and red zones.

Figure 50   Sample assessed risk waveform taken from the equilibrium zone of the 20 year simulation period for the electrical trade group showing the amber and red zones only on site H.
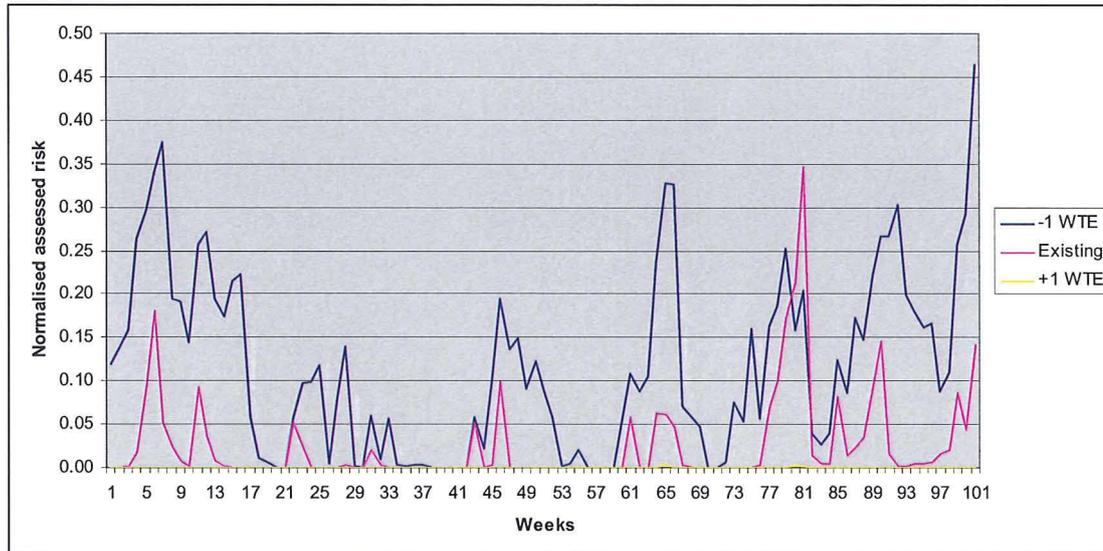
From the profiles developed within Figure 50, the simulation model allowed the observation of five main themes:

a.   Variation of WTE on assessed risk levels

The graph demonstrates the effects of increasing or decreasing the existing available man hours with respect to the assessed risk by plus or minus one whole time equivalent (WTE). Because the methodology employed to compute the resultant risk used utility functions, the areas beneath the graphs are a true ratio of the assessed risk in relation to that generated between the three levels of WTE .

b.   Involuntary trespass into the amber and red zones

The graph shows the remaining assessed risk as a series of 'spikes'. This demonstrated that the short intermittent and involuntary trespass into the amber and red zones as a naturally occurring event.

c.   Level of outstanding planned preventative maintenance in stock

Around week number 81 it is observed that there is an overlap between the graphs for -1 WTE and that for the existing complement of staff. It shows that as the WTE is

decreased the risk reduces. This apparent anomaly was seen to occur within other simulations for other trade groups within other sites; but only rarely.

The suggested cause of the phenomenon is that each simulation was run totally independently. The content of the planned preventative maintenance tasks in stock waiting to be completed varied with respect to time, as this level was dependent on the week's available man hours. For the majority of the simulations this stock variation appeared to keep in step between the various levels of WTE, but occasionally a stock accumulation occurred that knocked this out of phase. It must be remembered by the observer that like any simulation model the outputs attempt to copy reality. They cannot claim to mimic real life, but only to come close enough to allow judgements to be based on their predictions within levels of confidence, as previously discussed within the Literature Review Chapter.

For the great majority of the time, as the available man hours were decreased by one WTE the level of assessed risk significantly increased, similarly the reverse occurred with the increase in available man-hours by one WTE.

d.   Removal of all assessed risk within the amber and red zones
The model clearly demonstrated that increasing the level of WTE to a point where all of the assessed risks were removed from the amber and red zones would need the available man hours to be increased to well beyond reasonable levels, and that this would leave an unacceptable excess of available man hours/week.

e.   Interpretation needed a 'Judgement Call'
The above demonstrated that the correct interpretation of what available man hours were acceptable relied upon a 'Judgement Call' on behalf of the Trust's most senior managers. This lack of precision within an all embracing algorithm to determine the directly employed labour resource need is recognised within Estatecode, which clearly states that estimates, should be reviewed using professional judgement and local site knowledge to reach the final decision (NHS Estates 2002). It was from profiles similar to that in Figure 50 that Trusts were able to make their 'Judgement Call', in relation to the level of risk that they were willing to accept. This in turn set the level of available man

hours that they were prepared to balance against the budget, which they considered appropriate when compared to all of the other known Trust risks.

4.4.3   Main simulation outputs for Trust X

Graphs showing the assessed risks against all seven levels of available man hours were developed for each of the three trade groups for each of the two sites H and C. This data was then presented to and discussed with the Trust's estate managers, together with an 'initial' assessment. These 'initial' assessments are summarised and shown within Table 18 and discussed below.

Table 18   A summary of the estates manager's initial assessments of existing manning levels for Trust X.

Trade Group	H	C
Electrical	Correct	Correct
Mechanical	Correct (ish)	Undermanned by 2 WTE
Semi-Skilled	Overmanned by 1 WTE	Overmanned by 1 WTE

Below is a summary of the profiles used by the estates managers in Trust X to assess their current manning levels against the Trust's specific needs. As stated they are a small proportion of the actual data created, and are contained within this Chapter to demonstrate the type of information that was developed as part of the simulation process. The totality of the output information was presented to the Trust within a Compact Disk for their (and others) further consideration. The final analysis suggested that the existing available man hours were delivering an acceptable level of assessed risk, and that any variation of available man hours of plus or minus one WTE would create an unacceptable increase in assessed risk or a waste of resource. To ensure continuity of presentation Figure 51 below is a direct copy of Figure 50 above.

Figure 51   Sample assessed risk waveform taken from the equilibrium zone of the
20 year simulation period for site H electrical trade group, showing the amber and
red zones only
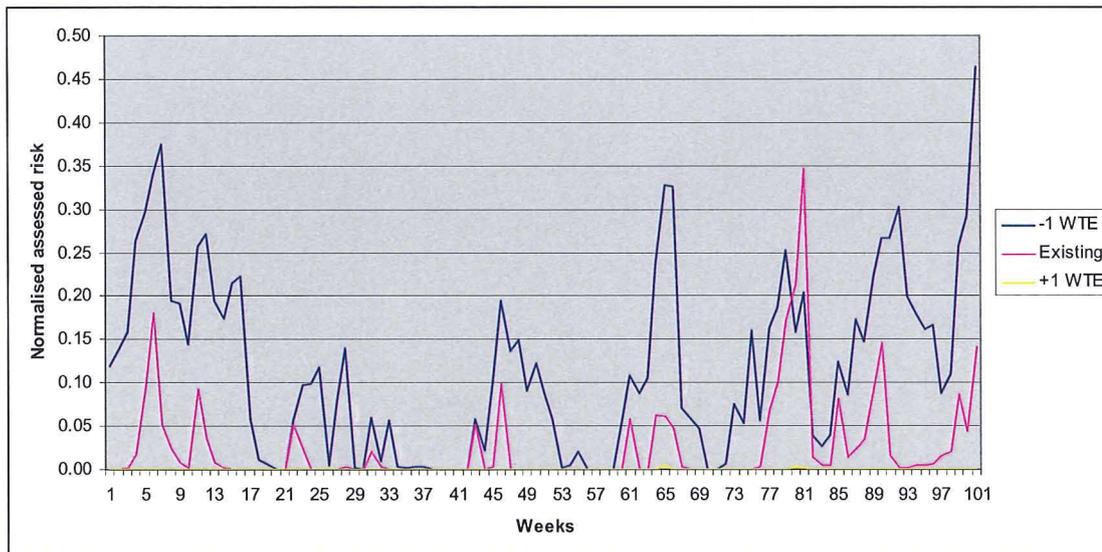
When the mechanical trade group within site H were subjected to the simulation program figure 52 was developed. This shows the variation of plus or minus one WTE was less clear cut as in the previous assessment for the electrical trade group, and therefore required further discussion and analysis in an attempt to form a firm opinion. However the Trust was advised that such an ideal solution may not be possible to obtain, and that a compromise may have to be made.

Figure 52   Sample assessed risk waveform taken from the equilibrium zone of the
20 year simulation period for site H mechanical trade group showing the amber
and red zones only
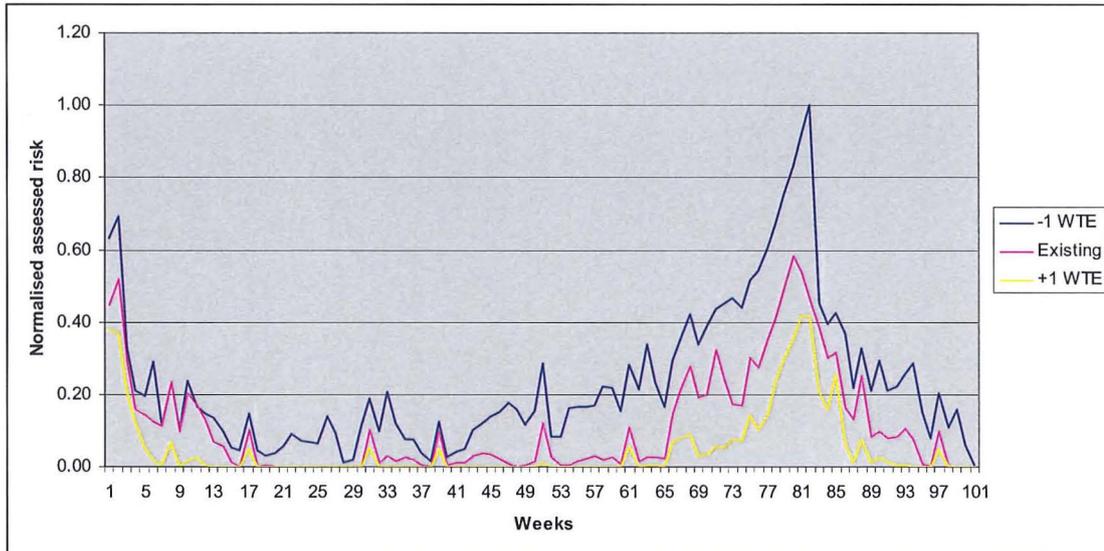
The graph in figure 53 relating to the semi-skilled trade group clearly demonstrated that by reducing the directly employed labour force by one WTE, there was little effect on the level of assessed risk that would be experienced. Additional data showed that any further reduction by say two WTE would without doubt create serious levels of assessed risk.

Figure 53   Sample assessed risk waveform taken from the equilibrium zone of the
20 year simulation period for site H semi-skilled trade group, showing the amber
and red zones only
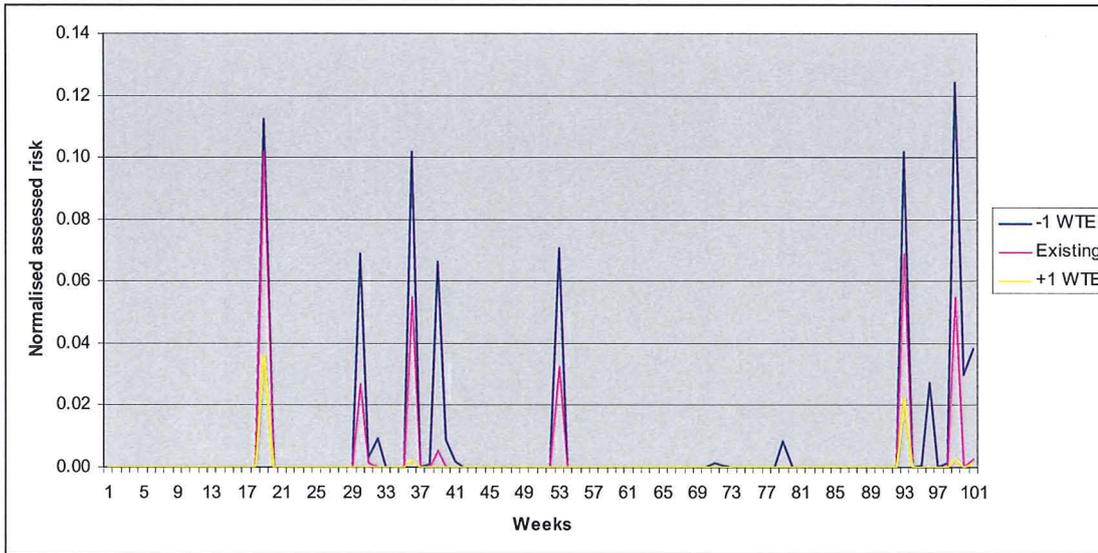
For site C the electrical trade group risk generation graph (in figure 54) attracted the same comments as with the site H electrical trade group, as the assessment suggested that the existing available man hours were delivering an acceptable level of assessed risk, and that any variation of available man hours of plus or minus one WTE would create an unacceptable increase in assessed risk, or a waste of resource.

Figure 54 Sample assessed risk waveform taken from the equilibrium zone of the
20 year simulation period for site C electrical trade group, showing the amber and
red zones only
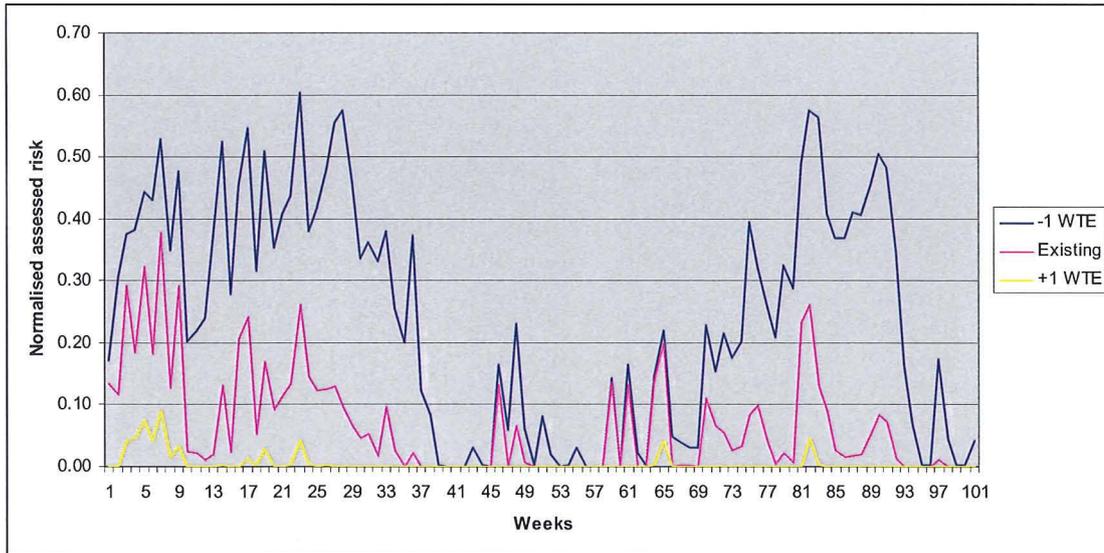
Serious concerns were raised with regard to the existing level of available man hours for the mechanical trade group in site C as shown in figure 55. It was clearly demonstrated that in this case the staffing levels were undermanned by between two or three WTE at least, as the assessed risks being generated were continuously within the amber and red zones throughout the equilibrium period.

Figure 55   Sample assessed risk waveform taken from the equilibrium zone of the 20 year simulation period for site C mechanical trade group, showing the amber and red zones only
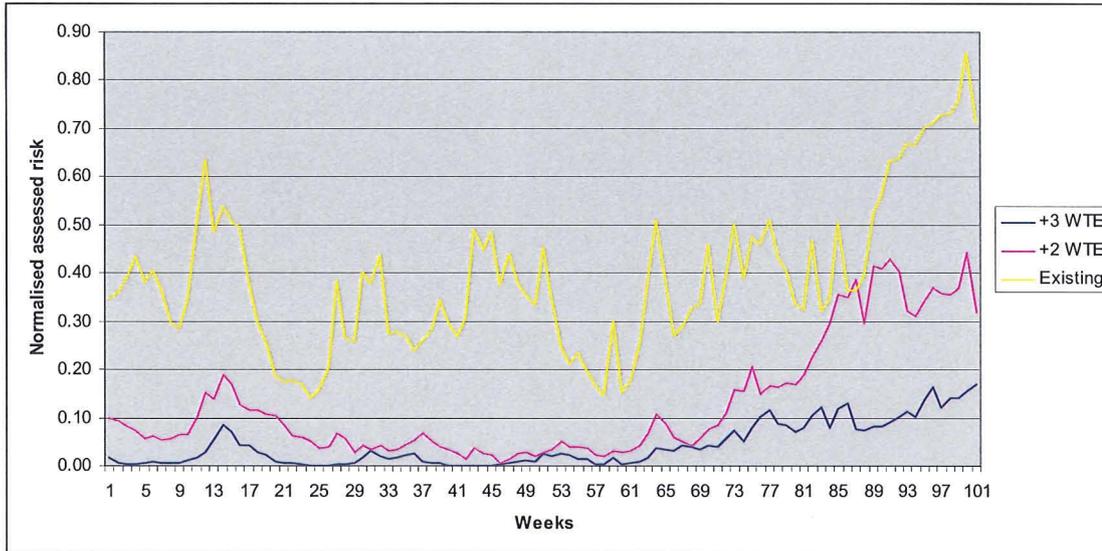
The site C semi-skilled trade group risk generation graphs, shown in figure 56 were assessed as being overmanned in excess of one WTE.

