The Development and Application of Novel Intelligent Scoring Systems in Critical Illness

Dr. Malcolm A. B. Sim

B.Sc.(Hons.), M.B.Ch.B.(Commended), F.R.C.A., F.R.C.P., F.F.I.C.M., E.D.I.C.

Submitted in fulfilment of the requirements for the degree of Doctor of Medicine

School of Medicine

University of Glasgow

June 2015

Abstract

Scoring systems in medicine are not a new concept. There are examples from the early 1950s, from around the same time as the polio epidemic in Copenhagen resulted in the birth of modern Intensive Care. Many scores have subsequently been developed specifically for Intensive Care patients. The majority summarise the overall physiological state of the patient in a variety of different ways.

A clinical interest in ascertaining whether haemodialysis causes cardiovascular instability in Intensive Care patients led to an initial simple experiment examining stability using a small number of cardiovascular parameters. It became apparent that to answer the question properly a physiologically based score which could be calculated automatically in real time, and which took into account the level of physiological or pharmacological support the patient was receiving would have to be developed, to counter or to mitigate the drawbacks of the main scoring systems in common use at the time.

This thesis describes the development and first stage in the validation of a novel physiologically based scoring system for Intensive Care patients which overcomes some of the major disadvantages of existing scores. The score was then used to investigate other clinical questions. Myocardial damage in Intensive Care is common and associated with a poor outcome. Aspects of the developed score were used to ascertain if it is possible to detect and predict myocardial damage occurring in Intensive Care patients based on physiological disturbance rather than a rise in biomarkers. The score was subsequently used to examine Intensive Care patient outcomes.

The introductory chapter describes the history of Intensive Care, the mechanism of data collection for patients in Scottish Intensive Care Units and its analysis to enable comparison of different units. Reviewing currently available scoring systems places this work in context and highlights the need for a new score. An overview of renal replacement therapy modalities follows, as an interest in cardiovascular stability during haemodialysis led to the idea for a new scoring system. Myocardial damage in Intensive Care patients is common and indicative of poorer outcomes. This is reviewed, as the developed score was used to detect and then predict where myocardial damage was occurring in critically ill patients, based on physiological disturbance rather than on raised biomarkers.

In Chapter 2, data from dialysis sessions in critically ill patients was collected, prc-processed, and analysed for cardiovascular instability. Using an arbitrary definition of

instability as a 20% change in mean arterial pressure or heart rate in either direction, 65% of dialysis sessions were stable and 35% unstable. This simple experiment suggested that haemodialysis is less cardiovascularly destabilising than previously believed. However a major deficiency was the lack of consideration of the level of physiological support required during dialysis. To investigate this and other clinical problems better, it became apparent that a new score would have to be developed.

Chapter 3 describes the development of a novel quantitative score which takes into account the amount of physiological and pharmacological support a patient is receiving. Physiological parameters were separated into those recorded regularly and those recorded intermittently. They were subsequently divided into ranges, scoring increasing points depending upon the degree of derangement. Ranges were based on an extensive literature search, currently available scores, and clinical opinion. Two key parameters viz. mean arterial pressure and oxygen saturation, were then weighted against a range of factors which can either increase or decrease their value. A score of instability could then be calculated by adding points for the weighted and unweighted parameters. After reflection using common clinical scenarios, some of the points scored in different ranges and weightings were revised to give the final quantitative score.

In Chapter 4, the quantitative score was tested against data sets from actual Intensive Care patients to produce graphs of overall cardiovascular stability against time. Although this approach did capture improvements and deteriorations it had several disadvantages. It captured the expertise of a single clinician only, gave an arbitrary number which could be difficult to interpret, and the emphasis given by the clinician to the relative importance of different physiological or pharmacological parameters would not be obvious to others. Clinical reflection led to a new approach to the problem, viz. the development of the 5 point qualitative scale described in Chapter 5.

Chapter 5 describes the development of a 5 point qualitative score for cardiovascular instability, underpinned by complex physiological rules, and capturing the expertise of several senior Intensive Care Clinicians. This is the Intensive Care Unit - Patient Scoring System (ICU-PSS). I scored data sets comprising thousands of predominantly hourly commonly recorded physiological and pharmacological parameters on a 5 point scale of cardiovascular stability (A to E). I also described rules in the form of different parameter ranges to indicate why I had scored time points as stable (A) through to unstable (E). These rules were incorporated into a computer programme which scored unseen data sets which I

also then scored. The computer's predicted A to E score based on these rules and my own score were compared in a confusion matrix. Mismatches with the computer prediction (based on my initial rules) were analysed and I either rescored the data if I considered that I had not assigned the correct level of instability, or modified the rule base. Through this process clinical expertise was better captured. This process was repeated with two other clinicians using my rules as a starting point. This led to further refinements of the rule base. The result was a sophisticated set of rules underpinning a 5 point, easily understandable scale of cardiovascular stability crystallising the expertise of 3 senior Intensive Care clinicians.

The ICU-PSS was tested in a discrimination experiment to ascertain if clinicians could agree with the score moving in a one step and two step change. This is the first stage in full validation of the score

In Chapter 6, the first stage in the validation of the ICU-PSS is described, using 10 clinicians from a city teaching and a district general hospital. It was hypothesised that if they were shown two consecutive hourly time points of physiological data from real patients and asked whether they were improving or deteriorating, they should agree with the ICU-PSS score in more than 50% of cases (random chance). In two discrimination experiments the consultants were, in random order, shown 4 examples of each type of two step improvement or deterioration in the score, e.g. A to C, and 4 examples of each type of one step change, e.g. E to D. In the two step experiment there was 92.9% agreement with the score, and in the one step change experiment, 90.9% agreement. Both were highly statistically significant.

Chapter 7 describes the first of the applications of the validated score. Myocardial damage is common in Intensive Care patients and is an independent risk factor for both short and long term mortality. The mechanism in Intensive Care patients is likely to be the so-called type II damage caused by extremes of physiological derangement leading to a myocardial oxygen supply and demand imbalance. I hypothesised that it should be possible to use aspects of the score to confirm and subsequently predict where this damage is occurs based on physiological disturbance alone rather than on a rise in cardiac biomarkers. Two clinicians agreed that a subset of the level E, D and C rules from the ICU-PSS occurring in 3 out of 5 consecutive time points would represent conditions likely to lead to myocardial damage in the critically ill. Data sets with known sequences of troponin rises were scanned to ascertain if the above conditions were met around the time of a troponin rise within a

sequence of troponin rises, given the natural decay of troponin. This was indeed the case in 75.8% of cases (95% CI: 57.7% to 88.9%). Similarly this set of conditions was applied to the same data sets, looking at time periods before a first troponin rise. These conditions were met in 87.5% of cases (95% CI: 61.6% to 98.1%). However, as the confidence intervals are wide (and also for the positive and negative predictive values of these tests), this early work is at best hypothesis generating. It will have to be repeated using much larger data sets.

In Chapter 8, the correlation between the mean ICU-PSS score and outcome was examined. A data set of patients was prepared from Ward Watcher with an approximate 50:50 split of medical and surgical diagnoses. The physiological data from these patients was extracted from CareVue and anonymised. A mean ICU-PSS score was calculated for different points during the patient stay. The data were analysed to ascertain if there were differences in mean ICU-PSS scores at different time periods among the survivors and non-survivors within the medical and surgical groups. There is a suggestion that the mean scores are different in certain patient groups between survivors and non-survivors. However, at the time this work was undertaken the computing system used was not yet able to apply appropriate statistical tests. Future work will focus address this problem and also examine the different proportions of the patient's stay spent in different categories of the score. This would avoid the difficulties above of converting ordinal to numerical data.

In a final analysis I ascertained the relationship between degree of any troponin rise and outcome, in the population of patients at Glasgow Royal Infirmary. In a study of 100 consecutive patients, troponin rises were grouped into three categories. These were low (0.04-0.19), medium, (0.2-1.99) and high (≥2.0 micromoles/litre). Intensive Care mortailty was 13.3%, 22.7% and 40% respectively. This association is consistent with findings from similar studies elsewhere in the literature.

In summary, I have developed a quantitative score of cardiovascular stability, and have developed, and partially validated, a more effective qualitative score for use in Intensive Care patients. I believe it overcomes the salient disadvantages of other currently available scores. I have demonstrated that it may be possible to confirm the presence of, and detect, where myocardial damage is occuring. Work thus far suggests that there may be an association between this score alone and outcome. Future work will focus on translating the score into a bedside monitor to give a continuous reading of the overall physiological state of the patient, to detect deterioration before it becomes clinically obvious.

Characteristic patterns of deterioration associated with impending myocardial damage will be displayed at the bedside with the prospect of earlier intervention aimed at preventing myocardial damage and its associated poor outcome.

Contents

List of figures

List of tables

List of abbreviations

ACHE: Architecture for Clinical Hypothesis Examination

AIDS: Acquired Immune Deficiency Syndrome

APACHE: Acute Physiology and Chronic Health Evaluation

APS: Acute Physiology Score

BP: Blood Pressure

CI: Cardiac Index

CO: Cardiac Output

CPAP: Continuous Positive Airway Pressure

CVP: Central Venous Pressure

DO2: Oxygen delivery

ECG: Electrocardiogram

HDU: High Dependency Unit

HR: Heart Rate

ICU: Intensive Care Unit

ICU-PSS: Intensive Care Unit-Patient Scoring System

ISD: Information Services Division

LiDCO: Lithium Dilution Cardiac Output

MAP: Mean Arterial Pressure

MODS: Multiple Organ Dysfunction Score

MPM: Mortality Probability Model

NHS: National Health Service

OECD: Organisation for Economic Co-operation and Development

PaO2: Partial Pressure of Oxygen in Arterial Blood

PEEP: Positive End Expiratory Pressure

POSSUM: Physiological and Operative Score for the enumeration of Mortality and Morbidity

SAPS: Simplified Acute Physiology Score

SICSAG: Scottish Intensive Care Society Audit Group

SIRS: Systemic Inflammatory Response Syndrome

SMR: Standardised Mortality Ratio

SOFA: Sequential Organ Failure Assessment

SpO2: Oxygen Saturation

SV: Stroke Volume

SVV: Stroke Volume Variance

SVR: Systemic Vascular Resistance

SVRI: Systemic Vascular Resistance Index

TISS: Therapeutic Intervention Scoring System

TRISS: Trauma Injury Severity Score

Acknowledgements

I would like to thank the following most sincerely for their assistance in this research.

Prof. John Kinsella (supervisor)	Professor and Head of University Section of Anaesthesia, Pain and Critical Care Medicine 2nd Floor, New Lister Building, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow, G31 2ER
Prof. Derek Sleeman (Collaborator)	Emeritus Professor, Computing Science Department, Meston Building, The University, Aberdeen, AB24 3FX
Dr. Laura Moss (Collaborator)	Research and Development Healthcare Computing Scientist, Department of Clinical Physics and Bioengineering, 2nd Floor, New Lister Building, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow, G31 2ER

I would also like to thank my consultant colleagues in the Intensive Care Units of the Royal Infirmary and the Western Infirmary (Glasgow), in Crosshouse Hospital (Ayrshire) and in Monklands Hospital (Lanarkshire), who so generously gave of their time to score data sets and test the scoring systems developed in this thesis. Without their assistance this work would not have been possible.

Glasgow Royal Infirmary:
Dr. Malcolm Booth, Dr. Lindsay Donaldson, Dr. Charlotte Gilhooly, Dr. Martin Hughes, Dr. Alex Puxty, Dr. Tara Quasim

Western Infirmary:
Dr. Mo Al-Haddad, Dr. Sandy Binning, Dr. Louie Plenderleith, Dr. Sarah Ramsay, Mr. William Tullett

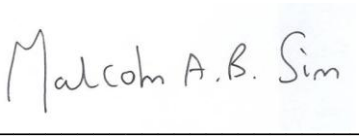
Crosshouse Hospital:
Dr. Jane Chestnut, Dr. Phil Korsah, Dr. Paul McConnell, Dr. Hugh Neil, Dr. Peter O'Brien

Monklands Hospital:
Dr. Pamela Dean

Finally, I would like to thank Mrs. Janette McBride for her indispensible secretarial and administrative support and Mr. John McWilliams for his invaluable information technology support.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature: 

Printed name: Dr. Malcolm A.B. Sim (11/6/15)

Dedication

I would like to dedicate this thesis to my wife Jennifer and to my parents Alasdair and Yvonne. I thank them for their unstinting support during the time spent undertaking the research this thesis required.

Chapter 1: Introduction

1.1. Abstract

Intensive Care is a relatively new specialty within the history of medicine. It is an expensive and precious resource. It relies more heavily than other disciplines on scoring systems to predict outcomes, guide therapy and compare the performance of different Intensive Care Units. Currently available scores have certain limitations. This thesis describes the development of a new quantitative and qualitative scoring system for critically ill patients, the first stage of validation of the qualitative score and in its clinical applications.

To set the work in its context, the history of critical care is described. There is a review of currently available and historical scores for use in the critically ill. The reasons why a new score might be useful are put forward. The mechanism of data collection for patients in Scottish Intensive Care Units and its reporting are included. As the work involved a collaboration with computing scientists a section is included on machine learning in healthcare with respect to the handling of very data sets. The process of validation of a new score in the absence of a gold standard is challenging. Methodology for validation is described.

The idea for a new score arose from my interest in, but an inability to characterise properly cardiovascular stability during haemodialysis. A review of renal replacement therapy in the critically ill is therefore provided, with a specific comparison of haemodialysis and haemofiltration.

The newly developed score was tested in a clinical practice. Initially this was to ascertain if it were possible to detect and predict where myocardial damage was occurring in Intensive Care patients from physiological disturbance alone (and confirmed by cardiac biomarkers. This is relevant as it is associated with a poor outcome. Therefore the introduction concludes with a review of myocardial damage in Intensive Care.

1.2. Definition of an Intensive Care Unit

An Intensive Care Unit (ICU) is a "*geographically defined area in the hospital providing care for critically ill patients with specialised personnel and complex equipment*" [1]. This is the definition by Vincent et al. in The European Society of Intensive Care Medicine "Guidelines for the utilisation of intensive care units." They further define the need for admission to such an area if patients have an "*unstable condition with impaired organ function"* or a "*high risk of developing serious and preventable complication*s". They should not be admitted if they have "*no chance of recovering to a reasonable quality of life*".

1.2.1. Levels of patient care

In 2000 the Department of Health published a review of adult critical care services in which it went further, defining the levels of care patients need during critical illness, while recovering from critical or at risk of critical illness [2]. A summary of these levels is as follows:

"Level 0 *Ward level care in an acute hospital.*

Level 1 *Patients at risk of deteriorating or those stepped down from higher levels of care who can be cared for on an acute ward with support from the critical care team.*

Level 2 *Patients requiring support for a single failing organ system, those requiring more detailed observation or those stepping down from level 3 care.*

Level 3 *Patients requiring advanced respiratory support or basic respiratory support plus support of two or more failing organ systems".*

The definitions were further refined by the Standards Committee of the Intensive Care Society of the United Kingdom [3]. This enabled the definitions to reflect the Critical Care minimal data set collected in England, which comprises 34 fields of administrative and clinical data, allowing it to analyse activity and guide capacity planning [4]. The refined definitions of the Intensive Care Society include detailed examples of what is meant by each level of care. They also highlight that the level of care which patients receive is not related to their location.

1.3. The history of Intensive Care

Critical care practised in such an organised and structured manner is a relatively new phenomenon. However, the recognition of the importance of organ support has much earlier roots.

In both World Wars various techniques were used to resuscitate injured soldiers on the battlefield. It was as a result of the 1952/3 polio epidemic in Copenhagen, Denmark that modern intensive care was born. H. Lassen described the treatment of the associated respiratory failure with "iron lung" negative pressure ventilators in a landmark paper in the Lancet [5]. Mortality was high at over 80%, and the Blegdan Hospital was overwhelmed. On the recommendation of a colleague, Mogens Björneboe, Dr. Lassen contacted the physician-turned-anaesthetist, Dr. Björn Ibsen. He described positive pressure ventilation via a tracheostomy in a 12 year old girl who was deteriorating with negative pressure ventilation [6, 7]. It is worth noting that Asclepiades of Persia is credited as the first person to perform a tracheostomy as long ago as 124 AD although the term tracheostomy itself was first used by Thomas Feyens (1567-1631)[8]. The girl lived and the management of these patients was changed to a high tracheostomy just below the larynx, regular suctioning or postural drainage of secretions, and positive pressure ventilation via a cuffed rubber tube. Mortality fell in this polio epidemic to 40%.

Following the successes in Denmark, there was widespread adoption of these techniques for the treatment of respiratory failure from a number of causes. Further advances in the management of other aspects of the critically ill followed, including relatively better monitoring. However, in the UK at this time, arrangements were not structured for the care of the critically ill, often occurring in side rooms on wards where primitive ventilators were moved to the patient, but with no dedicated medical staff to look after them. This is described in a transcript of a Witness Seminar held at University College London by the Wellcome Trust Centre for the History of Medicine in 2010[9].