Figure 56   Sample assessed risk waveform taken from the equilibrium zone of the 20 year simulation period for site C semi-skilled trade group, showing the amber and red zones only
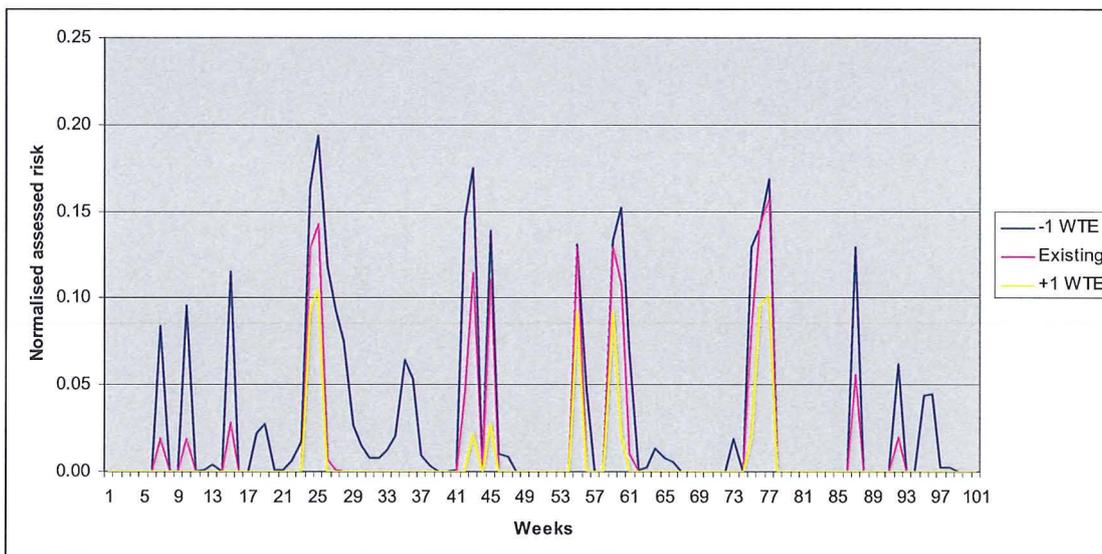
By reducing the available man hours there was no significant increase in the level of assessed risk. However, when the whole of the 20 year simulation for this trade group was viewed, four large periodic increases in risk were experienced when the manning level was reduced. But this was interspersed with relatively long low levels of risk. Therefore again the 'Judgement Call' was not clear cut but did suggest that the existing manning level was high.

4.4.4 Sensitivity of results to uncertainty in assumptions

It was only when the simulations using real data had been completed, could the actual sensitivity of the model's outputs and design assumptions be examined and observed. In order to ease the readability of this results section the simple tests and short outputs designed to analyse the sensitivities of the model have been included within this chapter. This section is split into the following discrete eight parts:

    a. Determination and observations of the steady state value
    b. Identification and observations of possible problem planned preventative maintenance tasks
    c. Rounding error identification and observation
    d. Interpolation verses integration within the software design
    e. The focus groups being risk adverse
    f. Probability utility sensitivity
    g. Impact utility sensitivity
    h. Crude assessment of man-power needs

a.   **Determination and discussion of the steady state value**

In practice when making any decisions, a number of 100 week periods were presented from different parts of the 20 year simulation time frame. In this way a cross-section of the outputs was viewed. For confidence to be assured that the model had indeed reached a steady state, the following three tests were applied.

    i   Total risk within each 100 week period
    ii   Standard deviation for each 100 week period
    iii   Consistency of each 'Judgement call'

i. Total risk within each 100 week period

The output was split into 100 week periods, where each period advanced one week at a time, thus forming 940 periods throughout the 20 year simulation cycle. This arrangement is shown in table 19 below:

Table 19   Table showing the 100 week periods.

Period	Start Week	End Week
1	1	100
2	2	102
3	3	103
-	-	-
-	-	-
-	-	-
-	-	-
938	939	1038
939	940	1039
940	941	1040

For each 100 week period the area under the resulting risk graph was calculated. From these data Figure 57 (below) was developed. This shows the total assessed risk within each 100 week period against each period start week.

Figure 57   The total assessed risk within each 100 week period for site C electrical trade group.

211
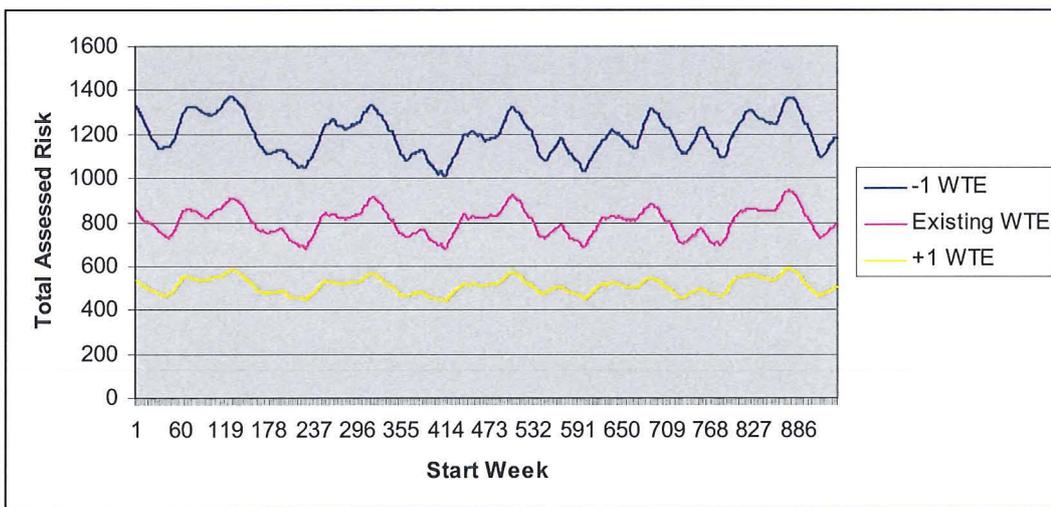
ii. Standard deviation for each 100 week period

As a normal distribution about the mean was assumed, the standard deviation was also computed for the period. As the areas and associated standard deviations stabilised, it was assumed that the overall 20 year simulation also stabilised.

iii. Consistency of each 'Judgement call'

The third and 'acid test' to confirm that the simulation had indeed steadied, was to take a number of 100 week periods as representative samples; and if the same 'judgement call' relating to the effects and acceptance of any man-power variation were made, then it would appear that the system had reached a steady state.

In practice the summation of the total assessed risk beneath each 100 week period, and resulting standard deviation computations were inconclusive. Therefore the whole process relied on the stability of the 'Judgement Call' from a number of 100 week periods to be the determining factor. This sounds crude and an unconvincing method to use, but the variation in risk generated by reducing/increasing the directly employed labour workforce by one WTE was pronounced in the majority of cases. Although the simulations were considered to have settled to their steady state by week 300, each of the focus groups decided to concentrate their analysis from the simulated year 10 onwards (weeks 500 to 1040), as a precaution.

b. Identification and observation of possible problem planned preventative maintenance tasks

This part of the analysis was formed to identify and discuss if an individual planned preventative maintenance task (or group of planned preventative maintenance tasks), was continuously generating high levels of assessed risk in comparison to others. This necessitated recording each level of computed risk for each planned preventative maintenance task for each week throughout the 20 year simulation period. This was repeated for each of the seven levels of the available man hours.

It was agreed previously between the focus groups and the researcher that the simulations had settled down into their steady state from year 10 onwards. This then formed a consistent timeframe within which all of the data could be compared. From year 10 onwards the total assessed risk generated between years 10 to 20 was

212

summated then normalised to ease presentation and interpretation. The graph below shows the risk generation for the existing man hour availability. Six similar graphs were produced for the -3 WTE through to +3 WTE scenarios. Again to aid the discussion the output from the electrical trade group for site H is shown in figure 58.

Figure 58   The normalised total risk being generated during the last 10 years of the simulation for each electrical planned preventative maintenance task for the existing available man hour levels for site H.
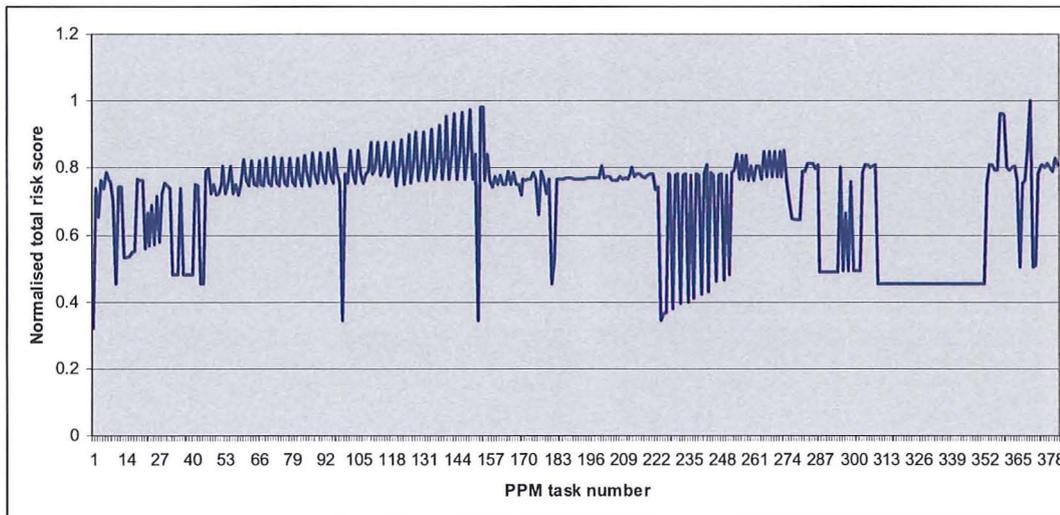
The figure shows a series of troughs notably at planned preventative maintenance task numbers 99, 152 and between numbers 310 and 352. By rearranging the graph into a descending order Figure 59 was developed. This rearrangement clearly shows that the total risk generated was held at a fairly constant value throughout the simulation period across the majority of the planned preventative maintenance tasks, with the exceptions being at the graph's extremities of left and right. By analysing these extreme points, it was found that for this level of available man hours a simple pattern at first appeared to exist, developing from the impact and probability profile shapes assigned to each planned preventative maintenance task.

Between numbers of 1 to 20 the planned preventative maintenance tasks were formed from similar values of impact, all with a monthly frequency (M), and all with probability shapes that generated the lowest levels of the rate of rise with time. planned

213

preventative maintenance task numbers of 326 to 366 were composed with impact descriptor values of two, with a corresponding yearly frequency (Y) and in the majority of cases assigned with probability utility shapes that created the greatest rate of rise with time. From planned preventative maintenance task number 367 onwards the impact descriptor was five with a corresponding frequency of weekly (W), and in the majority of cases was assigned with probability utility shapes that created the greatest rate of rise with time. The mid range of the graph was created from a mixture of impact and frequency values with varying probability utility shapes.

Figure 59   The normalised total risk being generated during the last 10 years of the simulation for each electrical planned preventative maintenance task for the existing available man hour levels for site H arranged in descending order.
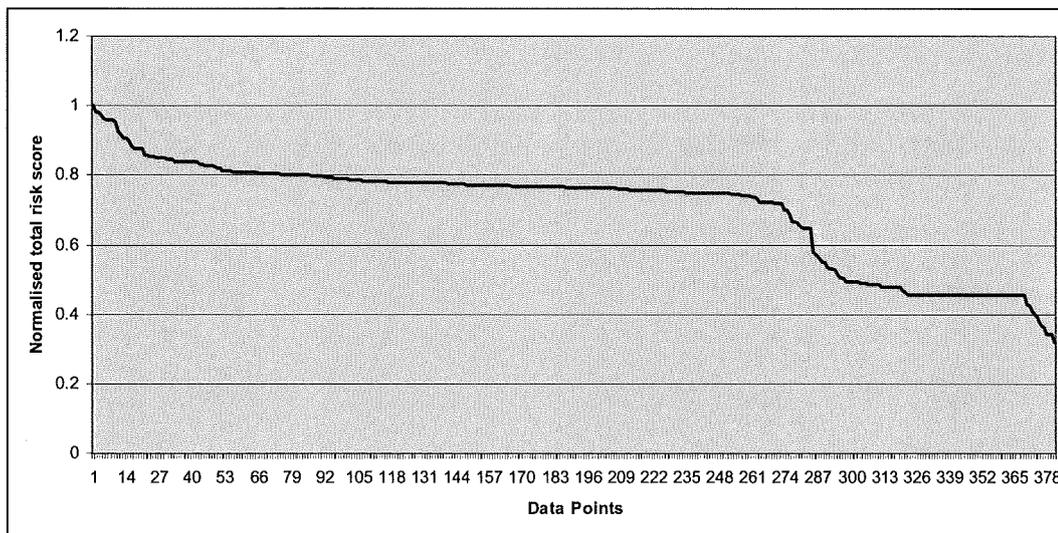
The procedure of developing and analysing a graph with descending values of risk, as in Figure 59 above was repeated for available man hour values of -3 WTE and +3 WTE. From this Table 20 was formed.

Table 20   Components of the assessed risk formed within figure 59 above

WTE	High risk end			Low risk end		
	Freq'	Impact	Prob' UF	Freq'	Impact	Prob' UF
+3	Low	Low	Med	High	High	High
Existing	High - Med	High	Low	High	High	High
-3	Low	High	High	Low	Low	High

For simplicity the frequency range was split into three zones. The high zone was included with weekly (W), two weekly (2W) and three weekly (T) frequencies, the mid - range as a monthly frequency (M), and the remaining quarterly (Q), half yearly (H), and yearly (Y) frequency as low. The impact was split into a low zone containing descriptors one and two, the mid-range being three with the high range being four and five. The probability utility function shapes split neatly into again three zones with those shapes that created rapid rates of rise with respect to time being classed as high, those with a low rate of rise as low and medium for those in between.

As can be observed from table 20 the initial assumption that a simple pattern could exist between the variables was shown to be false, as the interaction was complex.

From a further analysis of the data, the possibility that certain planned preventative maintenance tasks were continuously generating high levels of risk was explored. The output files from the simulations recorded the number of times that each planned preventative maintenance task was the maximum value for each week, i.e. the number of hits where the simulation identified that a particular planned preventative maintenance task was deemed to be generating the greatest risk for the week. The researcher recognises this methodology is crude, but it assisted in forming a picture of what was happening within the model throughout the simulation period. Again using the existing available man hours as a datum point the graph within Figure 60 was formed.

Planned preventative maintenance task number one received 41 hits out of a possible 520 throughout the simulation period, and was composed of the highest frequency weekly (W), with an impact descriptor five the maximum possible, and was assigned a probability shape that generated one of the greatest rates of rise possible. This again indicated that a simple pattern would exist, but when the files for the remaining six levels of available man hours were examined no simple pattern was visible, as they all displayed various maximum values at differing planned preventative maintenance task numbers, and with various values of impact, frequency and probability utility rates of rise.

From the hunt for possible problem planned preventative maintenance's, none were found. But the investigation did show that a simple pattern of impact, probability profile

shape, and frequency combinations appeared not to exist, and that the resulting assessed risk profiles appeared to be as a result of a complex interaction of values.