There was an increasing recognition of the inadequacy of care, and recommendations were made in a 1962 Department of Health publication "Progressive Patient Care" [10]. One of the recommendations was that "*between 2% and 5% of a hospital's acute beds should be earmarked for care of patients who were severely ill or required specialist acute care*". It also recommended that patients should be grouped together, and treated according to their level of dependency.

Financial support to hospitals followed, and the first Intensive Care Units were established. Greater understanding of the physiological disturbance resulting from disease processes, its manipulation and correction, followed in the 1970s to 80s. This was paralleled by large advances in monitoring techniques and equipment.

1.3.1. Modern Intensive Care

As of 2012 Critical Care patients in Scotland are managed in 25 Intensive Care Units with 274 actual beds. This figure includes cardiothoracic and neurosurgical capacity. There is funding in place for the equivalent of 183.8 of these beds [11]. It is worth noting that in the UK in 2010 health funding accounted for 9.6% of gross domestic product which was higher than the Organisation for Economic Co-operation and Development (OECD) average [12]. Despite this in 2010 the number of acute care beds per 1000 population was 2.4 compared to the OECD average of 3.4. Specifically for Intensive Care in 2010 there were 3.5 ICU beds per 100 000 of the population. This is fewer than our European neighbours e.g. Germany had 24.6 ICU beds per 100 000 of the population in the same year [13].

In Scotland, during the last period for which data are available (2012), a total of 13103 patients were admitted to ICUs or combined ICU / High Dependency Units (HDU) and 26977 patients to HDU. Mean bed occupancy in ICUs or combined units was 71.4%. Of the patients admitted to ICUs and combined units 20% died before hospital discharge [11].

1.3.2. Description of Glasgow Royal Infirmary Intensive Care Unit

The majority of the work leading to this thesis was carried out at Glasgow Royal Infirmary. This is a teaching hospital of over 1000 beds serving the population of the North East of the city. It has a 20 bed Intensive Care Unit which deals with all common medical and surgical conditions needing Intensive Care admission. The regional burns service is located in the Royal Infirmary, as is that for hepatobiliary disease, and in particular the management of complicated pancreatic conditions. As a result the ICU is also a tertiary referral centre for these specialties.

1.4. The need for Scoring Systems in Intensive Care

The cost of Intensive Care is high. The average expenditure on a patient who is admitted to Glasgow Royal Infirmary ICU is £4828 [14]. In the period from April 2011 to March 2012 the Royal Infirmary ICU admitted 1048 patients, accounting for 5137 bed days with a total cost of £5,059,447 [14]. This equates to a total cost per bed day of £984.90 with an average length of stay of 4.9 days [14]. These sorts of costs are repeated throughout the country. Intensive care is therefore an expensive and precious resource.

It would clearly be unethical to randomise a critically ill patient to receive or not receive critical care support. Given the enormous cost of this resource a number of scoring systems have been developed to aid decision making on suitability for admission, quantifying

disease severity, predicting outcome, comparing performance between different units, improving quality of patient care and as research tools.

1.4.1. Mechanism of data handling and collection for scores in Scotland

A national audit infrastructure exists to collect and analyse the data which are fed into a number of commonly used scores. This is managed through the Scottish Intensive Care Society Audit Group (SICSAG) founded in 1992 [15]. A national database has existed since 1995.

Data are collected prospectively using the purposely designed Ward Watcher system. Included are all general adult Intensive Care Units, approximately 90% of High Dependency Units and all Combined Units [11]. These data are validated by the Information Services Division (ISD) of the NHS National Services Scotland [16]. Missing data or queries about e.g. patient outcomes, discharges and treatment are identified and highlighted to individual ICUs by local and regional audit coordinators. All of the SICSAG data between 1998 and 2012 have been linked to the SMR01 data set held by ISD. This data set relates to general and acute inpatient day cases. Every patient appearing in the SICSAG database should have an SMR01 entry relating to the same admission. The advantage of this linkage is data that is enhanced by providing fields such as hospital or overall outcome. As of the 2013 SICSAG audit of critical care, reporting on the year 2012, overall 96% of SIGSAG entries have been linked to the equivalent SMR01 entry [11].

The most commonly used score into which data are fed is APACHE II (Acute Physiology and Chronic Health Evaluation). This will be fully described in a review of scoring systems to follow. One of the end products of this score is an ability, when combined with a patient diagnosis, to give a predicted mortality. Unit predicted mortality can be calculated and compared with actual mortality to give a standardised mortality ratio (SMR) [11]. These and other data are published in the SICSAG annual Audit of Critical Care in Scotland. The first audit was published in 1998.

1.5. A review of scoring systems applicable to Intensive Care

1.5.1. Early scoring systems used in Intensive Care

Scoring systems are not a new concept. Virginia Apgar published a score in 1951 to examine the state of the newborn [17]. It comprises 5 variables and is performed at 1 and 5 minutes after birth. Its ease and simplicity mean it is still routinely done 60 years after it was conceived. In 1976 Ranson described a score for predicting the severity of acute

pancreatitis [18]. The most famous of the early scores from the early 1970s still in use worldwide is the Glasgow Coma Score of 1974 [19]. It documents the conscious level of a patient based upon the best motor, verbal and eye response. Originally out of 14 it was subsequently revised to give a score between 3 and 15 points.

Early scores therefore tended to focus and prognosticate in patients with a single diagnosis. The 1980s saw a rapid increase in developments in Intensive Care. New technologies and therapies were expensive, even more so as described earlier. This led to the development of scores of a more global nature which could help prognosticate as to who might benefit from admission to ICU.

1.5.2. Types of scoring systems in Intensive Care

For the initial search of the scoring systems a sample of the leading textbooks in critical care was reviewed. All the original descriptions were obtained and papers in which the scoring systems were described. An *Ovid Medline* search was also performed to identify any other descriptions of scoring systems that had been potentially missed, but none were identified.

There is no agreed classification of types of score[20] but it is useful to group them as physiological, intervention based / therapeutic weighted, assessment of organ failures, or disease specific / miscellaneous scores. Mortality rates in intensive care are far higher than on general hospital wards. This is why, quite reasonably, many scoring systems use mortality at various stages in the patient stay as the primary outcome measure.

1.5.3. Physiological Scores

1.5.3.1. Acute Physiology and Chronic Health Evaluation (APACHE) score

Of the physiological scoring systems, one of the best known is APACHE II (Acute Physiology and Chronic Health Evaluation II) which was developed in 1985 by Knaus et al. [21]. As the APACHE II score is the score most commonly used in Scotland, I will include more detail about its development and use than on other scores. The APACHE II score superseded the original prototype APACHE [22]. The original system was developed on the premise that the severity of acute disease could be measured by quantifying the degree of abnormality of different physiological variables. The original score contained 34 such variables, the degree of derangement of which scoring 1 to 4 points, summed to produce an acute physiology score (APS). The greatest derangement of each variable within the first 24 hours after admission to ICU was used to score points. This time frame

was chosen to ensure the greatest chance of all relevant parameters being recorded and available for scoring. The 34 variables were selected after a literature review and weighted by a panel of clinicians based on their clinical experience. It was recognised that chronic disease of differing severity decreased the likelihood of surviving Intensive Care. The initial APACHE score therefore incorporated a 4 letter code (A-D) representing the severity of chronic disease. This was done by means of a health questionnaire.

The original APACHE score was found to have a direct correlation with hospital mortality [22]. It was also useful for comparing the success of different treatment programmes and for evaluating the outcome of Intensive Care [22]. However, it soon became apparent that it contained too many variables, was unnecessarily complex and lacked robust validation.

1.5.3.2. Acute Physiology and Chronic Health Evaluation II (APACHE II)

A desire to simplify and validate the original APACHE score was therefore the driving force behind the development of APACHE II [23]. The number of physiological variables comprising the acute physiology score was reduced from 34 to 12. Firstly, ones which were not often recorded, e.g. serum osmolarity, were deleted. A set of essential clinical variables was then established. By a process of multivariate analysis of the original APACHE system, variables which added little to survival prediction were also deleted. These included urine output, albumin and glucose. The authors postulated that they added little to the core parameters as they were more heavily influenced by intervention than actual disease severity. The authors appreciated that a patient's physiological reserve decreases with increasing age and points were now awarded for 5 different age ranges. The chronic health questionnaire was replaced by chronic health points for severe organ insufficiency (liver, cardiovascular, respiratory and renal) or immuno-compromise with a weighting for nonoperative or emergency postoperative procedures versus elective postoperative procedures. The significance of emergency surgery as a predictor of worse outcome was also now appreciated.