Figure 60 **The number of times that each planned preventative maintenance task was deemed to be generating the greatest risk for a given week.**
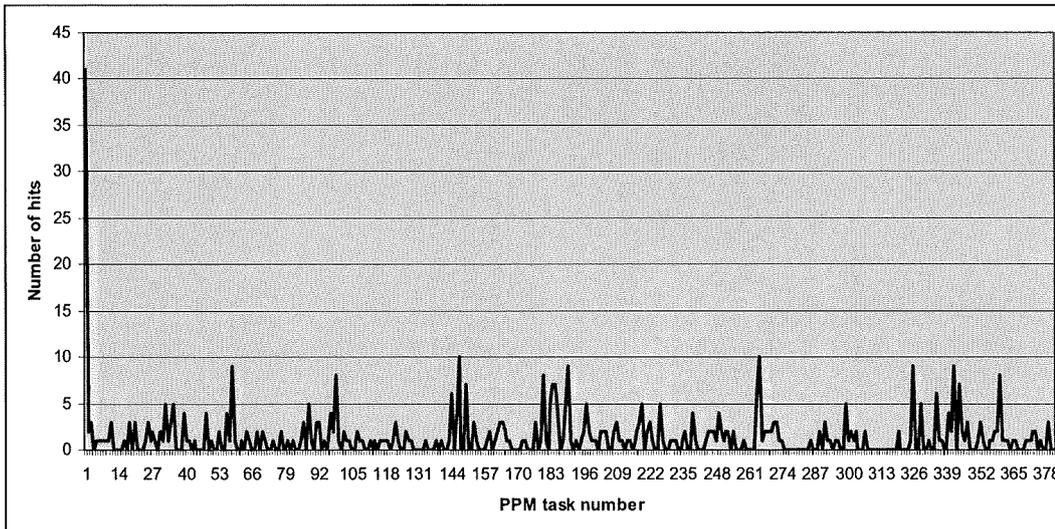
c. Rounding error identification and observation

Again once the simulation outputs had been formed from real data, it was possible to determine and discuss if the model was prone to rounding errors.

Numbers are stored within a computer with limited precision; this loss of precision of a value is termed a 'round-off' error. As complex arithmetical operations are performed, such as within a simulation model, these errors can combine to become a considerable source of model error (Beddall 2006). As numbers are stored in a binary format, their precision is influenced by the size of computer memory assigned to store the binary bits, where the larger the memory, the greater the precision. The Fortran compiler used to create the simulation model normally operated at what is termed 'single' precision, which generated seven significant figures. To test if the model was suffering from rounding errors and creating incorrect predictions, the simulations were rerun using 'double' precision. This is where the memory allocated to storing the numbers is increased to create a number with 15 significant figures.

216

Although the model's program was inherently prone to round-off errors due to the physical design of computers, there were no observed differences between the outputs from the single or double precision simulations for seven significant figures. Therefore the researcher considers that rounding errors were not a significant problem within the simulation model designed.

d.   Interpolation verses Integration within the software design

This part of the results assesses the significance of the error caused by employing interpolation of the probability utility in place of an integration technique. From reference to the probability of failure descriptor part of the NHS 5x5 criticality grid, the estates officers (in meetings chaired by the researcher) assigned a probability utility shape to each individual task. As with assets (plant and equipment) types were grouped together to share a common utility shape. For site H and site C the results found that only 22 of the possible 72 utility shapes were used. Each of these 22 shapes was assigned a reference letter, and their usage pattern formed into Figure 61 below.

Figure 61   The probability utility shape usage
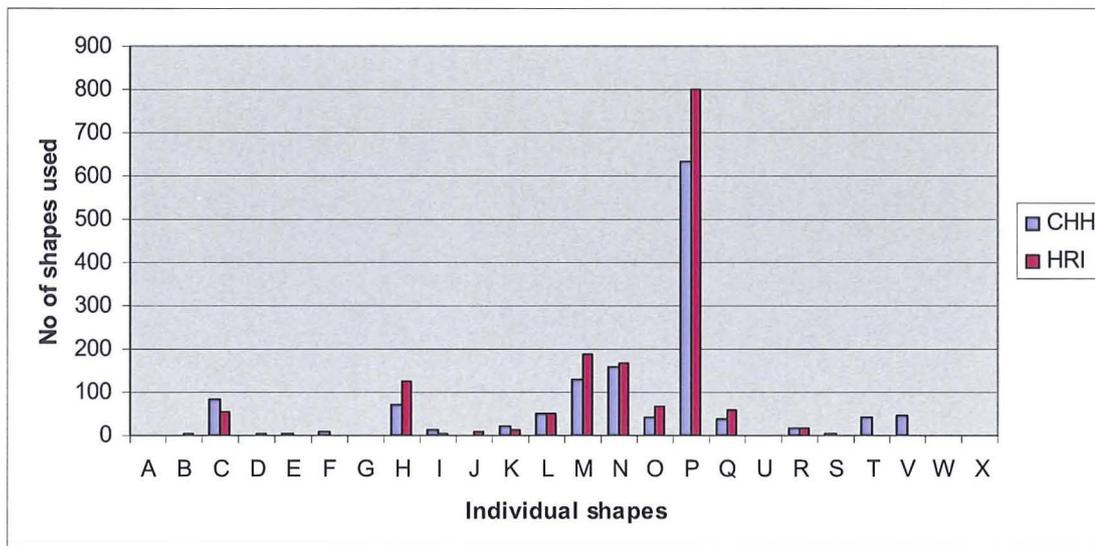
A trend line was fitted to the most commonly used utility shape as shown in Figure 62 below. The researcher assumed that the trend line was the true probability utility formed by the focus group. It was then assumed that the difference between areas formed by

the trend line calculated by integration, and that formed by the straight line interpolation technique would give an indication of error.

Figure 62 *The trend line and straight line shape formed by the probability utility shape P.*
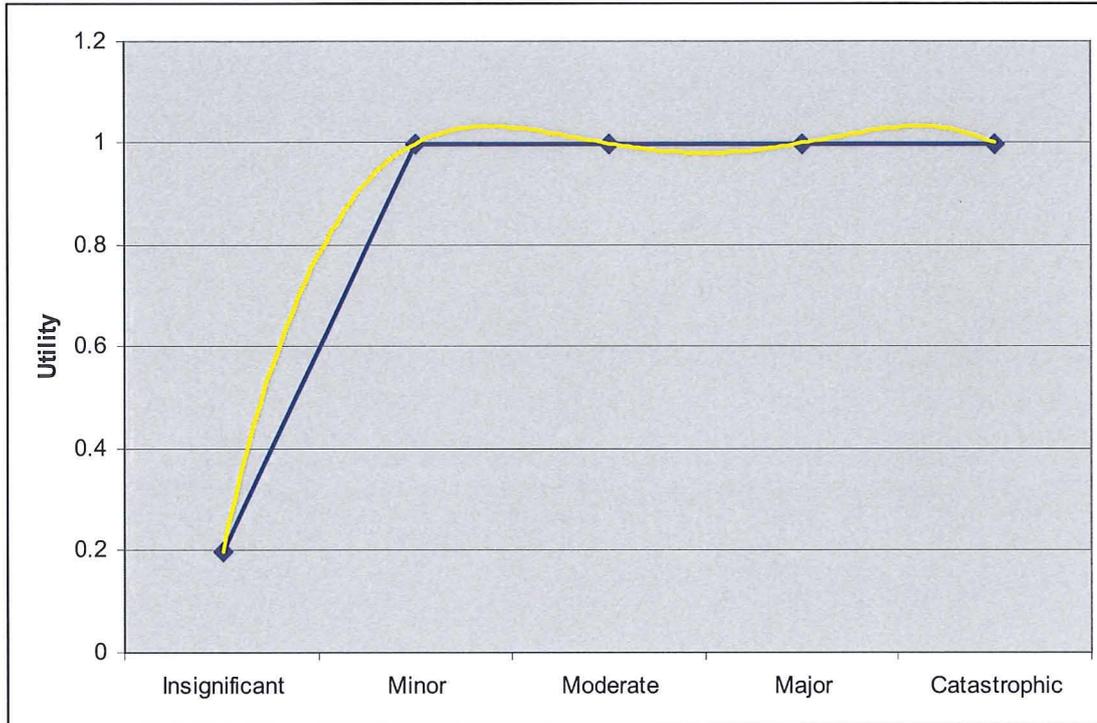
The lowest frequency of inspection was three weekly (T). This was the worst case scenario, in relation to interpolation accuracy. Therefore the difference in areas under shape P was computed for this frequency, first by Simpson's Rule (as the formula for interpolation between points), and then by integrating the trend line algorithm. The resulting calculations are shown below.

Table 21 was formed for each of the thirteen weeks analysed for the three weekly (T) planned preventative maintenance task. It shows for each week the probability descriptor rating and corresponding utility value.

Table 21   Table showing various values for each week

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	T			T			T			T			T
Utility value	0.2	0.46	0.73	1	1	1	1	1	1	1	1	1	1
Probability descriptor rating	1			5			5			5			5

Area approximates to $\frac{S}{3}[(F + L) + 4E + 2R]$

Where:

$$
\begin{aligned}
S &= \text{width of each strip} \\
F + L &= \text{sum of the first and last ordinates} \\
4E &= 4 \text{ x the sum of the even numbered ordinates} \\
2R &= 2 \text{ x the sum of the remaining odd-numbered ordinates}
\end{aligned}
$$

As the time steps are each one week in length, and there are 3 weeks between frequencies, then $S = \frac{1}{3}$. Therefore the approximate area under the probability function between the probability of failure rating limits of 1 and 5 is shown below, followed by a corresponding calculation using integration:

Approximate area $= 1/9[(1 + 5) + 4(.46 + 1 + 1 + 1 + 1 + 1) + 2(0.73 + 1 + 1 + 1 + 1)]$

$= \textbf{3.61 units}^2$

$$y = -0.0333x^4 + 0.4667x^3 - 2.3667x^2 + 5.1333x - 1$$

Approximate area $= \int_{1}^{5} -0.0333x^4 + 0.4667x^3 - 2.3667x^2 + 5.1333x - 1$

$= \textbf{3.77 units}^2$

The difference between the areas computed by the two methods for the worst case scenario was -0.42 percent, therefore the researcher considers that any error caused due to the adoption of the interpolation technique, was negligible.

e.   The focus groups being risk adverse

If the focus groups were risk adverse they would be overestimating the probability and impact utility values. Therefore as part of this discussion it was important to estimate the sensitivity of both of these utilities on the final risk profile results produced. The tests for both the probability and impact utility sensitivities follows

219

f.   Probability utility sensitivity

The method chosen was to simulate site H Mechanical planned preventative maintenance schedule five times, each run spanning a 10 year cycle, and analysing the last 100 weeks of each simulation for any significant difference. The first simulation was with the original shape P, which accounted for 54 percent of the profile shapes that were assigned by the focus group across the 497 individual planned preventative maintenance tasks within this schedule. This shape was also the most risk adverse shape that the focus groups could have chosen. The simulation was repeated by substituting for shapes O, N, M and W. A comparison of these five utility shapes is shown within Figure 63 below.

Figure 63   The five probability utility shapes used to assess their relative sensitivities
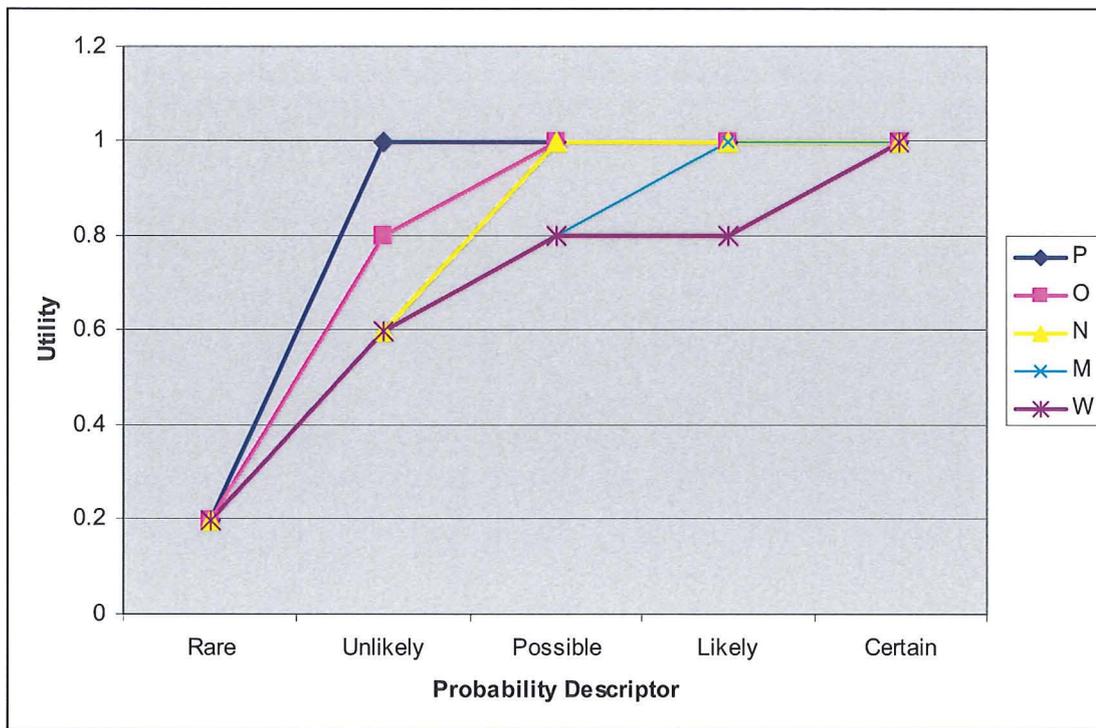
The resulting areas under each of the five risk profiles developed for the red and amber zones were then compared. The probability utility shape P was used as the reference to which all other areas were compared. For shape O there was no detectable difference in area. For the remaining three shapes equal differences to shape P were recorded of -4 percent and -5 percent for the current manpower level and for a reduction of 37.5

hrs/week respectfully. When the manpower was reduced by 75 hrs/week the difference for the three shapes became -16 percent. Figure 64 was developed to observe these differences.

Figure 64 *The resulting risk profiles from the adoption of probability shapes P and W.*
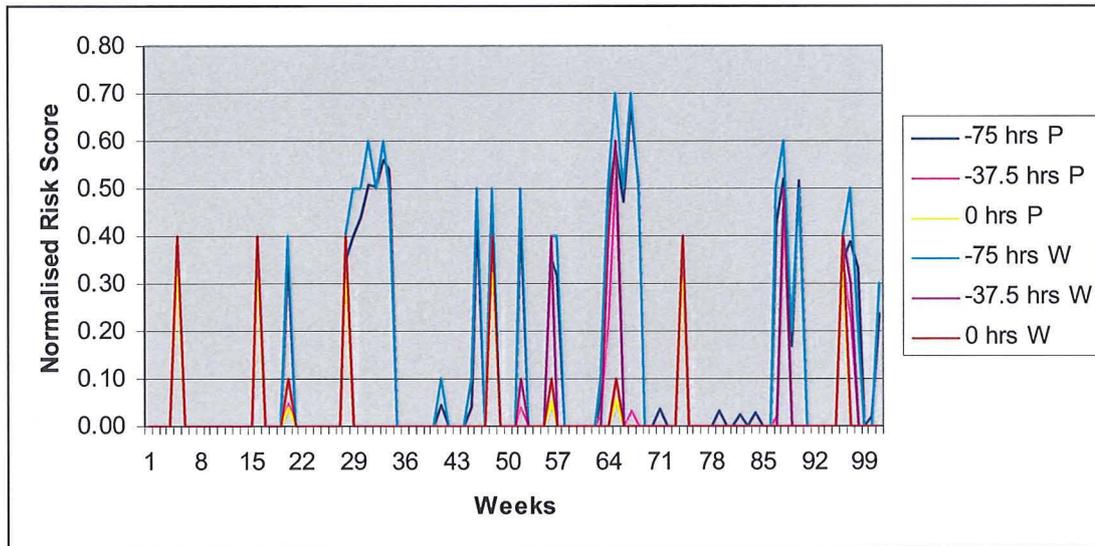
It is difficult to distinguish the differences visually between the risk profiles, except for the 75 hr/week manpower reduction. This is an important observation as the decision making process employed to determine the effects of reducing the available manpower to that acceptable, relies in part on the visual inspection of the resulting risk profile graphs. The researcher considers that the differences between the probability utility shapes of P and W are extreme. Thus the researcher assumes that the methodology is not significantly prone to a focus group being risk adverse, as the effects between the simulated extreme values were considered slight by the researcher.

g.   Impact utility sensitivity

This section explores the sensitivity of the focus group's impact utility on the risk profile output. Within the research instrument designed to meet the needs of this objective the boundary lines between the red, amber and red zones were formed by splitting the maximum risk score (25 from the NHS 5x5 criticality grid) into three equal parts. Thus a risk score below 8.33 was deemed to be in the green zone, between 8.33 and 16.66 the amber zone and above 16.66 the red zone. This simple split sufficed for the accuracy

221

that was required.. However, when assessing the risk a much more refined methodology was required to determine the split between the three zones. From reference to the NHS 5x5 Criticality Grid, reproduced below in Table 22, it can be seen that the matrix is split into coloured zones representing differing levels of risk severity as a guide to its use.

Table 22   The NHS 5x5 Criticality grid

		Probability				
		1	2	3	4	5
	1	1	2	3	4	5
	2	2	4	6	8	10
Impact	3	3	6	9	12	15
	4	4	8	12	16	20
	5	5	10	15	20	25

Score Range	Risk Ranking
1 - 6	Low
7 – 10	Moderate
11 – 16	Significant
17 - 25	High

Department of Health, 2004 Page 18

As discussed within the Methodology Chapter the grid is comprised of ordinal scales only. Therefore they have no mathematical meaning except to denote rank order. Within this research these scales were replaced with probability and impact utility scales. The resulting risk score within the matrix was a true ratio assessment of risk. Within the Methodology Chapter the Analytical Hierarchal Process (AHP) was used to develop impact utilities of four individuals. Their Consistency Ratio (CR) values ranged from 0.087 to 0.0017, where an acceptable value was deemed to be less than 0.1 (Saaty, 2001). Therefore two criticality grids were designed that used these two extreme utilities of impact. These are shown below in Tables 22 and 23.

As all of the probability utility functions begin with the value of 0.2, then this research project regarded any probability <0.2 to be generating risk within the green zone. The determined significance of risk is a function of an interpretation of the descriptors for the relative probability and impact.

Table 23   The criticality grid developed from a low CR value (0.017) for the assessment of impact ratios

		Probability Utility				
		0.2	0.4	0.6	0.8	1
Impact Ratio	0.0395	0.0079	0.0158	0.0237	0.0316	0.0395
	0.067	0.0134	0.0268	0.0402	0.0536	0.067
	0.119	0.0238	0.0476	0.0714	0.0952	0.119
	0.296	0.0592	0.1184	0.1776	0.2368	0.296
	0.478	0.0956	0.1912	0.2868	0.3824	0.478

The researcher considered that it was essential to honour the judgments made within the national NHS grid in Table 22, as reflected within the colour coding of the matrix cells. The numerical rule observed within the grid was that no risk score value within a colour band (say yellow) could have a greater numerical value than those risk scores within a higher coloured band (say pink). Therefore when the risk scores were computed within Table 23, the colour coding of some of the cells were changed by the researcher to conserve the numerical rule dividing the different colour bands, and the national judgements. When the grid was constructed using the ratio set for impact with the high CR value of 0.087, the grid in Table 24 was produced, which again created slightly different coloured zones.

Table 24   The criticality grid developed from a high CR value (0.087) for the assessment of impact ratios

		Probability Utility Function				
		0.2	0.4	0.6	0.8	1
Impact Ratio	0.031	0.0062	0.0124	0.0186	0.0248	0.031
	0.069	0.0138	0.0276	0.0414	0.0552	0.069
	0.104	0.0208	0.0416	0.0624	0.0832	0.104
	0.274	0.0548	0.1096	0.1644	0.2192	0.274
	0.521	0.1042	0.2084	0.3126	0.4168	0.521

In order to maintain a simple approach for the focus groups the researcher amalgamated the yellow and pink zones into one amber zone, thereby conserving continuity of presentation with the research instrument created within the first research objective. From the two grids developed, the boundary splits between the red, amber and green zones were formed, as shown within Table 25 below.

Table 25   The determination of the zone boundaries

| Zone | Risk Score Range | |
	High consistency ratio value (0.087)	Low consistency ratio value (0.017)
Red	.521 - .219	.478 - .287
Amber	.219 - .054	.287 - .047
Green	.054 -	.047 -

The sensitivity of the differing consistency ratio values was assessed by running two simulations with different consistency ratio values of 0.087 and 0.017, and comparing the areas (via Simpson's rule) under the risk profiles for the combined red and amber zones, for the last 100 weeks of a ten year simulation. The results gave an identical assessment of area correct to four decimal places, for both the current manpower and when the manpower figure was reduced by 37.5 hrs/week. The only difference detected was when the simulation reduced the manpower by 75 hrs/week which gave a 0.05 percent variation in area. The researcher thus considered that the variation in consistency ratio below 0.1 produced insignificant effects.

h.   Crude assessment of manpower needs

From the researchers experience a formula that has been used for many decades within NHS estates departments' circles to assess the man-power needs is shown below.

$$(\text{Reactive hrs} + \text{planned preventative maintenance hrs})/(37.5(52\text{-}6)) = \text{WTE/yr}$$

The earliest use of this algorithm is unknown, and no reference documentation explaining its formulation has been found. But it has been handed down from engineer-to-engineer throughout the years. The reasoning behind the figures is that the total artisan workload for each trade group is the summation of the reactive and planned preventative maintenance hours needed for a typical year. This is divided by the total hours for a week, being 37.5, multiplied by the weeks of the year, with the total 52 weeks being reduced by 6 weeks. This reduction is used to account for any unproductive time such as holidays, sickness and other diverted time. The output computes the manpower needs as whole time equivalent (WTE) needs per year. Table 26 below shows the base data for the total hours per year for each of the trade groups.

224

When this simple formula was applied to the six Trust trade groups, using the data contained within Table 26, and compared to the assessments made using the risk based methodology created from this research; the results within Table 27 were formed.

Table 26   The total hours per year for planned preventative maintenance, reactive and available man-power per year

Site	Trade group	planned preventative maintenance total hours/year	Reactive total hours/year	Available man-power total hours/year
H	Mechanical	4598	7678	12595
	Electrical	3225	10504	15981
	Semi-Skilled	7217	9944	19639
C	Mechanical	8681	7491	12489
	Electrical	3835	6980	11825
	Semi-Skilled	6855	5662	12972

Table 27   Results from applying the simplistic formula to assess man-power needs against the existing man hour availability

Site	Trade group	Simple formula required WTE/year	Risk assessed methodology requirements
H	Mechanical	-0.2	Correct
	Electrical	-1.3	Correct
	Semi-Skilled	-1.4	- 1 wte
C	Mechanical	+2.13	+ 2.5 wte
	Electrical	-0.6	Correct
	Semi-Skilled	-0.3	-1 wte

The table shows that in all cases (except C Semi-Skilled) the simplistic formula underestimated the man-hours needed per trade group to generate an acceptable level of risk from their activities. The researcher considers that the assessed differences were due to the simple static formula (which takes no account of risk) being applied to a complex dynamic situation. However it does show that the simplistic formula has a use as a rough estimate of resource need, and could be used as a starting point when assessing resource need.

4.5   Chapter four Summary

First research objective

The results presented clearly showed that the NHS estates departments analysed did indeed have problems, and where they were. The results also showed that these previously unknown risks ranged from moderate through to substantial/intolerable levels.

Second research objective

The results clearly showed what resources were needed in terms of man-power, outsource costs and time frames.

Third research objective

The results clearly showed the effects of reducing the man-power below an acceptable level and the resulting increase in risk. Also shown were the naturally occurring peaks of risks which were formed by the rapidly changing dynamics of the available resources, set against the demands of the reactive and planned preventative maintenance workloads. Also presented are the results of an extensive sensitivity analysis of the model's outputs against the design assumptions made, and confirmation of the robustness of the model's predictions.

Fourth research objective

As previously explained, due to time restrictions no results with regard to this question have been developed for inclusion into this thesis, due to time and resource constraints. It was for this reason that the phrasing for this objective excluded the word 'implement'

Chapter Five   Discussion Chapter

5.1   Introduction

The previous chapter reported that in every Trust the results confirmed that there was a major problem with the managerial systems and overall governance arrangements, and that this was so severe that deficiencies were placing the Trusts, patients, staff, public and stakeholders at substantial/intolerable levels of risk.

This chapter discusses the research unpinning the methodology used to create the results, the techniques employed and the robustness of the outcome. Following the successful analysis, discussion and conclusion of the procedures the researcher then shows that the simulation models created as part of the research instrument within the third research objective could be 'reverse engineered', to assess the trade group needs of a Trust if they had no reliable raw data.

5.2   First research objective

Evidence was presented within Chapter two against the assumption that plant and equipment failures were totally independent from managerial failings, as the collapse of managerial systems can lead to hazardous conditions quite separate to the type and condition of the infrastructure. It was argued that this had been recognised throughout the world within the Healthcare industry (Emslie, 2005). Which in turn had created a focus towards managerial system design, organisation and operations (Reason 1990; Department of Health, 2000).

The types of managerial failures discussed were split into active and latent, where the former were those actions committed 'at the sharp end' by the operatives that are the immediate causes of the incident. Latent failures however are caused by senior managers and lie dormant within the managerial system for a long time (possibly years) until triggered by an active failure (Ramanujam & Goodman, 2003). At this point they can combine to breach the system's defences. It was the need to expose these latent conditions before failure could occur (Reason, 2005) that led to the design of research instrument within the first research objective. The output of this objective was intended to uncover poor managerial systems relating to normal day-to-day activities, which in

general result in driven failures, not random failures (Tweeddale, 2003), with focus being on the system as opposed to the traditional person centred view (Reason, 2000).

The topics investigated were agreed prior to any analysis by the Trust as posing the greatest unaddressed perceived risk. The scope of the analysis was limited to operations characterised by pre-specified routines and generally accepted norms (Ramanujam and Goodman 2003). In all cases a multi-professional focus group was formed within each Trust, which consisted of those with experience based on the technology and managerial systems in use, and the environment that the system was operated within (Aven & Kørte, 2003; Backlund & Hannu, 2002).

For each focus group the researcher acted in a process consultation role (Schein, 1997; Schein 1998; Belton & Stewart, 2003), mindful of the main influencing criteria that could effect the group's assessment of risk (Cook & Slack, 1991; Tversky & Kahneman, 1974; Mussweiler & Strack, 2000), whilst actively encouraging the participants to challenge their collective risk assessments so as to gain confidence in the robustness of their risk scores (Armstrong, 1985).

In recognition of the failure of the standard criticality grid to adequately facilitate the assessment of risk when applied to a managerial system, due to the inability of people to accurately predict the percentage probability of an outcome (Tversky & Kahneman, 1974; Conrow, 2003), a series of utility functions was developed (Keeney & Raiffa, 1976; USA Department of Defence, 2002). The assessed risk scores resulting from the process were graded into the standard traffic light system of red, amber and green. This methodology created an almost instantaneous 'negative feed-back loop', from the model's initial outputs to the Trust's managers and back into the process to adjust/fine tune the assumptions.

In all cases the results from this research objective were readily accepted by each of the 31 Trusts, as a reasonable representation of 'where they were', in relation to conformance to the NHS national and local acceptable standards. The success of this part of the process hinged on a combination of factors, all being brought together into one package, underpinned by academic research, and where the inputs and outputs were created by local people addressing local needs. In the main, these factors

228

consisted of a structured and common approach agreed by all of the Trust's senior managers, via a multi-professional team, before any data gathering or analysis began. This in itself created an atmosphere that allowed the smooth transition of emphasis from a perceived estates' problem, into one of a Trust responsibility. This focused the various stakeholders onto taking personal responsibility for the issues raised.

The simplicity of the model aided an understanding how the model worked by all concerned. This together with the active encouragement to test various 'what-if-scenarios', with an almost instantaneous 'negative feed back loop', and creating a repeatability of assessed risk scores through the use of utility functions, played a major contribution to the model's acceptance in the opinion of the researcher. When coupled with the model's ease of use, its robustness, ease of interpretation of outputs, together with its clarity in terms of inputs and outputs from Statute law and regulations, down through national guidance to local needs, all combined to produce a unique parcel of techniques, and created an exceptional managerial tool. The acceptance of the process by all of the Trusts involved, and their willingness to redistribute their limited resources based on the output is testimony to its success.

This recognition was further reinforced by the inclusion of the research instrument design created and employed within the first research objective within the demonstration project for excellence in the construction industry, following the recommendations within the Egan report, entitled Rethinking Construction (Egan, 1998). The report was commissioned by the Deputy Prime Minister and published in July 1998. The main theme of the document was that through the application of best practices a collective improvement in performance could be achieved. This led to a series of demonstration projects to exemplify some of the innovations advocated within the report. Project number 234 publicised the redevelopment at the Royal Hospitals Phase 1, Northern Ireland, and the use of the methodology developed within this research. The web site gives a background to the redevelopment and describes the methodology employed to assess the hospital estate department's operational engineering maintenance risks. The report then shows the output risk profile as developed as part of this research (Constructing Excellence Limited, 2007).

As evidence of this achievement the greatest accolade that could be bestowed was the request from the Trusts to rerun the exercise some 18 months later, to revisit their risk profile, confirm their progress towards closing the identified gaps and to facilitate and guide their strategic planning process i.e. an extension of the ' negative feed back loop'. An example of the output from such a series of exercises is shown in Figure 65 below. This shows the movement from the left to the right of the risk profile as the Trust concentrated their limited resources to closing the gaps within the red zone from the year 2001, through years 2003, 2004 and finally 2006. The style of the presentation of Figure 65 is not strictly correct as the x-axis should be multi-layered, but this method of presentation has found favour with the Trusts because of its simplicity.

The results have shown that in all cases where the research instrument within the first research objective was used it uncovered numerous unknown latent conditions of non-conformity within the Trust's managerial systems. These deficiencies ranged from minor to those posing substantial /intolerable risks.

Figure 65   ***The reduction in risk over the years from 2001 to 2006 presented as a sketch for the Belfast Royal Victoria Hospital Group.***
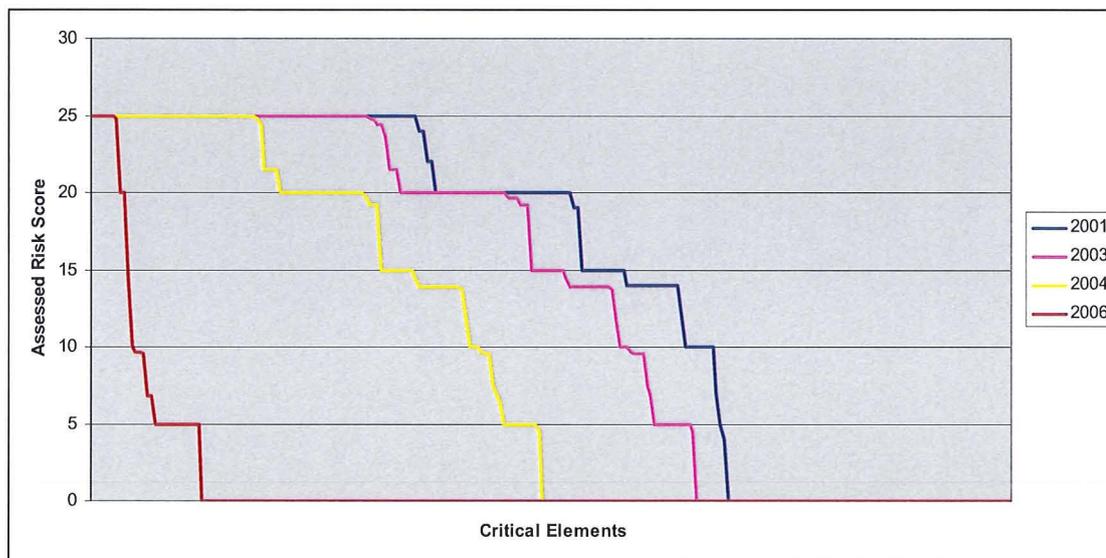
The first research objective process was specifically designed for use within NHS Trust estates departments. However because of its popularity its use has been extended into the areas known as 'soft facilities'. These consist of such topics as cleaning, catering,

transport, pest control, waste management, etc. It is beyond the scope of this thesis to discuss the use of the research instrument developed as part of the first research objective into these other areas further, except to note that due to the robustness of the academic work underpinning the model's design, repeatability, auditability and simplicity of use, the concepts were readily transferred into other areas, and have the potential for further expansion.

These results and their acceptance by each organisation appear to confirm that: the managerial system of internal control within the NHS **does not** give adequate assurance that NHS healthcare building services engineering day-to-day maintenance activities conform to the national guidance. This confirmation led the research to address the second, third and fourth research objectives.

5.3   Second research objective

If the Trust considered that the numbers of identified managerial gaps were small, and the managerial and technical skills needed to close the gaps were also small and not complex, then the organisation did not partake within this second research objective. But if the number of identified gaps was large, and required a considerable managerial resource and complex level of technical skills to close these gaps then the research instrument developed as part of this objective was employed; which occurred in the majority of cases. This methodology was designed with reference to the construction industry and standard project management techniques, as no definitive literature was found that directly addressed the problem of assessing the manpower, budget and timeframe needed to close the managerial gaps uncovered. As such the Trust's focus groups were presented with a Gant Chart (designed by the researcher) that they populated. This guided them to take each gap in turn, break it down into tasks, sub-tasks and in some cases sub-sub-tasks; with each having the manpower and managerial/technical skills assessed alongside their respective estimates of days needed for completion. Again, an almost instantaneous 'negative feedback loop' from the output of this objective to the Trusts' managers and back into the process was created. This allowed those populating the framework with data to quickly adjust/modify the assessments until acceptable.