The process of simplifying the number of variables and changing their relative weightings used the methodology designed by Gustafon et al. [24] who described a strategy of developing a replicable index. Up until this point there had been no clear descriptions in the literature of methodology for developing severity indices, nor of different panels of clinicians producing severity indices with similar performance characteristics. By illustrating the development of a heart disease severity index the authors outlined their

method and a test of the method's transportability. To that end, some of the weightings assigned to the original APACHE variables were changed, as it was appreciated that a low Glasgow Coma Scale and acute renal failure were very poor prognostic signs.

To calculate the APACHE II score, the acute physiology points are added to those for increasing age and chronic ill-health to give a maximum score of 71. The variables are illustrated in table 1-1 (adapted from Knaus W.A. et al. APACHE - acute physiology and chronic health evaluation: A physiologically based classification system[22].)

Table 1-1: Variables used in the calculation of the APACHE score

Score Component	Variables with maximum score
Acute Physiology Score (APS) Maximum = 60	Temperature (^{0}C) - (4) Mean Arterial Pressure (mmHg) - (4) Heart Rate (Ventricular Response) - (4) Respiratory Rate (non-ventilated or ventilated) - (4) Oxygenation, A-aDO if FiO2 >0.5, PaO2 (mmHg) if FiO2 <0.5 - (4) Arterial pH - (4) Serum Sodium (mMol/L) - (4) Serum Potassium (mMol/L) - (4) Serum Creatinine (mg/100ml), points doubled if acute renal failure - (4 / 8) Haematocrit (%) - (4) White blood count (total/mm^3 in 1000s) - (4) Glasgow Coma Score (GCS). Score = 15 minus actual GCS - (12)
Age Points Maximum = 6	Age (yrs) < 44 - (0) 45-54 - (2) 55-64 - (3) 65-74 - (5) ≥ 75 - (6)
Chronic Health Points Maximum = 5	History of severe organ system insufficiency or is immuno-compromised plus a. Non operative or emergency postoperative patient - (5) b. Elective post operative patient - (2)

The score was validated by examining its association with hospital mortality in unselected ICU admissions from 13 hospitals in the United States between 1979 and 1982. 5815 ICU admissions were included. It was shown that there was a relationship between an increasing APACHE II score calculated within the first 24 hours and hospital mortality. However, the score alone could not predict a specific risk of death unless a diagnostic category was included. Patients on admission to ICU in the above analysis were assigned to a specific diagnostic category according to their principal reason for admission. The

overall risk of death varied with the diagnostic category assigned and whether the patient had received emergency surgery. By a process of multiple logistic regression, diagnostic category weightings were derived. In other words, the APACHE II score could be combined with specific diagnostic categories to give predicted hospital mortality. The individual risk (R) of hospital death is given by the equation (R/1-R) = -3.517 + (APACHE II score x 0.146) + (0.603 if post emergency surgery) + (Diagnostic category weight). A list of principal diagnostic categories leading to ICU admission is given in the appendix of the original paper [21].

1.5.3.3. Acute Physiology and Chronic Health Evaluation III (APACHE III) Score

The APACHE III prognostic system was published in 1991 by the same authors [25]. The aim was of improving the risk prediction available with APACHE II, to make a distinction between predictive estimates of mortality for groups of patients versus individual mortality estimates, predict unit length of stay and to examine the relationship between timing of ICU admission and outcome.

The APACHE III score comprises (as in APACHE II) physiological variables (up to 252 points), age (up to 24 points) and chronic ill-health (up to 23 points) which are summed to give a maximum score of 299. This score can be combined with a single disease category to perform a relative risk stratification. Interestingly, after multivariable logistic regression analysis, there are up to 17 physiological variables in this score as opposed to the 12 in APACHE II. The added variables were blood urea nitrogen, urine output, serum albumin, bilirubin and glucose i.e. some variables previously considered not to add to the predictive power of the score. It was also appreciated, on analysis of their large data set of 17440 ICU admissions, that the predictive power of extremes of physiology had been underestimated, in particular hypotension, and that a narrower range of physiological variables should be assigned a zero weighting. Seven chronic health comorbidities were found to be statistically useful, but not in elective postoperative patients, and are thus excluded from the calculation of the score for these patients.

The result was a system with two major components, an APACHE III score and an APACHE III predictive equation. The APACHE III score can be used to provide initial risk stratification for severely ill patients within defined patient groups. The APACHE III predictive equation uses the APACHE III score and reference data on major disease

categories, plus the patient's location before Intensive Care admission, to provide an individual risk estimate of hospital mortality for different Intensive Care patients.

It is worth noting that, in an analysis by Woods et al [26] of 22 Scottish ICUs over a two year period, unit length of stay, predicted by APACHE III, did not correlate well with the actual length of stay. The length of stay in Scottish ICUs was consistently less than that predicted by a system based on American practice. Further, in a comparison of 5 intensive care scoring models using data from 22 general ICUs in Scotland, Livingston et al. concluded that the APACHE II score's calibration made it the most suitable for comparison of mortality rates [27]. It is worth noting, however, that new coefficients for the APACHE II score have been created for analysis of Scottish data, rather than those used in Knaus's original paper. This is because standardised mortality ratios (SMRs) in Scottish Units have been falling, coupled with varying mortality prediction accuracy [28]. The new coefficients have been used since November 2012, although the latest report from SICSAG shows SMRs based upon both the original and the recalibrated coefficients.

1.5.3.4. Simplified Acute Physiology Score (SAPS)

Following APACHE other physiological based scores have emerged. Le Gall et al in 1984 described the Simplified Acute Physiology Score (SAPS) [29]. This utilises 14 clinical and biological variables in a simple score to classify patients into groups of increasing risk of death. It was evaluated in 679 consecutive patients admitted to 8 Intensive Care Units in France. The classification into groups of increasing risk of death was shown to hold, irrespective of diagnosis. The authors argued that it was less time consuming to calculate, given its simplicity, yet compared well to the more complex Acute Physiology Score.

1.5.3.5. Simplified Acute Physiology Score II (SAPS II)

In the early 1990s the number of scoring systems available to clinicians was growing. Increasing complexity meant a longer time and greater expenditure were required to collect data to input into the scores. There is merit in simplicity, and in 1993 Le Gall et al published their new Simplified Acute Physiology Score (SAPS II) [30]. This European and North American study analysed data from 13152 patients. It comprises 17 variables, 12 of which are physiological. These are all readily available. The aim had been to have a purely physiologically based score leading to a prediction of hospital mortality which was independent of diagnosis. However, the authors found that the model performed more favourably if combined with 3 underlying disease variables (metastatic cancer,

haematological malignancy, AIDS), age, and type of admission. It is still true to say, however, that the SAPS II score is independent of the primary diagnosis.

1.5.3.6. Mortality Probability Model (MPM)

The Mortality Probability Model (MPM) was described by Lemeshow et al in 1985 [31]. Unlike SAPS or APACHE this was the first attempt at a score which was purely statistically derived. That is to say, the relative weights of variables were not subjectively determined. An analysis was undertaken of 755 general medical and surgical patients. 137 variables were collected at admission, and 75 at 24 hours after admission. Using statistical techniques the relative importance of each variable was determined and only those with a strong association with outcome retained. This resulted in 7 variables collected at admission and 7 at 24 hours. Again, unlike APACHE and SAPS, this model could be applied at the time of admission.

Although I have classed this as a physiologically based score there is a greater emphasis on condition-based variables than in APACHE e.g. the presence or absence of a cardiac arrhythmia. Further the physiological variables are recorded as affirmative or negative rather than as an actual number.

Lemeshow published an updated form of the model, the MPM II in 1993 [32]. Using two much larger data sets, 19124 patients in total from multiple ICUs were analysed. This resulted in two models, MPM_0 at admission MPM_{24} at 24 hours. Again, as in the MPM, the variables are recorded as simple yes or no answers. For example, in relation to blood pressure the model states "*record whether the systolic blood pressure was noted to be less than or equal to 90mmHg within 1 hour before or after ICU admission*". MPM_0 requires the collection of 15 and MPM_{24} a further 8 variables. Both models were shown to be good systems for reliably estimating hospital mortality. At that time MPM_0 was, by definition, the only model for estimating hospital mortality which was independent of treatment.