Estimates of levels of confidence were assigned to the manpower levels and budget needs, which related to a probability distribution. From this a Monte Carlo simulation was run to 'fine tune' the overall estimate. Following this, approximations of time scales to close each of the gaps were made. A natural output from this objective was the creation of small work packages that could be graded in order of priority, and assigned to individual estates officers to complete as part of the Trust's/department's strategic objectives, and/or their Personal Development Plans (PDP's), and/or Personal Objectives. Throughout this objective the researcher again practised the art of 'process consultation'. Teaching the focus groups the techniques developed as part of this research, allowed them to apply their new skills to areas outside of the scope of this thesis (Saaty 2001). Testament to the success of this objective is again from Trusts prioritising and reallocating their limited resources by being guided by the outputs from this objective.

5.4 Third research objective

The output from this objective was a profile that depicted the Trust's assessment of the effects of varying the maintenance revenue resources with respect to the assessed risk (Standards New Zealand, 2005). Continuing the theme developed within previous chapters, Figure 66 shows how this effect was formed via the electrical trade group within site H.

As the available man-hours were increased the simulated risk fell and eventually began to stabilise at a minimum level. Any further increase in resource was found to have negligible effect on the risk being generated. The shape of the graph is a virtual copy of that predicted within the Australian/New Zealand risk management standard (Standards Australia, 2005). Prior to this research the creation of this graph had never been achieved for an actual NHS site. From the development of Figure 66 it would appear that, for the first time in the 60 year history of the NHS estates departments, they were now able to assess the effect of varying resource levels with respect to risk. This in turn would suggest that they could be guided to the optimum level of manpower for their organisation. But as highlighted within the literature review, the words 'lead towards' were used by Sherwin to emphasise that the optimum level was a moving target continuously influenced by new data/environments (Sherwin, 2000). That said, the graph does give a very good indication of the current resource needs. From the researcher's

experience (and that of the Trusts), the rate of change of maintenance need that would effect the 'judgement call' relating to this predicted optimum level of man-power resource was slow. Therefore the graph has been considered by all as a very useful tool.

Figure 66   An estimation of the total risk as the resource (manpower) varies.
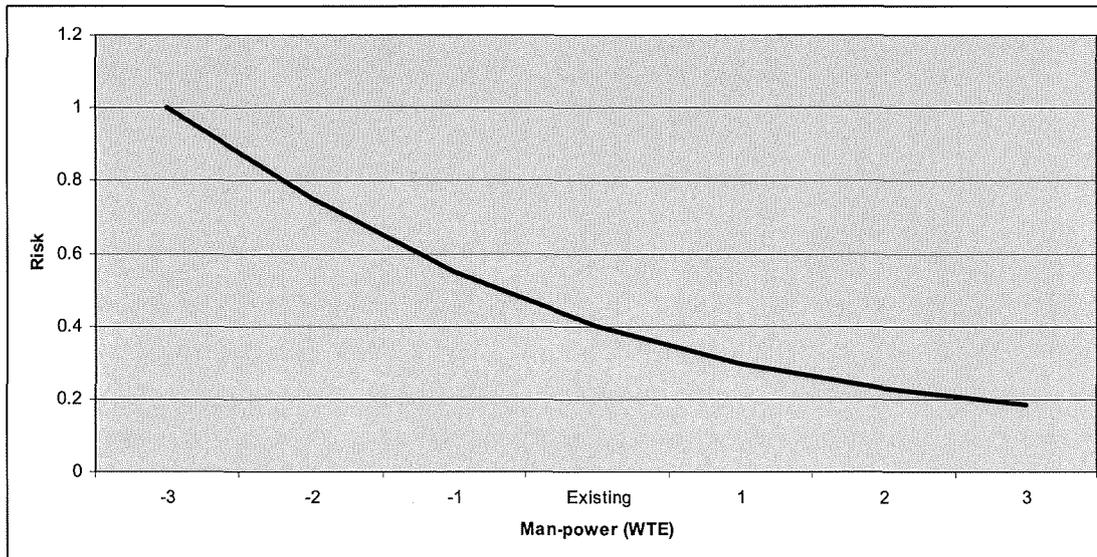
Like the previous two objectives, the methodology designed gave an almost instantaneous negative 'feed back loop'. This again encouraged the Trusts (and the researcher) to probe and question their assumptions and to learn how the processes between resource levels and assessed risk were intertwined. This ability to rapidly question the counterintuitive inputs and outputs of the simulation model and hence overcome the resistance to previous policy decisions, owes much to the speed of these feedback loops. As Sterman discusses in detail, 'learning in and about complex systems is a feed back process' (Sterman, 2000).

The simulation model demonstrated the dynamic non-linear complexity of the real world that reacts in both anticipated and unanticipated ways. It allowed the participants to develop and conduct controlled simulation experiments that would be unethical to conduct in practice. Sterman goes on to argue that prior to developing a systems model (such as that developed within this third research objective), organisations would be

making decisions with limited information, misperceptions and flawed cognitive maps, leading to unscientific reasoning, judgmental errors and biases, all within an environment of personal defensive routines and interpersonal impediments to learning, that leads to the inevitable implementation failures. His detailed arguments supporting such a stance are too lengthy to reproduce within this thesis. With over 30 years of personal experience operating in a senior managerial capacity within the NHS, the researcher supports these statements.

5.4.1 Data gathering

In all cases this part of the exercise uncovered some serious latent deficiencies within the existing planned preventative maintenance schedules. The reasons why this large and important part of the organisation's workload and responsibility had been left to disintegrate to such a state that it had to be rewritten, are beyond the scope of this thesis. But the researcher suspects that it was as a result of the removal of the detailed external scrutiny that accompanied the NHS Labour Management system (Ramanujam & Goodman, 2003; Reason, 2005). However this need for external review does support all of the findings from the outputs from the first research objective which uncovered large areas of latent defects within the managerial systems of estates departments. And it demonstrated to each Trust the need for them to re-design their respective monitoring procedures.

Again in all cases the researcher realised that to push the planned preventative maintenance reviews forward he had to intervene and act as a catalyst to lead the estates officers through the process; but being careful to practice the 'process consultation' techniques as discussed earlier (Schein, 1997; Schein, 1998; Belton & Stewart, 2003). It was during these review meetings that the researcher gained the distinct impression that he was teaching the younger members of the managerial team the fundamental aspects of how to construct the planned preventative maintenance system and the philosophy behind the techniques.

5.4.2 Assessment of impact and probability

The researcher considered that of all of the tasks that could cause a variance in opinion within the third research objective, it was the assessment of impact and probability assigned to each planned preventative maintenance task which would do so most often,

and that if agreement could not be reached, then the entire design of the research instrument would have to be abandoned. In all cases the researcher's concerns were shown to be ill-founded.

During this exercise extensive reference was made to the risk criticality grid descriptors within NHS guidance already mentioned. These were readily understood by the group as a whole, as they had the experience to make qualified judgements relating to the likely impact levels; as such this section passed without major discussion. The researcher expected that it would be the effort to reach agreement about the probability profile shapes that would cause the greatest problem. But each group started slowly, questioning themselves and each other, discussing in detail each situation, until a consensus was reached (US Army Corps of Engineers, 2000; Bilal, 2001; Keeney & Raiffa, 1976). Throughout each of the focus groups were comforted in the knowledge that they had the opportunity to review the profiles selected after the initial simulation, and to observe the effects and significance of any changes. Throughout, the researcher consciously restricted his comments and opinions, practicing the art of 'process consultation', and periodically advising the participants of his role,.

As agreed, the involvement of other professional senior managers within the process was restricted until the estates officers were confident with the base data and simulation outputs. Therefore it was the estates officers that initially assessed the planned preventative maintenance need and the values of impact. Following the successful initial simulations, all planned preventative maintenance tasks and assessments of impact were discussed with the Trust's technical and clinical advisors, their counsel sought, and the simulations re-run. Thus the final output was from the joint consultation and agreement of a Trust multi-professional team, as explicitly required by the relevant national guidelines (NHS Estates 2002). It was the outputs of these final simulations that were presented to the Trust Boards, who under their corporate governance responsibilities, decided the resources to allocate to the subject of day-to-day maintenance, and consequently the level of assessed risk that they were willing to accept.

5.4.3 Examination of the results created from the simulation

Following the successful simulation of the raw data the researcher found that he was in total agreement with all of the academic publications cited within this thesis relating to the modelling of systems, in that it is only when a model is created and used that true knowledge begins to be developed of the processes being investigated.

Sterman argues that no simulation model can be validated or verified, as *verify* indicates truth and *valid* implies being supported by truth. He then goes on to quote various leading figures that have over the years agreed that, 'all models are wrong and are simplifications of the real world'. Therefore the researcher looked for multiple points of contact between the real world and the model, so as to determine the model's limitations (Sterman, 2000). This approach consisted of testing the integrity of the modelling process and the sensitivity of results to uncertainty in assumptions.

The integrity of the methodology was checked against the basic principles of good model design, as recommended by numerous publications for such simulation models, but mainly the comprehensive work by Sterman and Vose discussed earlier. As such the model was designed and used to promote enquiry, without any preconceptions about the output values, or including hidden assumptions. It was used by the widest community possible and supported multiple viewpoints, within focus groups and with direct participation in the process. The flexibility of the model was such that it supported sensitivity and policy analysis, and throughout the empowerment of the Trust managers was promoted to challenge the model's outputs.

Without doubt the simulation model allowed the estates managers (and others), to observe and question the effects of the assessed risk generated from varying resource levels. This enabled them to use evidence based information in their discussions with the other professional groups throughout the Trust, and to advise the Trust board of the consequences of any variation to the existing estates' artisan complement. This in turn allowed them to qualify their decisions with risk based evidence when allocating resources to their day-to-day maintenance activity.

5.4.4 Using the model

As the researcher became familiar with using the model, a common pattern quickly began to emerge when analysing the outputs. For each trade/site the first set of risk profiles that was analysed was for the two test files numbers 1 and 2, to confirm that the program had operated correctly. From hereon in the researcher and the Trust estates officers were confident that the remaining files would give a reasonable representation of the true risk that would be experienced from available man hour variations. Of the various output files, the most common to initially focus on were those that reflected the amber and red risk zones (A & R) for the medium level of WTE. From this starting point all other files were examined. In none of the Trusts examined was it found necessary to extend the simulations beyond the extreme boundaries of plus or minus 3 WTE.

The researcher actively encouraged the individual estates officers to create various personal 'what if scenarios', so that they could learn from the model the effects of policy and procedural decisions. In one case the model was used to predict the likely man-power needed for a site that was to reduce in physical size by approximately 25 percent, due to a clinical rationalisation programme intended to close various buildings. This 'game playing' assisted in gaining confidence with the model in readiness for the presentation of the results to the other senior managers within the Trust.

Once the Trust estates officers were confident with the outputs, preparations were made to present these to the other Trust senior managers outside of the estates function. These presentations were given with much of the technical detail removed. The main output for this was the simulated risk profile that assisted the Trust to make a 'Judgement Call', based on the acceptable level of risk to accept in relation to the risk generated from the day-to-day activities from their directly employed labour staffs. An example of such a profile discussed previously is given below in Figure 67.

237

Figure 67   *Example risk profile showing the Red & Amber zones*
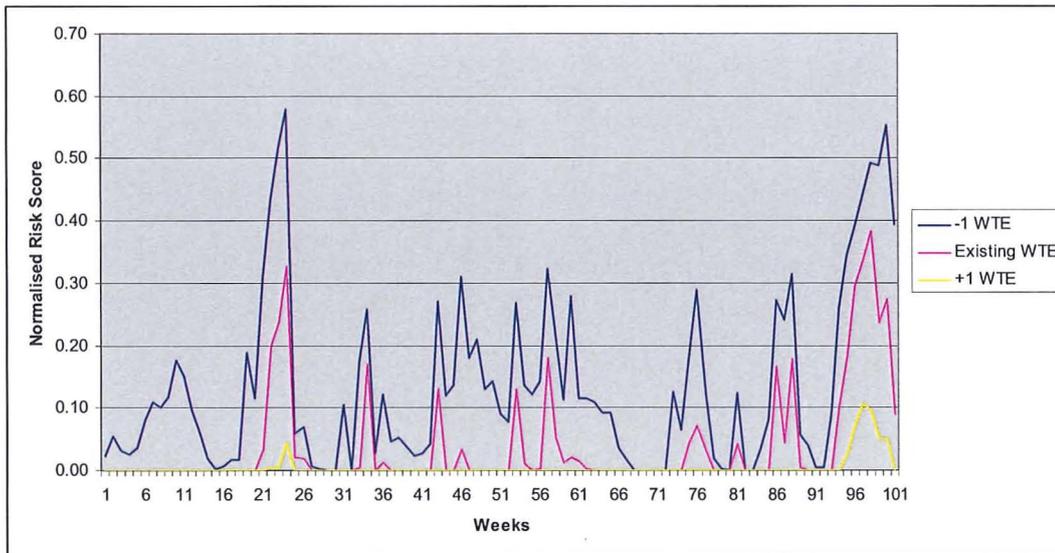
Based on outputs such as these, the Trusts made a 'judgement call', and decided to accept the level of risk generated by the existing directly employed labour resource, or not. In order for them to have the confidence to make such a decision they had to be assured that the simulation model was robust and reflected what would probably occur in practice. This necessitated that the simulation outputs were shown in detail (on request) from the base data onwards. In parallel with this their robustness had to be demonstrated, by including all of the assumptions and variables that could affect the final result.

In all cases the researcher advised the Trusts not to accept the outputs of the model as the definitive projection of risk, and not to immediately implement policy changes. Caution was advocated, so that any change to the existing arrangements should be gradually based on factual information, as opposed to that predicted by the model. It was suggested that small incremental changes be made, and their effects observed and cross -checked against those anticipated. In this way any unknown error within the simulation process would be small and quickly contained. This recommendation was readily accepted by all participating Trusts.

5.5   Fourth research objective

As previously stated the implementation of the methodology related to the development and design of the generic monitoring regime created as part of this research has not been implemented into a Trust, due to the time constraints imposed for the completion of this thesis. But work is now underway to put the system into practice at Trust N, the agreed date for total completion for the estates maintenance department being December 2008.

5.6   The four research objectives

The previous sections questioned the integrity of the process and the sensitivity of results to uncertainty in the assumptions. Through this procedure confidence in the processes was achieved by the consistency of the results and outputs, being in concord with the collective experiences of the focus groups, regarding the likely effects of the variation of resource with associated health and safety risks in the real world.

One important flaw inbuilt into the whole process was that it only assessed the individual health and safety risks associated with each planned preventative maintenance task as a disparate unit. Those risks that were considered were summated into a total risk score. This developed a detailed insight into the separate components that create the risk, and guided managers to concentrate on those areas of concern to develop risk reduction measures. But no estimation of the financial or legal prosecution risks was taken into account. Also the model deliberately ignored the possibility of the risks combining in unforeseen ways and creating an unexpected and greater risk (Tweeddale, 2003). This problem of merging risks has been recognised by the nuclear, petrochemical, aeronautical, offshore platforms and defence industries in their use of Probabilistic Risk Assessment systems, as previously discussed (NASA, 2002; National Patient Safety Agency, 2005; Marx & Slonim, 2003; Wreathall & Nemeth, 2004).

Therefore the simulation model's output could only claim to represent the minimum level of assessed risk in terms of health and safety issues, and indicate that the true risk exposed by the organisation would be greater than this. In the knowledge that all simulation models were inherently flawed, and the limitations of the model determined, the forgoing did give assurance to the Trusts and the researcher that the natural tolerances displayed by the model were acceptable for the environment in which the

model was to be employed. Thus the model was a useful tool to be used within their strategic planning process.

Two unpredicted issues began to emerge during the discussion of the four research objectives. These were: the method of funding the day-to-day maintenance activities of a Trust's maintenance department; the second concerned the possibility of creating synthetic input data for use within the research instrument within the third research objective. In order to protect the integrity of the 'story line' relating to the history of this research programme these two topics are presented within this chapter of the thesis.

5.6.1   Revenue Budgets

Traditionally the day-to-day maintenance activities have been funded with a fixed, as opposed to variable, revenue budget. But as this research has shown, the level of risk generated from a Trust's maintenance activities varies throughout a calendar year; as the reactive maintenance needs increase, the available man-hours to complete the planned preventative maintenance tasks decrease, and the risk rises in consequence. This process is shown within Figure 68 below. This constant interaction and variation between the three waveforms of reactive and planned preventative maintenance needs, and the available man-hours, generates a risk profile that also varies. In order to contain the assessed risk within pre-set levels of acceptability the resource budget must be flexible.