1.5.3.7. Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity (POSSUM)

Copeland et al published the Physiological and Operative Score for the enumeration of Mortality and Morbidity (POSSUM) in 1991[33]. It is different from the previous scores described, as it was designed to enable comparisons between general surgical patients having a wide variety of operations, by risk adjusting them based upon their physiological condition. In this score, 12 physiological parameters, shown by multivariate analysis (out

of an original 62) to independently predict outcome, are analysed, and graded into categories scoring 1, 2, 4 or 8 points depending upon their degree of derangement. These are combined to give a POSSUM physiology score. Six operative parameters are similarly graded. These include blood loss, the presence of peritoneal contamination and malignancy status. The physiology and operative severity scores are combined in the following formula to predict mortality risk (given by the letter R):

Ln R/1-R = -7.04 + (0.13 x physiological score) + (0.16 x operative severity score).

Prytherch et al in 1998 claimed that the POSSUM score over-predicts mortality, especially for patients with a low risk (5% or less) [34]. They modified the original POSSUM logistic regression equation by analysing 10,000 general surgical cases between 1993 and 1995. 2500 cases from the 10,000 were used to modify the equation and the new equation was tested on the remaining 7500. This formed the Portsmouth- POSSUM or P-POSSUM score. They showed that the new equation fits the observed in-hospital mortality better. The revised equation is:

Ln R/1-R = -9.065 + (0.1692 x physiological score) + (0.1550 x operative severity score).

1.5.4. Intervention based / Therapeutic weighted scores

Intervention based (or therapeutic weighted) scores assume that critically ill patients require more medical and nursing intervention than those who are less unstable. That is to say, the amount of intervention is a surrogate for the severity of illness. The most widely known of this type of score is the Therapeutic Intervention Scoring System (TISS), devised by Cullen et al in 1974 [35]. Up to 76 different therapeutic activities score a certain number of points, e.g. central line insertion, the need for renal replacement therapy, number of vasoactive drugs etc. It therefore allows an assessment of cost as well as severity, but there were drawbacks to this type of score. It was time consuming and the 76 therapeutic items did not always reflect the amount of care the nursing staff would need to give an individual patient. Although it correlated reasonably well with severity of illness, its use for that purpose decreased with the advent of more specific systems such as APACHE[22]. It remained a useful tool for quantifying nurse workload and resource utilisation. Miranda et al revised the original TISS-76 in 1996. Using 10,000 random records of TISS-76 items from 903 consecutive ICU admissions, and through a process of multivariable regression analysis, they reduced the number of variables to 28. This is the simplified TISS or TISS-

28 [36]. This system is widely used, and a study by Lefering et al in 2000 showed that, in an analysis of 1986 patients equating to 10,448 observation days, the TISS-28 adequately reflects the amount of critical care provided, and in the context of a surgical ICU, may also provide useful information about prognosis[37].

1.5.5. Assessment of organ failure scores

1.5.5.1. Sequential Organ Failure Assessment (SOFA)

By its nature, organ failure is common in the critically ill. It is not surprising that a group of scores have been developed specifically looking at organ failure and outcome. Vincent et al. published the Sequential Organ Failure Assessment (SOFA) score in 1996[38]. It was created during a consensus meeting of the European Society of Intensive Care Medicine in 1994, and revised in 1996, with the aim of producing a simple and continuous (sequential) score that could be widely used. It looks at 6 organ systems (respiratory, cardiovascular, renal, hepatic, central nervous system, and coagulation). As would be expected, as the number of organs which have failed increases, so does the mortality. Both the highest and mean SOFA scores are good predictors of outcome. Unlike APACHE II it is calculated on the day of admission and subsequent days in ICU. Other advantages over APACHE II are that it takes into account the level of cardiovascular support (dobutamine, noradrenaline or adrenaline) that the patient is receiving.

1.5.5.2. Multiple Organ Dysfunction Score (MODS)

The Multiple Organ Dysfunction Score (MODS) [39] was published in 1995 during the refinement period of the SOFA score. It looks at the same 6 parameters as SOFA, the major difference being the method of cardiovascular assessment. MODS utilises the pressure-adjusted heart rate, defined as the heart rate multiplied by the ratio of right atrial pressure to mean arterial pressure. Given the similarity of the data collected, it is unsurprising that both give similar mortality predictions. However, in a study of 949 patients by Bota et al. comparing the outcome prediction of the two scores, cardiovascular dysfunction was better related to outcome using the SOFA, rather than the MODS model [40].

1.5.6. Miscellaneous scores

For completeness, it is worth mentioning briefly several scores which do not fit neatly to the above classification, but nonetheless are of relevance to critical care.

Examples include the TRISS (Trauma Injury Severity Score) for severe trauma [41]. This score combines a weighted Revised Trauma Score, Injury Severity Score, and a score for the patient's age, and it also takes into account the mechanism of injury (blunt or penetrating trauma). The CURB-65 score is validated for predicting mortality in community acquired pneumonia [42]. It makes an assessment of confusion, urea, respiratory rate, and blood pressure. The rate of death at 30 days increases with an increasing score. The rule of nines for the assessment of burn area is now several decades old but still in widespread use [43]. The original Baux score [44] from 1961, (age plus percentage of burn), for predicting mortality after a burn, was not modified until 2010 by Osler et al. The new revised Baux score now includes smoke inhalation as a contributing factor for mortality prediction [45].

1.6. The need for a novel score

As can be seen, numerous elaborate and sophisticated scoring systems exist. However, they all have drawbacks and limitations of varying degree. In general, the statistical analysis underpinning the variables selected, and the relative weights given, often come from studying large American databases of patients treated in the early 1980s. These may not be entirely applicable to a European population. Several scores e.g. APACHE II, give only a snapshot of what is occurring during the first 24 hours after admission, and take no account of how much the patient is being supported when interpreting the physiological data collected, e.g. scoring a "normal" blood pressure when the patient is receiving high doses of vasopressor or inotrope. All the scores described fall short of what would be regarded as the "ideal score".

1.6.1. The properties of the ideal scoring system

In 1998 Saxon Ridley described an ideal scoring system [46] as one that is:

- *Validated* (where its ability to predict mortality is tested on a different population from that used to create the score).
- *Calibrated* (how closely the score's mortality estimation correlates with actual mortality over the range of probabilities).
- *Accurate* (its ability to discriminate between patients who will live and die). Interestingly, physicians are in general better at discriminating who will survive or not survive at the very ill / very well end of the spectrum. Scoring systems tend to perform better in the mid-range of mortality risk [46].

- *Reliable* (which relates to intra- and inter-observer agreement for e.g. in data collection. If there is wide variation in the choice of primary diagnosis, the system may be unreliable).
- Has *content validity* (a measure of the comprehensiveness of the score).
- Has *methodological rigour* (bias is avoided by basing the score on a large data base of consecutive patients).

Other important properties of an ideal score would be that it:

- Is based upon routinely collected data.
- Takes into account treatment effect.
- Can be calculated and displayed in real time.
- Is repeatable an infinite number of times.
- Is automated.
- Is widely applicable.
- Does not exclude patients groups e.g. patients less than 18 years, burns, pregnancy, readmissions to ICU.
- Is possibly diagnosis independent.
- Can be used as a research tool.
- Predicts outcome.

1.7. Applicability of previous scores to the work in this thesis

In this thesis I describe the development of a novel quantitative and then qualitative score of cardiovascular instability in Intensive Care Patients. As well as setting the work of this thesis in context, the initial reason for reviewing the scoring systems was to establish which criteria were used and the weights attributed to the derivation of each variable in measuring the severity of critical illness. Some of the ranges used could then be incorporated into the novel quantitative score. The first scoring systems in general were used to predict or explain mortality, rather than cardiovascular stability. Although they are discussed in the introduction of the thesis, to start with, outcomes were a secondary consideration as the initial aim of the research was to devise a new score to capture instability. Of particular interest was the physiological component of each score and whether it was validated for calculation once e.g. APACHE II, more frequently e.g. SOFA or if any scores summarising the physiological state of a patient came close to what this research was attempting to create i.e. a score that could be repeatedly calculated. In table 1-2 the main scores reviewed are summarised along with their strengths, weaknesses and applicability to the novel quantitative score described in chapter 3 of this thesis.

Table 1-2: Applicability of previous scores to the work in the thesis

Score	Outcome	Strengths / weaknesses	Helpful for my new score?
APACHE II	Mortality	Calculated once in 24 hours Burns excluded Readmissions excluded Age < 16 excluded	Some ranges from the acute physiology score incorporated
SAPS II	Mortality	Easier to calculate than APACHE II (less variables) Independent of primary diagnosis	Ranges for heart rate and temperature useful as a guide.
MPM	Mortality	Could be calculated at 0 and 24 hours	Interesting as statistically derived, but mainly condition-based variables so of less use for a new score.
POSSUM	Mortality	Allowed comparison between general surgical patients having a wide variety of operations.	Useful guide to ranges for heart rate
SOFA	Morbidity (subsequently association with mortality)	Characterises organ failure as a continuum rather than present or absent	Useful guide to ranges for adrenaline / noradrenaline
TISS / TRISS / CURB-65/ Rule of 9s	Various	-	Included for completeness

1.8. Introduction to machine learning in a healthcare setting

The work in this thesis involved collaboration with computing science colleagues from Aberdeen University. During this time, large quantities of data were extracted from the Electronic Patient Records of Intensive Care patients at Glasgow Royal Infirmary, processed and analysed. This posed many challenges. However, with greater experience, the handling and processing of the data became more sophisticated. An introduction to this field of computing science is described below.

Making computers more intelligent is the branch of computing science known as Artificial Intelligence [47]. To become intelligent you have to learn and machine learning is the science of computational methods for inducing knowledge through a process of "*accumulating, altering and updating knowledge within intelligent systems"* [48]. There has been an interest in modelling algorithms to analyse large datasets since the first primitive computers were introduced in the 1950s. Machine learning can be regarded as a development of Artificial Intelligence with the aim of pattern recognition within data and with the learned patterns performs meaningful inferences. Machine learning applications are therefore useful for pattern recognition, classification problems and in prediction [49]. There are two main types of inference. "Classification" is where unseen data is analysed and a determination made as to which of a number of classes it belongs to. "Regression" is where unseen data are used in the prediction of behaviour in one or a series of random variables [50]. The determination of physiological stability or instability is a form of "classification".

There are several major obstacles to be overcome for classification or regression inferences to be meaningful. Large amounts of data are often involved in these processes. The quality of the data is important e.g. inconsistency in the units used for recording a parameter, inconsistently recorded parameters and missing data. This can lead to some models being developed or conclusions being drawn which are not credible (or that are blindingly obvious to a clinician). A lot of sophisticated software for processing is required but it only works if it actually reflects the sophisticated manner in which data is actually interpreted by clinicians in real time. However, there is also a real need to actually avoid opinion (or biases) in the traditional sense of the word i.e. what is really needed is the ability to capture expertise as it is reliably applied in the clinical situation.

1.9. Methodology for assessing validity

The thesis describes the development of a quantitative and then qualitative score. There are many challenges in the validation of a novel score, particularly in the absence of a previous

gold standard. An overview of the major types of validity and how they may apply to developing a scoring system is therefore provided here.

Validity derives from the Latin word *validatis* meaning strong. Of all types of validity, face validation (sometimes known as surface validity) is perhaps the easiest (but weakest) form to undertake. It examines whether a test covers subjectively what it is supposed to be measuring. For example if you set an examination for students in a subject and show the examination paper to a group of colleagues then they may agree that "on the face of it" the examination is a fair reflection of the subject matter to be covered. Face validation is often considered a minimum validation requirement [51]. A novel rule based physiological scoring system designed to capture improvement or deterioration in a patient could be shown to expert colleagues and they may agree that on the surface that it adequately captures improvements or deteriorations. However, this approach has disadvantages. As a subjective method it is inherently weak. Just because other experts agree that on the surface a scoring system captures changes in the state of a patient, it does not mean that any of them are correct. Further, they could be influenced by the manner in which questions about a new score are put to them. Finally colleagues may not wish to contradict other colleagues' work.

Another non-statistical validity related to face validity is content validity. Here the issue is adequacy of sampling and it is therefore a measure of how much an "empirical measurement reflects a specific domain of content" [52] i.e. does the test represent all aspects of the construct being studied? Bachmann in 1990 summarised the difference between face and content delivery with "face validity is the appearance of real life (and) content relevance the representation of real life" [53]. Content validity is used predominantly in devising educational tests and in psychology. It has to rely on experts who are familiar with the area in which the test is measuring. Typically to confirm content validity, experts would be shown the measurement tool and asked to provide feedback as to how well it measures the construct being considered [52]. For a physiological score of instability experts would have to provide feedback that the parameters used in the score to capture instability was a fair representation of parameters that could cause instability. The major problem with content validity is that it still relies upon consensual professional judgement as to whether the test content adequately covers the domain in question [54]. Further there are no agreed upon criterion for determining content validity.

Criterion Validity assesses the degree of correlation of a test with a gold standard [51](the best available test [55]) and can be divided into concurrent and predictive validity. For a test to have concurrent validity it has to be simultaneously applied to a previously validated gold standard test for the phenomenon under investigation and the results compared. For the test to have predictive validity it has to predict an outcome from the phenomenon under consideration compared with a previously validated gold standard test applied in the same manner. Traditionally the degree of concurrent validity has been assessed by linear regression and correlation statistics. However Bland and Altman whilst conceding that the correct statistical test is not obvious, state that correlation measures "*strength of a relation between two variables, not the agreement between them*". [56] They are of the opinion that the use of precision and bias statistics is more appropriate. However it does appear that linear regression and correlation coefficients are still commonly used tools. For example, in a recent study quantifying the concurrent validity of hamstring length measures the authors used a combination of linear regression, correlation and kappa statistics [57]. In a test of a new scoring system of instability there is unfortunately no gold standard so criterion validity cannot be used.

Construct validity of a test is established by demonstrating that the test or measurement tool "*measures the variables or constructs that it proposes to identify or measure*". [58] In other words does the test measure what it claims to be measuring? This form of validity is useful when no universally criterion exists. Researchers have to identify other measures that would theoretically support the concept (or construct) being measured [59].Two subtypes of construct validity are convergent and discriminant validity. Convergent validity refers to the degree to which two measures that should be related are, whereas discriminant validity is a test of whether measures that theoretically should be unrelated are indeed unrelated. With a new scoring system of instability if two lines of physiological data are unrelated (because the scoring system judges them to be in different categories of stability) and a clinician agrees that there has been either improvement or deterioration then this could be argued to be a form of convergent validity. Similarly if two lines of physiological data are judged to be the same (because they place the patient at the same level of stability) and a clinician agrees that there has been no change then this is a form of convergent validity.

The narrative above highlights that for a novel score of stability there is no one form of validation test that can be applied. To successfully validate a score in the absence of a previously validated gold standard would require a number of tests. Table 1-3 summarises the main types of validity and their applicability to a novel instability score.

Table 1-3: Types of validity and their applicability to a novel instability score

Type of validity	**Pragmatic definition in relation to a novel instability score**	**Can it be used to validate a new score of instability?**	**Comments**
Face validity	Score appears on the face of it to be a reasonable score of instability	Yes	Simple to do and useful. A weak form of validity as relies on expert opinion. They may all be incorrect.
Content validity	The score takes into account what most would regards are the key parameters of instability	Yes	Important to do but still relies upon professional consensual judgement.
Criterion validity	Score is correlated to a previously validated gold standard of instability	No	Cannot be done as there is no gold standard.
Convergent validity	When there is no change in level of instability in the score a clinician agrees	Yes	Useful in the absence of a gold standard.
Discriminant validity	When there is a change in the level of instability as judged by the score the clinician agrees	Yes	Useful in the absence of a gold standard.

For completeness it is worth discussing *reliability*. Where validity is the extent to which a test or measurement actually measures what it is supposed to measure, reliability is the extent to which a test or measurement gives consistent results [60]. An analogy is darts thrown at a dartboard. If all the darts hit the bull's eye then there is high reliability and validity. If all the darts hit the bottom of the board there is high reliability but low validity. If there is a scatter of darts then there is low validity and low reliability. Theoretically if the relevant variables for the score have been collected, they have been collected correctly, the equipment recording the variables is in good working order and the data is processed properly then the score should be reliable (even if not fully validated).

1.10. Cardiovascular stability during renal replacement therapy

1.10.1. Incidence of Acute Kidney Injury

The incidence of acute kidney injury in Intensive Care in some studies approaches 70% [61]. This figure is based on studies using the Risk, Injury, Failure, Loss and end-stage renal disease (RIFLE) criteria for acute kidney injury, defined by the Acute Dialysis Quality Initiative in 2004 [62], and refined by the Acute Kidney Injury Network in 2007 [63]. Overall, 12% of patients in Scottish Intensive Care, or combined Intensive Care and High Dependency Units, require some form of renal replacement therapy. The figure for Intensive Care units which do not have a high dependency component is slightly higher [64].

1.10.2. Modalities of renal replacement therapy

There are two principal modes of renal replacement therapy which are used in Intensive Care Units, viz. continuous haemofiltration, and intermittent haemodialysis. There are also various hybrid techniques.