Figure 68   *A simple cogitative diagram showing the effects of increasing the reactive maintenance needs, or available man-hours on the assessed risk generated.*

240
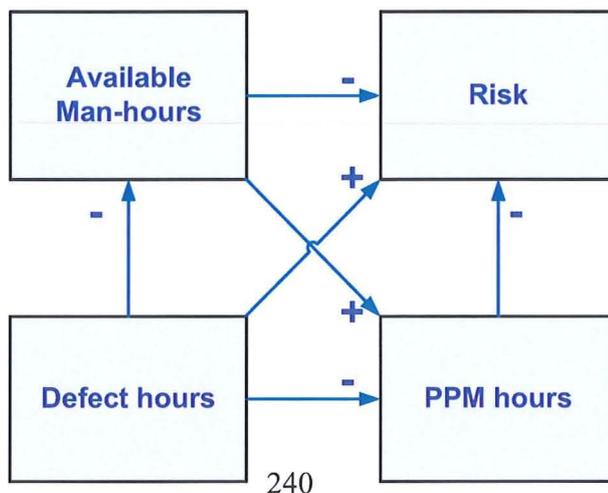
5.6.2   The development of synthetic data

Following the successful use of the simulation programs, the researcher began to question whether the simulation process could be 'reverse engineered', to assess the trade group needs for a Trust, which had no robust man-hour and/or reactive waveform input data. In order to address this question each component of the raw data obtained to-date was analysed

The raw data developed as part of this research consisted of three sets, the planned preventative maintenance, reactive and man-hour waveforms. The addition of the planned preventative maintenance waveform to the reactive waveforms creates the man-hour waveform needs for a site. The planned preventative maintenance waveform is peculiar to each Trust, as each site is individual in terms of its building configuration, layout and building services. Therefore for any simulation this waveform should (and could) be built up from first principles for each individual site, as described within the Methodology Chapter. This left the two unknown waveforms for reactive maintenance needs and the man-hour availability to be created synthetically. In order to develop such a 'reverse engineering' process, the naturally occurring parameters of the two waveforms were sought.

5.6.3   Development of the reverse engineering process research instrument

The base information relating to the reactive maintenance needs for a Trust follow simple requirements. These were itemised by Al-Zubaidi and Christer (1997), as being such tasks as repairing dripping taps, changing light bulbs, etc, and discussed earlier. The researcher considered it acceptable to assume that the rate of generation of these defects and the man-hour needs for their correction would be common from Trust to Trust. This then assumed that a performance indicator could be developed for reactive man-hour needs/m^2/year.

The reactive data from the Trust **X** for each of their two acute sites had been financially audited via the national NHS Labour Management System. This information then formed part of the yearly published statutory accounts for the Trust. Therefore the researcher used this data to form a performance indicator set.

The defect hour waveforms appeared random in nature, as observed by Al-Zubaidi and Christer (1997) who recorded these tasks over a three month period. In this research the observation period was extended to three years, and appeared to confirm their findings of the apparent randomness. This led the researcher to assume that a synthetic reactive waveform could be developed with a random type shape.

It was assumed by the researcher that over time the available man-hour waveform would follow a similar random type pattern to the reactive waveform.

An important part of the analysis was to determine if there was a relationship between the reactive and available man-hour waveforms. Figure 69 below shows graphically these two waveforms for the Trust X electrical trade group on site H as an example.

Figure 69 **The original available man hrs and defect hrs waveforms for the Trust X electrical trade group in site H.**
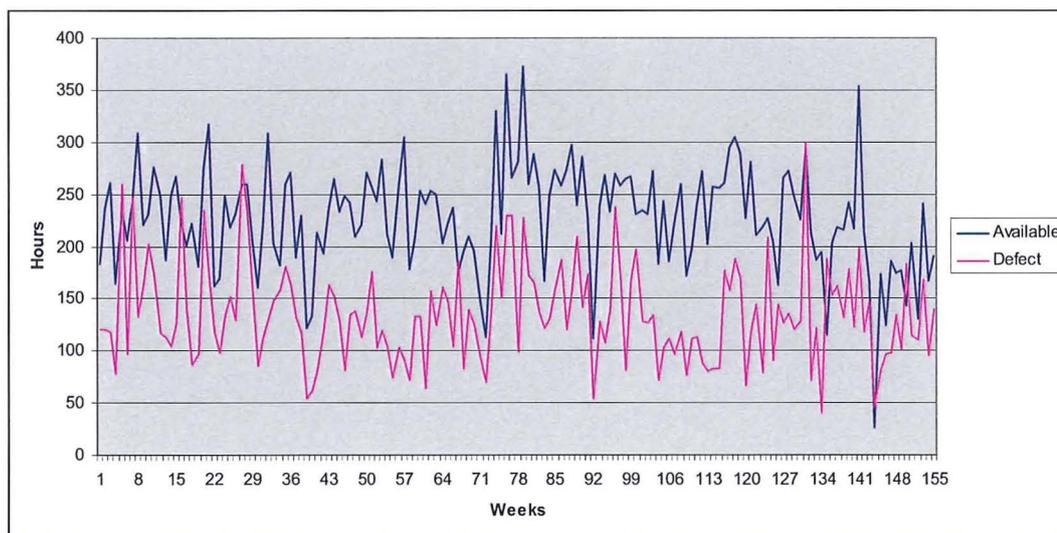
A covariance test was undertaken on the two waveforms for each trade group within Trust X. This test determined if large values in one data set were associated with large values in the other data set. If this were so then it would give a large positive covariance. The reverse would be true if large variations in one data set were matched with small variations within the other data set; then a negative covariance would be computed. If

the covariance were zero than it could not be assumed that there was any relationship. The results of the covariance tests are shown within the table below.

Table 28   *The covariance between the waveforms for reactive and available man-hour waveforms*

Site	Trade	Covariance
H	Electrical	592
H	Mechanical	15
H	Semi-Skilled	19
C	Electrical	171
C	Mechanical	294
C	Semi-Skilled	364

The results within table 28 show that the relationships for each of the respective trade groups indicated a positive weak relationship only. This supported the view that the two waveforms could be synthetically created from random type data.

Any synthetic waveform must reflect the same parameter limits as the original waveform that it was designed to mimic, and to affect the output of the simulation model to a similar degree. The simulation model output was the 'judgement call' made by the Trust's multi-professional group in relation to the staffing level, and the resultant level of assessed risk posed to the Trust, staff, patient, public and other stakeholders. This was a qualitative issue that referenced the level of risk generated, as opposed to a quantitative value. Therefore in order to give a robustness and reliability to the experimental data, the researcher again chose to measure the effect on the total level of assessed risk, summated from the last 10 years of the 20 year simulation model.

As a first step in creating synthetic data it was necessary to describe the original man-hour availability and reactive waveform shapes mathematically, to determine the parameters of the synthetic waveforms. The mean and standard deviations of the original data were computed as a simple aid to describe their properties. From reference to the Trust **X** data, the waveforms for the two sites and six trade groups Table 29 were generated. The planned preventative maintenance waveform data was included for completeness. As discussed earlier the researcher suggested that it was not necessary

243

to synthetically create this waveform as in all cases it could be formed from the data for the site being analysed.

Table 29   ***The mean and standard deviations of the three original waveforms.***

Site	Trade group	Mean			Standard Deviation as a percentage of the mean		
		Man-hours	Reactive	planned preventative maintenance	Man-hours	Reactive	planned preventative maintenance
H	Mechanical	242	147	88	17.0	56.8	25.6
H	Electrical	307	202	62	21.6	34.4	56.8
H	Semi-Skilled	377	191	138	17.3	30.5	27.9
C	Mechanical	240	144	166	22.9	33.7	36.6
C	Electrical	227	134	73	22.1	33.8	24.6
C	Semi-Skilled	249	108	131	22.9	48.5	32.1

In order to test the suggestion that synthetic waveforms could be formed and used to simulate the effects of varying resource levels against an existing or future work load scenario, it was necessary to create the synthetic waveforms, test for covariance, simulate their effects, and to plot the outputs against the original waveforms for the reactive and man-hour figures. If the synthetic waveforms generated a level of risk similar to the original waveforms, and created the same 'judgement calls' relating to the acceptable level of resource/risk ratio, then it would suggest that synthetic waveforms could be used.

As any synthetic data must reflect the natural parameters of the original waveforms, the figures within table 29 suggested that for a given mean value of reactive and available man-hours the synthetic waveforms should lie within the following given parameters:

Standard deviation
The standard deviations of the synthetic reactive waveforms should span between 25% and 60% about the mean; similarly the synthetic available man-hour waveform should be between 15% and 25%.

Shape

From the work conducted previously it was shown that there was little covariance between the reactive and available man-hour waveforms, therefore the shape of these could be created synthetically.

5.6.4 Procedure

From the foregoing, four synthetic waveforms were created, simulated against each other and their outputs from the last 10 years of the 20 year simulation, compared to that of the original waveforms, as shown within Table 30 for all risks within the red and amber zones. In line with the previous presentations, site HRI and the trade group electrical are shown as an example to aid the discussions.

Table 30   *Simulated and original summated risk for the last 10 years of a 20 year simulation for the red and amber zones.*

Defect SD60 and Av_hrs SD25	Defect SD60 and Av_hrs SD15	Defect SD25 and Av_hrs SD25	Defect SD25 and Av_hrs SD15	Original Defect And Av_hrs	WTE
2078	2096	5141	7127	4801	-3
972	985	2651	3563	2405	-2
370	459	988	1086	827	-1
159	216	226	155	171	Acceptable
20	59	13	3	9	1
0	1	0	0	0	2
0	0	0	0	0	3

By normalising the figures within Table 30 the graph in Figure 70 was formed. This shows the variation in predicted risk for various levels of man-hour resource for the synthetic waveforms against that generated from the original waveforms.

Figure 70   *Change in assessed risk for a variation of resource availability for the red and amber zones.*
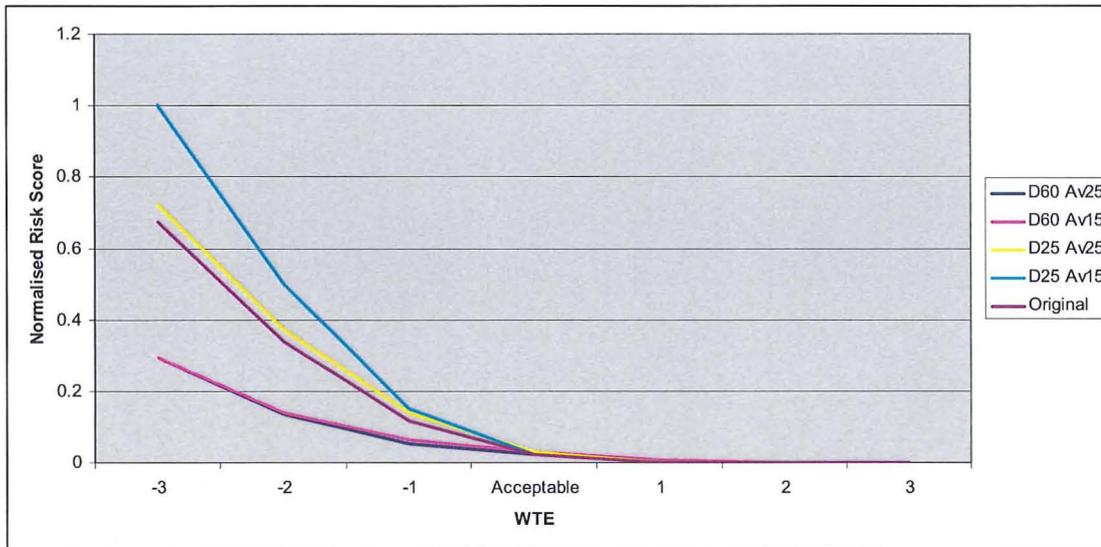
Where the 'judgement call' accepted the generated level of risk for a given resource WTE, the simulated waveforms gave a close approximation to that generated from the original waveforms for reactive resource need and man-hour availability. As the resource was reduced this closeness of fit against the original began to splay out.

These results again suggest that synthetic waveforms could be used to estimate the effects of varying the resource against the assessed risk generated. But in order to give confidence to the assessments a range of waveforms must be simulated and their outputs must converge to a closeness of fit around the acceptable resource predictions.

When the tests were repeated with an increased mean the summated risk increased dramatically. Similar results were obtained for the reactive maintenance waveform and for other groups. Thus, from this series of experiments the researcher was confident that synthetic waveforms could be produced, where no covariance exists between the two waveforms. These results gave a high level of confidence to produce the final graph and output from this, an estimation of the value of risk with respect to a variation in resource levels.

5.7 Follow-up studies

As with any research it is important to devise follow-up studies to confirm (or otherwise) the success of the exercise. The research instruments from the first and second research objectives are being successfully expanded into other non-clinical areas within the NHS. In the case of the third research objective research instruments, such a follow-up process is planned. But unlike the follow up studies for the previous two objectives, the naturally long timescales involved with the implementation of this objective could not be completed until at least two years or more beyond the time frame for this thesis, and thus before any definitive conclusions could be published. The fourth objective relating to the monitoring system is currently being developed for implementation into two NHS Trust estates departments, with the anticipation that further designs will be created for other NHS departments over the forthcoming years.

This problem of follow-up studies was recognised by Sterman (2000), when attempting to ask, "whether the modelling process helped change the system for the better", as it was difficult to determine if it was the model, the process or a combination of external factors that changed people's mental pictures and beliefs.

5.8 Selection of sample Trusts

In the year 2000 financial support for this research was refused by the NHS; the only way that it was allowed to precede was if it was 'self-funding'. In other words only if Trusts were willing to pay for the studies to be conducted within their sites, then the research could continue. Since then this research has survived on this basis up until the present day.

This funding arrangement and the sample of Trusts participating within this study have not affected the validity and/or success of the academic approach to the methodologies employed throughout each of the four research objectives, or their results and contributions made to mankind's 'body of knowledge'.

But without doubt it cannot be stated that the results truly represent a random type sample of NHS Trusts. That said, in all cases the methodology has uncovered previously unknown latent conditions that their own multi-professional managerial teams (comprised of their most senior managers) have categorised as Substantial/Intolerable.

247

When these results are viewed alongside the Department of Health publication in the year 2000 condemning the fact that the NHS is directly responsible for approximately 850,000 adverse events and wasting about £2B in the process (Department of Health 2000); and the National Audit office report in 2003 that cites the Commission for Health Improvement (CHI), where this organisation found that, in a third of the organisations CHI had reviewed, there was a 'lack-of-connection', between the policies that the board had approved and what was actually happening in the wards, departments, etc (Audit Commission, 2003); then a disturbing picture of inadequate internal control procedures emerge.

5.9   Discussion summary

This Chapter considers the research cited within the Literature Search Chapter, the techniques employed as created within the Methodology Chapter, and the robustness of the outcomes as tested within the Results Chapter. Following the successful analysis, discussion and conclusion of the procedures the researcher then summarises the findings.

The outputs from the first research objective were shown to uncover previously unknown latent gaps within the managerial systems. In some cases these gaps were judged by the relevant Trust senior managers to be generating serious/intolerable levels of risk to their organisations, patients, public, staff and stakeholders. The output from this objective allowed the Trusts to refocus their limited resources to 'where they would do most good', Testimony to the value of this output was received by the Trusts requesting follow up studies to confirm progress in closing the gaps, and through national recognition and publication.

The outputs from the second research objective estimated the resources needed to close the managerial gaps identified. This gave a budget expenditure and risk reduction profile against a time line with which to project manage the closure of the managerial failings.

The third research objective outputs allowed the Trusts to determine what revenue resources were needed to keep the risks associated with maintenance to an acceptable level. As a direct consequence of this research, it is now possible for the first time in the

60 year history of the NHS for evidence to be shown of the consequences of varying resource in terms of the effects on risk.

The simulation model specifically developed for this objective shows that even for reasonably resourced organisation, peaks of risk naturally occurred within the red and amber zones with respect to time. This suggested that maintenance revenue budgets should be variable and not fixed year-on-year.

The sensitivity of the simulation model was tested and this demonstrated that only negligible effects were experienced by: the use of interpolation in place of integration within the simulation software to speed its design; the sensitivity of the probability utility to the focus groups being particularly risk adverse; and the focus groups' judgement in assessing the ratio differences between the impact descriptors.

This objective also questioned if the simulation models created could be 'reverse engineered', to assess the trade group resource needs of a Trust if they had no reliable raw data.

This led to the development and use of synthetic data for available man-hours and reactive hours, and showed that it was possible to produce reasonable results from such information to within small tolerance limits of accuracy, when compared against actual values derived from original data.

The fourth objective which was designed to monitor the day-to-day managerial activities was not discussed in depth within the Chapter, as it was not possible to implement the methodology into a Trust within the acceptable time scale of this research programme.

Chapter Six   Conclusions

6.1   Overriding conclusion

This thesis has uncovered major systematic failings within NHS estates departments; whilst some of these failings were unknown, others were suspected to exist. This exposure has been achieved only by moving beyond the simple basic operational performance methodologies in common practice within the NHS, into complex and hitherto unused models of analysis within these types of organisations.

This research began with the suggestion that NHS estates departments were not complying with the recognised NHS codes of practice. This led to the first research objective:

> Develop and implement a robust, transparent and repeatable methodology to assess if the managerial systems of internal control within the National Health Service (NHS) give adequate assurance that the day-to-day building service engineering maintenance activities conform to their respective national guidelines.

The results of this research conclusively show that of the Trusts analysed as part of this research, not one had adequate managerial systems of internal control, and in all cases each Trust's own senior managers (comprised of multi-professional focus groups) considered that:

> The previously unknown systematic failings of their governance systems were exposing their organisation, patients, staff, public and stakeholders to substantial/ intolerable risk.

These failings were not the cause of a momentary lack of focus, but had developed over a considerable period of time, creating a history of systematic failings. This research described the concepts of the performance management procedures specifically developed for this study. It has shown the underlying theory and background to the development of the processes, how they were applied, the results they produced and how this new found knowledge is able to feed into the decision making processes of the organisation.

250

6.1.1 Novel academic contributions to the subject area

First research objective   Use of utility functions

The novel academic contribution of the research instrument developed and successfully employed is the use of utility theory into the subject of risk assessment of a managerial system compared to the standard criticality grids generally in use. The research has shown the ease of use, transparency, robustness and repeatability of the utility function methodologies employed, as the need to assess the percentage probability of failure has been removed and replaced with an assessment of percentage compliance, to assess risk of a partially compliant managerial system when assessed against national and local standards. This has been further enhanced by the adoption of the diagrammatic methodology within focus group situations and the speed of response compared to other risk assessment systems. Although the research instrument has been specifically developed for use within an NHS environment it is applicable to a range of managerial situations external to this organisation.

Second research objective   Use of typical construction project planning techniques

This research has found that it is unusual to employ Gantt charts to break down a managerial deficiency into tasks, sub-tasks and sub-sub-tasks, etc, assign probable levels of resource need in terms of man-power, budget and timescale. The design of the research instrument successfully used for the completion of this objective has formed a novel academic contribution to similar problems faced by organisations, as this technique has applications in a wide variety of establishments.

Third research objective   Use of simulation techniques to assess risk vs. resource.

The novel academic contribution made through the successful development and implementation of the research instrument specifically for this third research objective has been made in the following four areas:

> **Area one:**   The traditional algorithm for the assessment of man-power need for a typical maintenance department has been shown through this research to be fundamentally flawed, and to constantly underestimate the man-power need.

251

Area two: The design and use of a simulation model has shown the previously unknown relationship between risk and values of resource at a macro (week-by-week) level, where it has been shown that due to the random type nature of available resource and reactive maintenance demands, that the risk trespasses into the areas graded as intolerable/substantial and major risk zones (red and amber zones) even at high levels of resource.

Area three: The high level output from the simulation shows for the first time, the relationship between the risks generated from a variation in resource for the maintenance function. Therefore it is now possible for senior managers to simulate their actions 'before' adjusting resource (budgets) in relation to the subject of building services, plant and equipment maintenance against their risk appetite.

Area four: The outputs have conclusively shown that to match the levels of risk appetite from year-to-year an organisation must adopt a flexible budget allocation.

Fourth research objective   Use of a driver

Although the research instrument developed specifically within this objective was not implemented during the timescale of this research, the unique contribution this research has made to the theoretical discussions has been the suggestion to the use of a driver within the monitoring system. This to govern the monitoring process through a web based system, which allocates managerial monitoring tasks to various hierarchal levels of managers both within and outside a department.

Overall contribution of all four research objectives

The uniqueness of the research is summarised within the combined presentation of all four objectives. As never since 1948 has NHS maintenance managers (or others), been able to show the seriousness (or otherwise) of their managerial systems when compared against national and local standards, nor have they been able to demonstrate the influence of resource changes with respect to an assessed risk against a patients safety.

6.2   Personal reflections (A history of managerial failure)

Over the 60 years since the NHS has been formed, each decade has borne witness to a series of initiatives to improve its managerial systems. Since 1978 the engineering function has been subjected to review, where hospitals were placed into a 'league table' as part of a quality control measure of their maintenance activities (Department of Health and Social Security 1973). Additional information was created through the introduction of the self financing engineering incentive scheme, which was commonly referred to as the bonus scheme (National Performance Advisory Group 1997). But this latter information flow ceased when Trusts were formed and the incentive schemes were disbanded.

Controls assurance was introduced into the NHS to improve the internal controls throughout the organisation as a single initiative, which attempted to develop an integrated internal control structure (Scrivens 2005). In 1998 an information strategy was introduced as a strategic approach to the use of information technology (Department of Health 1998). This was followed by the introduction of Corporate Governance around the year 2000, roughly in parallel with the publication that claimed that the NHS was wasting in excess of £2 billion/year, and which estimated that 850,000 adverse events occurred per year directly affecting patients (Department of Health 2000). At this time it was recommended that the NHS should take a systems approach as opposed to a person centred methodology to reduce this waste and number of incidents.

In the year 2002 Trusts were forced to sign a Statement of Internal Control alongside the accounts of the Trust, with the aim of ensuring that the Trust Board was receiving regular assurances that its managerial systems were functioning correctly (Department of Health 2002). But in 2003 the observation was made that at the heart of many public service failures was poor corporate governance. The NHS itself went on to state that there was a 'lack-of-connection' between the policies that the Trust Boards had approved and what was actually happening in the wards, departments, etc. From 198 published clinical governance inspections, it was clear that half of the Boards evidently did not receive information from service activities (Audit Commission 2003).

The Health Care Standards Support Unit was formed in 2004 to provide support to the Department of Health initiative on standards (Scrivens 2005). This organisation effectively operated in parallel with the NHS Litigation Authority, which was originally

253

formed in 1995 to handle claims of non-clinical and clinical negligence against the NHS. It is estimated that the current theoretical cost of paying all outstanding claims immediately would be in the order of £9.09 billion for clinical claims, and approximately £0.13 billion for non-clinical claims (NHS Litigation Authority 2008). In 2005 a call was made for new approaches to be formed towards an integrated audit (Scrivens 2005). In 2006 new guidance was issued to ensure that the basic building blocks were in place for integrated governance by 2008 (Department of Health 2006). The year 2008 is now here, but despite all of the initiatives created, reports commissioned and published, guidance material issued and audits undertaken, this study has shown that:

> None of the Trusts taking part in this research had adequate governance systems in place to advise them (and others), that their estate departments were not performing to the national standards.

6.3   The outputs from this research

This research has successfully broken down this governance problem into fine detail to analyse the activities of a number of estates departments.. Central throughout the whole of the research procedure has been the identification of potential latent errors. Although these errors seldom create adverse consequences themselves, they do create the environments which make such outcomes more likely.

It is from the techniques designed within the first research objective, that NHS Trusts are now able (for the first time in 60 years), to identify and assess their deviations from the predetermined and accepted national practice, and to gauge the level of risk that they pose, using a reliable, simple, repeatable and auditable method. Prior to this, Trusts were trying to apply un-calibrated ordinate scales of probability to a 5x5 criticality grid which was subject to personal and political bias, with no transparency of methodology or repeatability.

In all cases the results from the methods developed as part of this research have uncovered previously unknown latent risks, that the Trusts own multi-professional focus groups have assessed as posing intolerable/substantial risks to their patients, staff, stakeholders, etc. This led directly to a redistribution of their limited resources to, 'where they would do most good'. But to enable this transfer, the resource needs to close each

254

of these managerial gaps in terms of man-power, cost and time had to be assessed; this need directed the research to develop the second research objective.. Prior to this, Trusts were attempting to assess their resource needs for such managerial tasks by applying little more than 'professional guesswork'. The output from this research gave Trusts the structured, detailed and auditable methodology to break down their managerial tasks into fine detail, assign levels of confidence to their assessments, and to produce estimates of need to acceptable levels of accuracy and auditability. This in turn facilitated the creation of future probable scenarios reflecting the reduction of assessed risk with respect to time, from various levels of resource specifically aimed at closing the identified managerial gaps.

Whilst the outputs from the first two research objectives identified the latent conditions creating the risks, and estimated the resource needs to close the managerial gaps to reduce the risks to an acceptable level, they did not give any indication of the revenue consequences needed to keep the gaps closed in terms of the directly employed artisan shop-floor staffs. Therefore this research naturally followed on into this area, and the third research objective was created.

Prior to this the, artisan man-power planning within NHS estates departments relied on a simple crude algorithm, that totalled the trade group yearly hours needed to complete all of the reactive and predictive maintenance work. The results of this research conclusively show that the heuristics supporting this simple model are false, and that in all cases the resulting resource estimates derived from the formula grossly underestimate the acceptable man-power need, and as a consequence create unacceptable levels of intolerable/substantial risk. This research has shown that this phenomenon is due to the previously unknown dynamic complex interchange between the levels of available directly employed artisan resource and workload.

To prove that the simple algorithm was inadequate, it was necessary for this research to develop, design and use a simulation model, which followed the resource availability and need patterns, and work allocation processes. As the technique uses a Trust's actual data, the process captures the effects of the trade's staff being interrupted, and the role of novelty and the subsequent disruption caused to their normal work flow pattern, specific to each Trust site in question. Thus the model reflects the role of resource and

workload quantity. The simulation model also allowed numerous 'what-if-scenarios' to be explored, that would have been unethical to test in practice. The research within this objective also shows that if a Trust's specific data was either corrupted or unavailable, synthetic data could be created and used within the model for man-power availability and/or reactive workload waveforms.

The outputs from this only predict the minimum level of likely generated risk, as no account is made of the increased levels created through a combined multiple component failure effect within the interactive and tight coupling of the healthcare environment. However, due to this research, it is now possible for the first time since the formation of the NHS in 1948, to predict to an acceptable and repeatable level of accuracy, the effect of varying a Trust estate's artisan staff's maintenance budget availability with respect to the resultant generation of risk.

The first three research objectives have successfully designed and implemented techniques to identify and close the managerial gaps caused through previously unknown latent conditions within Trusts' managerial systems. The final and fourth part of the research process was to develop and design a generic methodology to monitor the day-to-day actual activities of an NHS estates department against the pre-agreed national and local maintenance standards. Although there is much NHS guidance literature suggesting that Trusts should implement such a monitoring system as part of their corporate governance procedures, there is no such documentation relating to the practical details of how to develop, design, implement and maintain such a structure.

Throughout the monitoring system framework designed as part of this research are processes that incorporate both risk management and internal auditing, together with joining of risk management ideology with corporate governance tasks. This is in the belief that to be effective the framework must be underpinned with good information management and that this facilitates better risk management, that in turn creates the environment for better decision management. It is from this basic information that scenario planning can be developed to enable strategic and operational decisions to be taken that would cater for all plausible futures.

By creating such an integrated and dynamic assurance framework the need for a traditional audit cycle is eliminated, as information is reviewed regularly throughout the year, and the organisation's exposure to risk and, managers' concerns about risk are constantly being evaluated. This allows the organisation's managerial, technical and general audit functions to be targeted to where the need is greatest, as opposed to following a pre-determined historical calendar of tasks. It also creates the atmosphere for the audit function to move away from a simple process control investigation, into the areas of business processes relating to the generation (or suppression) of significant strategic and operational risk, i.e. the analysis of how the risks are being generated and managed. From such a structured approach the costs associated with operational processes and the associated need for monitoring, verses the impact of failure can be estimated, thus identifying the scope for, and facilitating the removal and burden of, unnecessary monitoring systems.

The advantages from the design of the assurance framework developed within this research are created from combining the risk management and internal/external audit functions into one body, for the determination and understanding of strategic and operational risk, coupled with the individually designed monitoring outputs for each manager in turn, that span both the vertical hierarchal managerial structure of the department, and the horizontal information pathways to other interested groups throughout the organisation. This is achieved by capitalising on the individual actors desire to 'do well', and to improve the services that they provide, and from encouraging ownership by training them in the concepts of risk and governance topics. This enables them to each take an active part in the system's development, to create solutions to their problems.

Due to the time constraints on this research no results were available for discussion and comment from this final research objective. However the researcher is currently developing, designing and implementing this monitoring system assurance framework into a major Trust across all facility groups, including their estates department. The anticipated time scale for completion for the estates section is September 2008.

6.4   Future research opportunities

Although this research is focused towards the engineering day-to-day maintenance function within the NHS, all the research instruments created within this research programme have the potential to be expanded into other departments within the NHS, and into external organisations. Therefore the suggestions for further research building on this thesis inherently develop for this extension into other areas.

The output from the first research objective has shown a demonstrable success story of implementing the system within 31 Trust estates departments (to date). Outside of the scope, but as a result, of this research programme the process has been successfully applied in the following areas within the NHS: catering, cleaning, transport, pest control and linen management. Further research could usefully be undertaken by applying these techniques into numerous fields, particularly the clinical areas of the NHS. This could highlight common managerial problem areas both within professional groups and those that appear across subject boundaries.

The outputs from the second research objective assessed the resources required to close the managerial gaps identified from the first research objective. Further research could follow up the assessments already concluded and compare them with the actual tasks that were undertaken and the resources consumed in terms of man-power, cost and time scale. From this, useful 'performance indicators' could be developed for like tasks and circulated for general use throughout the NHS. Again this development and approach could be expanded into other professional groups.

The simulation model created as part of the third research objective has the potential to be expanded into areas other than purely NHS estate artisan staffs, as other departments (inside and outside the NHS) will operate with similar routine (periodic), novel and non-novel tasks, all of varying quantities. A typical example would be within the domestic cleaning departments of hospitals.

The simulation model currently only analyses the man-power requirement for the directly employed artisan staff groups. It does not attempt to simulate the activities of the department's managerial staffs. But NHS estates departments are no different from the majority of other organisations in terms of their managerial functions, in that they are

comprised in the main of routine (periodic) and non-novel tasks. Currently there is no recognised managerial man-power formula in use within the NHS to assess these day-to-day managerial resource needs. As each organisation has its 'tipping point' beyond which the performance of the department rapidly collapses, it should be possible to construct a simulation model, similar to that within the third research objective of this research, to study the effects of varying resource and assessed risk relating to the managerial function. From the experience gained from the model's development, design, construction and use, it should be possible to expand this approach into other managerial departments.

This research is continuing with the development and implementation of the fourth research objective assurance framework into a large acute hospital's facilities department, with the expected completion date being April 2009. Discussions are also underway to pilot the system within a clinical area and beyond. Whilst the current focus is within the NHS, there is no reason why the system could not be expanded beyond these confines.

By combining the four research objectives when analysing the day-to-day man-power needs for the management function, it should be possible to create indicators that would signal when the available resource level is likely to reach the critical level before the 'tipping point' is breached through a quantity induced crisis, as this type of event may provide an early warning system through an identifiable 'temporal signature' created from the interaction of certain performance indicators.

The outputs from all four objectives are designed around a 'judgement call' made by a multi-professional focus group comprised of the Trust's senior managers. Their individual and combined risk taking/adverse attitude would influence the decisions taken relating to their acceptable resource/risk ratio. Understanding the group's (Trust's) behaviours, attitudes and beliefs would prove useful to those wanting to understand and influence their behaviour. Of particular interest would be to know if there were any relationships between this acceptable resource/risk ratio and the Trust's financial control. Expanding this research still further, an analysis could be undertaken to ascertain if this acceptable resource/risk ratio changed as a result of the passing of the Corporate Manslaughter Act in April 2008, or any other external influence.

Throughout the research information has been collected, analysed, computed and produced to inform the decision making processes employed by the Trust. The richness of this valuable information may be enhanced through the use of 'Fuzzy Numbers', to give a breath of variance to the final preferred option.

6.5 The possible future

April 2008 saw the Corporate Manslaughter and Corporate Homicide Act come into force, which means that Trusts are now criminally liable for any deaths where there is a 'gross breach of the relevant duty of care' in the way that their activities are organised or managed (Parliament, 2007). It is the opinion of this researcher that this single act of legislation will do more to force good governance into the NHS than all of the efforts over the last 60 years, as each person on a Trust board will be seen publically as responsible.

Although outside the scope of this research the resulting techniques for each of the four research objectives can be used for any department within the NHS to assist in the development of good governance systems. The researcher is convinced that it is this four part process that is the bedrock from which good corporate governance systems can be built within the service, and it is what the NHS has needed since 1948.

Reference List

Al-Zubaidi, H. & Christer, A. (1997), "Maintenance manpower modelling for a hospital building complex", *European Journal of Operational Research,* vol. 99, no. 3, pp. 603-618.

Al-Zubaidi, H. (1997), "Assessing the demand for building maintenance in a major hospital complex", *Property Management,* vol. 15, no. 3, pp. 173-183.

Armstrong, J.S. (1985), *Long Range Forecasting,* 2nd edn, John Wiley & Sons, New York.