The process of haemofiltration was first introduced by Henderson in 1967 [65]. Continuous haemofiltration was described by Kramer in 1977 as a technique for the management of fluid overload, in patients who were unresponsive to diuretics in the Intensive Care Unit [66]. During this process a positive hydrostatic pressure forces water and solutes across a semi-permeable membrane. The solute is cleared by convection. A recent study revealed that continuous veno-venous haemofiltration remains the first line modality in 65% of Intensive Care Units in the UK [67]. By contrast, the less frequently used haemodialysis has earlier roots. In 1948 Bywaters described the first haemodialysis in the UK for patients with acute kidney injury [68]. Using counter current flow, solute diffuses down a concentration gradient through a semi-permeable membrane.

1.10.3. Comparison of haemodialysis and haemofiltration

Although there is no conclusive evidence that diffusive therapy (intermittent haemodialysis) is superior to convective therapy (continuous haemofiltration) in terms of outcome, haemodialysis does have some practical advantages. A study by Srisawat et al. showed that intermittent techniques were on average cheaper than continuous techniques [69]. They took into consideration nursing costs, dialysate and fluid replacement costs, anticoagulant and extra-corporeal circuit costs. Other advantages include less time in which a patient's blood is passing through a hazardous extra-corporeal circuit, less nursing input required and better and more rapid solute clearance.

1.10.4. Haemodynamic instability during intermittent haemodialysis

Despite the apparent advantages of intermittent haemodialysis, continuous convective techniques remain the norm, perhaps because of the widely held notion that they offer greater haemodynamic stability, and that any technique which causes cardiovascular instability will lead to a worsening of other organ failures.

The evidence to support this view-point is weak. In a widely quoted study from 1993, Davenport et al randomised 32 consecutive patients to receive intermittent machine haemofiltration, or a continuous technique, either arteriovenous haemofiltration or arteriovenous haemofiltration with dialysis [70]. Measurements included cardiac index, mean arterial pressure, pulmonary artery occlusion pressure, and tissue oxygen delivery. His group reported a fall in all of these parameters, leading to their conclusion that continuous forms of renal replacement therapy are preferred due to increased cardiovascular tolerability. It is worth noting that they used arteriovenous haemofiltration or haemodiafiltration, as opposed to the venovenous modes used in current practice. The patient population was not typical, as the centre was a quaternary centre for liver transplant, all the patients in the study also having hepatic failure.

A study by Vinsonneau, comparing continuous venovenous haemodiafiltration with intermittent haemodialysis in patients with multi-organ failure and acute renal failure published in 2006 [71], showed no significant difference in arterial hypotension between the two groups. This was a well constructed, prospective, multicentre trial with 360 patients from 21 Intensive Care Units in France. Moreover, the definition of hypotension was wide, with either a drop of systolic arterial pressure of >50mmHg from the baseline value, or a systolic arterial pressure of <80mmHg.

Most other studies in the literature in this area are of small scale, and are crossover in design. Further, they are of short duration, and have recruited small numbers of patients [72].

1.11. Myocardial infarction in Intensive Care

1.11.1. Diagnostic difficulty

The diagnosis of myocardial infarction in the Intensive Care Unit can be challenging. We are dealing, on the whole, with a population of patients who are sedated, intubated, and ventilated. Traditional symptoms such as chest pain, which may alert the clinician to the possibility of ischaemia, may not be apparent. Instead the clinician may have to rely on physiological disturbance, e.g. hypotension, hypoxia or cardiac arrhythmia, as an initial indicator of ischaemia [73]. However, both of these strategies have their limitations. With the advent of more specific and sensitive markers of myocardial injury, e.g. troponin (a regulatory protein of the thin actin filament), it is possible to detect myocardial injury more easily in the absence of overt ischaemia.[74]

In recent years there have been several consensus conferences to refine the diagnostic criteria for acute myocardial infarction. In 1999, the European Society of Cardiology and the American College of Cardiology recommended that cardiac troponins (I or T) are the preferred markers for the diagnosis of myocardial injury. They further added that detectable increases in biomarkers of cardiac injury were indicative of injury to the myocardium, but were not synonymous with an ischaemic mechanism, and, as such, could not mandate the diagnosis of myocardial infarction.[75] In 2007 the recommendations of a further global taskforce were published [76]. Myocardial infarction was classified into five clinical types. Type 2 is most relevant to intensive care, i.e. *myocardial infarction secondary to ischaemia due to either increased oxygen demand or decreased supply,* e.g. hypotension or arrhythmias. This would be more common than spontaneous myocardial infarction related to ischaemia due to a primary coronary event (type one). Specifically for Type 1 and 2 the term myocardial infarction should be used when there is evidence of myocardial necrosis in the context of a clinical setting consistent with myocardial ischaemia. In practice, this requires a rise in a cardiac biomarker (preferably troponin) with at least one value above the 99th percentile of the upper reference limit, plus one out of - symptoms of ischaemia, ECG changes indicative of new ischaemia, the development of pathological Q waves, or imaging evidence which can take the form of new loss of viable myocardium, or new regional wall motion abnormality.

In 2012, the Third Universal Definition of Myocardial Infarction was published [77]. The 5 types of myocardial infarction were refined with new imaging and ECG criteria added. The Taskforce behind the guidelines was careful to define these criteria in line with modern management of those suspected of having a myocardial infarction, with particular

emphasis on the clinical context, e.g. a myocardial supply / demand imbalance in the critically ill [78]. In contrast, the observation of ECG changes, or abnormality on echocardiography, characterising a territorial myocardial infarction, is a rare occurrence in our clinical practice [79].

1.11.2. Significance of myocardial injury in the critically ill

There is a growing body of evidence that raised cardiac biomarkers are independent risk factors for in-hospital, short, and long term mortality, even after adjustment for severity of disease. Babuin et al demonstrated a 30 day mortality of 35% in patients with a rise in the cardiac biomarker troponin T of ≥ 0.01 micrograms/litre, and a mortality of 14% without elevation [80]. In a study in 2008, Lim and colleagues systematically screened 103 consecutive patients on admission to intensive care with troponin measurements and ECGs [81]. These were repeated at serial intervals until death or discharge from Intensive Care, for a maximum of two months. The ECGs were screened for evidence of ischaemia as per the European Society of Cardiology and the American College of Cardiology guidelines for the purposes of diagnosing myocardial infarction. 35.9% had a myocardial infarction, 14.6% had an elevated troponin only, and 49.5% had no troponin rise. They showed that patients with an elevated troponin had a higher hospital mortality than those who had no rise. Of note was that screening detected a large proportion of myocardial infarctions not diagnosed clinically (62.2% ultimately diagnosed). Outcomes were similar in patients diagnosed with myocardial infarction clinically and in those patients where the infarction had been detected by screening alone.

It has been shown that myocardial infarctions, and troponin rises per se, are far more common than previously imagined in intensive care patients. This evidence comes from systematic screening studies. In one recent prospective study, the authors demonstrated a troponin rise in 47% of a critically ill cohort [82]. The incidence may be as high as 71 % depending upon the particular troponin used and the level selected to define myocardial damage [83]. There is little evidence based guidance on the management of intensive care patients who have an isolated troponin rise, or even a myocardial infarction. Early diagnosis, e.g. from subtle physiological disturbance, and early intervention are likely, but have not yet been proven, to be important.

1.12. Aims of this research

The initial premise of the quantitative (and later qualitative) scores was to capture cardiovascular instability in Intensive Care Patients in a more sophisticated manner than was possible before. The need for such scores arose from an initial interest in trying to establish whether haemodialysis was a cardiovascularly destabilising therapy. As the work developed, there was an interest to ascertain, if having devised scores of instability, whether they could have wider applicability. This is where the research was expanded into prediction of myocardial events and some very preliminary work examining mortality outcomes.

In this thesis I therefore describe:

- A simple experiment to quantify cardiovascular instability during haemodialysis, and the need for a sophisticated scoring system.

- The development of a quantitative score to capture cardiovascular instability in the critically ill.

- The development and initial validation of a qualitative score to capture cardiovascular instability in the critically ill.

- The wider applications of the qualitative score: Association and prediction of myocardial events and preliminary work on outcomes in the critically ill.

Chapter 2: Quantifying cardiovascular instability during intermittent haemodialysis and the need to design a sophisticated scoring system

2.1. Abstract

2.1.1. Background

Acute renal failure requiring renal replacement therapy is common in critically ill patients. Haemofiltration is often favoured over haemodialysis as the method of renal replacement therapy, due to a belief that haemofiltration is less cardiovascularly unstable. There is little evidence in the literature to support this view. In a small proof-of-concept study, cardiovascular stability was characterised in a cohort of patients requiring haemodialysis in Glasgow Royal Infirmary Intensive Care Unit.