Audit Commission (2003), *Corporate Governance Improvement and trust in local public services,* 1st edn, Audit Commission Publications, Wetherby.

Aven, T. and Kørte, J. (2003), "On the use of risk and decision analysis to support decision making", *Reliability Engineering & System Safety,* vol. 79, pp. 289-299.

Ayyub, B. 2001, Elicitation of Expert Opinions for Uncertainty and Risks, 1st edn, CRC Press, London.

Backlund, F. & Hannu, J. (2002), "Can we make maintenance decisions on risk analysis results?", *Journal of Quality in Maintenance Engineering,* vol. 8, no. 1, pp. 77-91.

Beckmerhagen, I., Berg, H., Karapetrovic, S. and Willborn, W. (2003), "Auditing in support of the integration of management systems: a case from the nuclear industry", *Managerial Auditing Journal* , pp. 560-568.

Beddall, A. (2006), *Computational methods in physics*, Department of Engineering Physics, University of Gaziantep, Turkey.

Belton, V. & Stewart, T.J. (2003), *Multiple Criteria Decision Analysis An integrated approach,* 1st edn, Kluwer Academic Publishers, London.

Bilal, A. (2001), *Elicitation of Expert Opinions for Uncertainty and Risks,* 1st edn, CRC Press, USA.

261

British Standards Institution (1991), *BS 5760 Part 5:1991 Reliability of Systems, equipment and components: Guide to failure modes, effects and criticality analysis (FMEA and FMECA)*, Redwood Books, Wiltshire.

British Standards Institution (1993), *Glossary of terms used in Terotechnology BS3811,* Fourth edition edn, British Standards Publications, London.

British Standards Institution (2002), *BS EN ISO 19011:2002 Guidelines for quality and/or environmental management systems auditing*, British Standards Publications, London.

BSRIA (2000), *Toolkit for Building Operation Audits Application Guide AG 13/2000*, The Chameleon Press Ltd, London.

Cadbury, A. (1992), *The Financial Aspects of Corporate Governance*, Burgess Science Press, Basingstoke.

Codling, S. (1998), *Benchmarking,* 1st edn, Gower Publishing Ltd, Aldershot, Hampshire, England.

Conrow, E.H. (2003), *Effective risk management,* 1st edn, American Institute of Aeronautics and Astronautics, Inc., Reston, Virginia.

Constructing Excellence Limited (2007)*, Redevelopment at the Royal Hospitals - Phase 1* [Homepage of © 2008 Constructing Excellence Limited, trading as Constructing Excellence in the Built Environment, registered in England No. 4641522], Available online:
http://www.constructingexcellence.org.uk/pdf/case_studies/royal_victoria_hospital.pdf
[Accessed on 9 November 2008].

Cook, S. & Slack, N. (1991), *Making management decisions,* 2nd edn, Prentice Hall, London.

Department of Health (1997), *The new NHS : modern, dependable,* HMSO, London.

Department of Health (1998), *Information for Health An information strategy for the modern NHS,* HMSO, London.

Department of Health (2000), *An organisation with a memory: report of an expert group on learning from adverse events in the National Health Service,* HMSO, London.

Department of Health (2002), *Assurance: The Board Agenda,* 1st edn, HMSO, London.

Department of Health (2003), *Building the Assurance Framework: a practical guide for NHS Boards,* 1st edn, HMSO, London.

Department of Health (2004), *A risk-based methodology for establishing and managing backlog,* 1st edn, HMSO, London.

Department of Health (2005), *National standards, Local action. Health and social care standards and planning framework. 2005/06-2007/08,* HMSO, London.

Department of Health (2006), *Integrated Governance Handbook A handbook for executives and non-executives in healthcare organisations,* HMSO, London.

Department of Health B (2006), *Health Technical Memorandum 02-01 Medical Gas Pipeline Systems,* HMSO, London.

Department of Health and Social Security (1973), *Estmancode Estate Management (Building, Engineering and Grounds) Practices Code for the National Health Service,* 1st edn, Department of Health and Social Security, Welsh Office, Wales.

Dr Foster Research Team (2006), *The Intelligent Commissioning Board,* Imperial College, London.

Dobson, J., Lock, S. and Martin, D. (2006),, *Complexities of Multi-organisational Error Management* [Homepage of Department of Computing, University of Lanchester], Available online: www.dirc.org.uk/publications/inproceedings/papers/82.pdf [Accessed on 9 November 2008] .

Donaldson, C. & Armstrong, J. (2000), *Toolkit for Building Operation Audits application guide AG 13/2000,* 1st edn, BSRIA, The Chameleon Press Ltd, London.

Drummer, G. and Winton, R. (1974), *An elementary guide to reliability,* 2nd edn, Pergamon Press, Oxford.

Egan, J. (1998*), Rethinking Construction,* HMSO, Norwich.

Emblemsvåg, J. & Kjølstad, L. (2006), "Qualitative risk analysis: some problems and remedies", *Management Decision,* vol. 44, no. 3, pp. 395-408.

Emslie, S. (2005), *International Perspectives on Patient Safety. Key information on selected notable patient safety programmes, initiatives and practices in place outside the United Kingdom,* National Audit Office.

Encycogov.com (2002), *What is Corporate Governance?* [Homepage of AcadData], Available online:
http://www.encycogov.com/WhatIsGorpGov.asp
[Accessed on 12 November 2008]

Ericson, C. (2005), *Hazard Analysis Techniques for System Safety,* 1st edn, John Wiley & Sons, Hoboken, New Jersey.

Fear, W. (2006), "High reliability organisations and safer health care", Available online:
http://www.saferhealthcare.org.uk/IHI/Topics/AnalysisandTheory/WhatWeKnow/High+reliability+organisations+and+safer+health+care.htm. [Accessed on 3 February 2007].

Gephart, (2004), Normal Risk, *Organization & Environment,* vol. 17, no. March, pp. 20-26.

Glaser, B. and Strauss, A. (1968), *The discovery of grounded theory,* 1st edn, Weidenfeld and Nicolson, London.

Health and Safety Commission (2000), *Safety of pressure systems : Pressure Systems Safety Regulations 2000,* HSE Books, Sudbury.

Heimann, L. (2005), "Repeated Failures in the Management of High Risk Technologies", *European Management Journal,* vol. 1, no. February, pp. 105-117.

Hirschberg, M. (2000), *The V Model* [Homepage of Software Technology Support Centre], Available online: http://www.stsc.hill.af.mil/crosstalk/2000/06/Hirschberg.html [Accessed on 9 November 2008].

Horner, R. M. W., El-Haram, M.A. and Munns, A.K. (1997), "Building Maintenance Strategy: a new approach", *Journal of Quality in Maintenance Engineering,* vol. 3, no. 4, pp. 273-280.

HM Treasury (2003), *The Green Book Appraisal and Evaluation in Central Government, Treasury Guidance,* HMSO London, Available online:
http://www.epractice.eu/files/media/media1950.pdf
[Accessed: 25 January 2010]

HM Treasury (2004), *The Orange Book Management of Risk Principles and Concepts,* HMSO London.

Johnston, R., Brignall, S. and Fitzgerald, L. (2002), "'Good enough' performance measurement: a trade-off between activity and action", *Journal of the Operational Research Society,* vol. 53, pp. 256-262.

Karapetrovic, S. and Willborn, W. (1998), "Integrated audit of management systems", *International Journal of Quality & Reliability Management,* vol. 15, no. 7, pp. 694-711.

Karapetrovic, S. and Willborn, W. (2000), "Generic audit of management systems: fundamentals", *Managerial Auditing Journal,* vol. 15, no. 6, pp. 279-294.

Keele University (2005), Baxter Award, Available online:
http://www.keele.ac.uk/news/week/archive/2005/0715/0715-01.htm
[Accessed: 9 November 2008]

Keeney, R.L. and Raiffa, H. (1976), *Decisions with multiple objectives : preferences and value tradeoffs,* Wiley, New York.

Kletz, T. (1997), *HAZOP AND HAZAN,* Third Edition edn, Institute of Chemical Engineers, Warwickshire England.

Labib, A.W. (2004), "A decision analysis model for maintenance policy selection using a CMMS", *Journal of Quality in Maintenance Engineering,* vol. 10, no. 3, pp. 191-202.

Marx, D. and Slonim, A. (2003), "Assessing patient safety risk before injury occurs: an introduction to sociotechnical probabilistic risk modelling in health care", *Qual. Saf. Health Care,* vol. 12, no. doi:10.1136/qhc.12.suppl_2.ii33, pp. i33-i38.

'McShee, D. (2007), *Guide to the Corporate Manslaughter and Corporate Homicide Act 2007,* Workplace Law Publishing, Cambridge.

Mendl, M. (1999), "Performing under pressure: stress and cognitive function", *Applied Animal Behaviour Science,* vol. 65, pp. 221-244.

Meyer, M.A., Mniszewski, S. and Peaslee, A. (1989), "Using three minimally biasing elicitation techniques for knowledge acquisition", *Knowledge Acquisition,* vol. 1, pp. 59-79.

Meyer, M. & Booker, J. 2001, *Eliciting and Analyzing Expert Judgment (a practical guide),* 2nd edn, SIAM Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Mjema , E. A. M. (2002), "An analysis of personnel capacity requirement in the maintenance department by using a simulation method", *Journal of Quality in Maintenance Engineering,* vol. 8, no. 3, pp. 253-273.

Moubray, J. (1997), *Reliability-Centred Maintenance,* 2nd Edition edn, Industrial Press Inc., New York.

Mussweiler, T. & Strack, F. (2000), "Numeric judgments under uncertainty: The role of knowledge anchoring", *Journal of Experimental Social Psychology,* vol. 36, no. 10, pp. 495-23.

266

NASA August (2002), *Probabilistic risk assessment guide for NASA managers and practitioners Version 1.1*, Office of Safety and Mission Assurance NASA Headquarters, Washington DC 20546.

National Patient Safety Agency (2005), *Seven steps to patient safety for primary care Step 3: integrate your risk management activity*, Department of Health, London.

National Performance Advisory Group (1997), *NHS Standard Work Measurement Data for maintenance Departments Report 1386/PPM*, HMSO, London.

NHS Audit Committee (2005), *NHS Audit Committee Handbook - Gateway ref 5706,* 1st edn, Department of Health, London.

NHS Estates (2002), *Estatecode*, The Stationary Office, United Kingdom.

NHS Litigation Authority (2008*), NHSLA Risk Management Standards for Acute Trusts,* Department of Health, London.

Noortwijk, J., Dekker, R., Cooke, R. and Mazzuchi, T. (1992), "Expert Judgment in Maintenance Optimization", *IEEE Transactions on Reliability,* vol. 41, no. 3, pp. 427-432.

Ortiz, N., Wheeler, R. and Breeding, R. (1991), "Use of expert judgment in NUREG-1150", *Nuclear Engineering and Design,* vol. 126, pp. 313-331.

Øien, K. (1998), "Improved quality of input data for mantenance optimization using expert judgment", *Reliability Engineering & System Safety,* vol. 60, pp. 93-101.

Parliament (2007), *Corporate Manslaughter and Corporate Homicide Act 2007,* HMSO, London.

Perrow, C. (1999), *Normal Accidents living with high risk technologies*, Princeton University Press, Chichester, West Sussex.

Ramanujam, R. and Goodman, P. (2003), "Latent errors and adverse organizational consequences: A conceptualisation", *Journal of Organizational Behaviour,* vol. 24, no. 7, pp. 815-836.

Rausand, M. (1998), "Reliability Centred Maintenance", *Reliability and Systems Safety,* vol. 60, pp. 121-132.

Reason, J. (1990), *Human error,* Cambridge University Press, Cambridge England ; New York.

Reason, J. (2000), "Human error: models and management", *British Medical Journal,* vol. 320, pp. 768-770.

Reason, J. (2005), "Safety in the operating theatre - part 2: human error and organisational failure", *Quality and Safety in Health Care,* vol. 14, pp. 56-60.

Reason, J. & Hobbs, A. (2006), *Managing Maintenance Error,* 1st edn, Ashgate, Cornwall, UK.

Roberts, K. & Tadmore, C. (2002), "Lessons learned from non-medical industries: the tragedy of the USS Greenville", *Health Care,* vol. 11, pp. 355-357.

Rudolph, J. and Repenning, N. (2002), "Disaster Dynamics: Understanding the role of quantity in organisational collapse", *Administrative Science Quarterly,* vol. 47, no. 1, pp. 1-30.

Russell, J. (2004), "12 Ways to add value to Audits", *Quality Progress,* vol. 37, no. 6, pp. 78-5.

Saaty, T. (2001), *Decision Making for Leaders,* 1st edn, RWS Publications, Pittsburgh.

Sagan, D. (2004), "Learning from normal accidents", *Organization & Environment,* vol. 17, no. March, pp. 15-19.

Saleh, J.H. and Marais, K. (2006), "Highlights from the early (and pre-) history of reliability engineering", *Reliability Engineering & System Safety,* vol. 91, no. 2, pp. 249-256.

Schein, E.W. (1997), *Process Consultation: Vol 2: Lessons for managers and consultants,* 1st edn, Addison-Wesley.

Schein, E.W. (1998), *Process Consultation: Vol 1 Its role in organisation development,* 2nd edn, Addison-Wesley.

Scrivens, E. (2005), *Quality, risk and control in health care,* 1st edn, Open University Press, Berkshire, England.

Sherwin, D. (2000), "A review of overall models for maintenance management", *Journal of Quality in Maintenance Engineering,* vol. 6, no. 3, pp. 138-164.

Spira, L.F. and Page, M. (2003), "Risk Management The reinvention of internal control and the changing role of internal audit.", Accounting, Auditing and Accountability Journal, vol. 16, no. 4, pp. 640-661.

Stamper, R. (2001), "Measurement or the semantics of numerical information", Management Decision, vol. 39, no. 7, pp. 583-593.

Standards Australia and Standards New Zealand (2004). *Risk Management.* AS/NZS 4360:2004. www.riskmanagement.com.au [Accessed on 9 November 2008].

Standards Australia (2005), Risk Management Guidelines Companion to AS/NZS 4360:2004, Standards Australia International Ltd and Standards New Zealand, Sydney and Wellington.

Sterman, J. (2000), Business Dynamics: Systems thinking and modelling for a complex world, 1st edn, McGraw-Hill, USA.

Strategy Unit Report (2002), Risk: Improving government's capability to handle risk and uncertainty, HMSO, London.

Taylor, F. (1911), The Principles of Scientific Management. Available online: http://en.wikipedia.org/wiki/Frederick_Winslow_Taylor

[Accessed on 9 November 2008].

Teigen, K. (1994), "Yerkes-Dodson: A law for all seasons", *Theory & Psychology,* vol. 4, no. 4, pp. 525-547.

The Institute of Chartered Accountants (1999a), *Internal Control: guidance for Directors on the combined code,* The Institute of Charted Accountants, London.

The Institute of Chartered Accountants (1999b), *Implementing Turnbull A Boardroom Briefing,* 1st edn, The Centre for Business Performance, London.

The National Health Service 2009, 23rd November 2009-last update, *History of the NHS* [Homepage of The NHS], [Online]. Available: http://www.nhs.uk/tools/pages/nhstimeline.aspx [2009, 27th November 2009] .

Tversky, T. & Kahneman, D. (1974), "Judgement under Uncertainty: Heuristics and Biases", *Science,* vol. 185, pp. 1124-31.

Tweeddale, M. May (2001), "Nourishing and poisoning a 'safety culture'", *Trans I ChemE,* vol. 79, no. B, pp. 167-173.

Tweeddale, M. (2003), *Managing Risk and Reliability of Process Plants,* 1st edn, Gulf Professional Publishing, USA.

U.S. Army Corps of Engineers 2000, *Methods for Expert-Opinion Elicitation of Probabilities and Consequences for Corps Facilities,* Institute of Water Resources, Alexandria.

USA Department of Defence (2002), *Risk management guide for DOD acquisition,* 5th edn, Defence acquisition university press, Fort Belfoir, Virginia.

Van der Schaaf, T. (2002), *"Medical applications of industrial safety science",* Quality and Safety in Health Care, vol. 11, pp. 205-206.

Waring, J., McDonald, R. & Harrison, S. 2006, *"Safety and complexity Inter-departmental relationships as a threat to patient safety in the operating department"*, Journal of Health Organisation and Management, vol. 20, no. 3, pp. 227-242.

Wharrad, H. (2004), Levels of Measurement [Homepage of University of Cambridge],. Available online:
http://www.ucel.ac.uk/showroom/levels_of_measurement/downloads/levels_notes.pdf
[Accessed on 12 November 2008].

Wreathall, J. and Nemeth, C. (2004), "Assessing risk: the role of probabilistic risk assessment (PRA) in patient safety improvement", *Quality and Safety in Health Care,* vol. 13, pp. 206-212.