2.1.2. Methods

Physiological data were collected from the Electronic Patient Records of 10 critically ill patients undergoing a total of 23 dialysis sessions. The data were anonymised, pre-processed and analysed. For the purposes of this experiment, cardiovascular instability was defined as a 20% change, in either direction, of heart rate or mean arterial pressure. In a second analysis, the dialysis sessions were examined to ascertain what percentage of mean arterial pressures stayed within an arbitrary “normal” range of 70-109, or moved around the range in different directions.

2.1.3. Results

Using the definition of a 20% change in heart rate or mean arterial pressure as representing cardiovascular instability, 65% of the sessions were stable and 35% unstable. Taking a normal mean arterial pressure as 70-109, 40% of mean arterial pressures changed from low to normal, 50% stayed within that range, and 10% changed from normal to high.

2.1.4. Conclusions

In this simple experiment there was a signal that haemodialysis was not a cardiovascularly unstable therapy. In fact, stability improved in a percentage of patients. The experiment did not take into account the amount of physiological or pharmacological support the patient was receiving to achieve the measured heart rate, or mean arterial pressure. This flaw is common to a lot of currently available scores. To overcome this problem, a new score would have to be developed.

2.2. Introduction

Acute renal failure requiring renal replacement therapy occurs in approximately 12% of patients admitted to Scottish Intensive Care Units [84]. Whereas renal units tend to use intermittent haemodialysis, critical care units more often use haemofiltration as the modality of renal replacement. This is due to a widely held belief that haemofiltration offers greater cardiovascular stability than haemodialysis, but with little evidence to support this view in the literature [71].

The definition of what constitutes hypotension is controversial. For example some clinicians define it as an arbitrary 20% drop in blood pressure [85]. In one of the most comprehensive multicentre prospective comparisons of haemofiltration and haemodialysis, Vinsonneau used a much broader definition of hypotension, namely a drop of systolic arterial pressure of >50mmHg from the baseline value, or a systolic arterial pressure of <80mmHg [71]. As can be seen, the definition of cardiovascular stability is extremely challenging. Even in national clinical guidelines (Scottish Intercollegiate Guidelines Network - Postoperative management in adults) this expert group could not define ranges for cardiovascular stability either for heart rate or for other cardiovascular variables [86].

The evidence to support haemodialysis being a cardiovascularly destabilising therapy is weak. In a widely quoted study from 1993, Davenport et al. randomised 32 consecutive patients to receive intermittent machine haemofiltration or a continuous technique, either arteriovenous haemofiltration or arteriovenous haemofiltration with dialysis [70]. Measurements included cardiac index, mean arterial pressure, pulmonary artery occlusion pressure and tissue oxygen delivery. His group reported a fall in all these parameters leading to their conclusion that continuous forms of renal replacement therapy are preferred due to increased cardiovascular tolerability. It is worth noting that they used arteriovenous haemofiltration or haemodiafiltration, as opposed to the venovenous modes used in current practice. The patient population was not typical as the centre was a quaternary centre for liver transplant, all the patients in the study also having hepatic failure. In the 2006 study by Vinsonneau mentioned above, comparing continuous venovenous haemodiafiltration with intermittent haemodialysis in patients with multi-organ failure and acute renal failure, there was no significant difference in arterial hypotension between two groups [71]. This was a well constructed prospective, multicentre trial with 360 patients from 21 Intensive Care Units in France. As above their definition of hypotension was wide. Most other studies in the literature in this area are of small scale

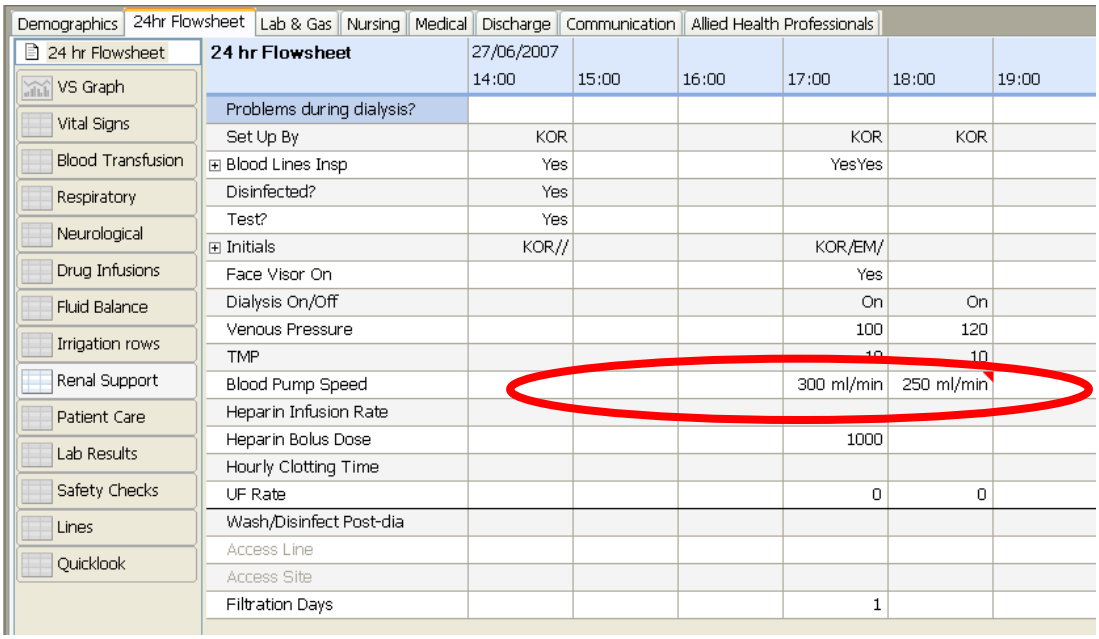
and are crossover in their design. Further they are of short duration and have recruited small numbers of patients [72].

With the advent of the means to analyse and process large quantities of physiological data, I hypothesise that, on closer scrutiny of key cardiovascular parameters, it is possible to refute haemodialysis causing cardiovascular instability.

2.3. Methods

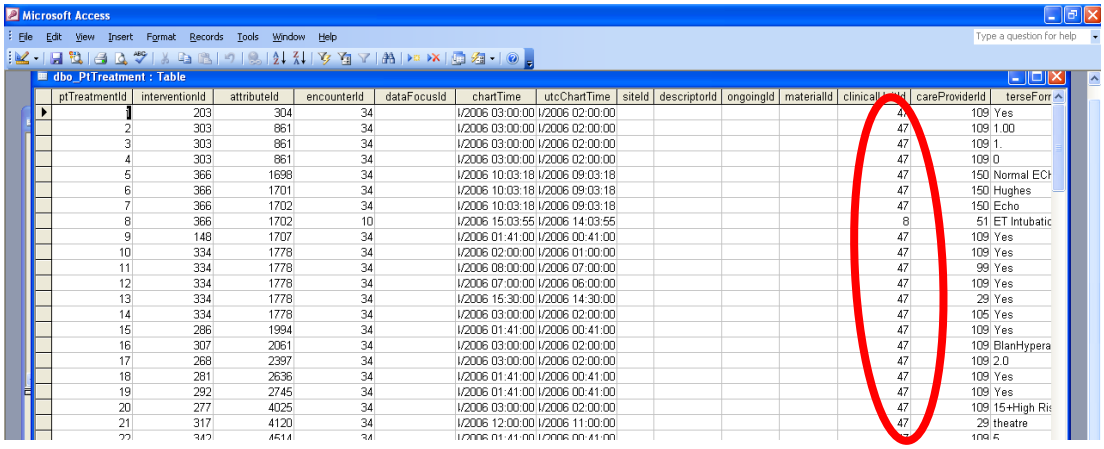
As a small proof-of-concept study, anonymised physiological data were collected from 10 patients in Glasgow Royal Infirmary Intensive Care Unit undergoing 23 dialysis sessions among them, in total. The data contained predominantly hourly time points. The Royal Infirmary ICU has dispensed with traditional paper based records and replaced them with an Electronic Patient Record, the Philips CareVue System [87]. This system allows the collection and storage of vast quantities of readily accessible patient data. These are entered into the electronic system by the nurse at a terminal beside the patient's bed and then verified, again by nursing staff. An example is shown in figure 2-1. Different physiological parameters or aspects of care can be accessed by clicking on the tabs in the far left column. In this example, the Renal Support tab has been clicked to reveal information about this patient's renal replacement therapy. The red circle indicates blood pump speed on dialysis, which is the method used by the CareVue administrators to extract data about patients receiving haemodialysis from the CareVue system (described below).

Figure 2-1: A screenshot showing the renal replacement therapy tab as displayed in the CareVue system



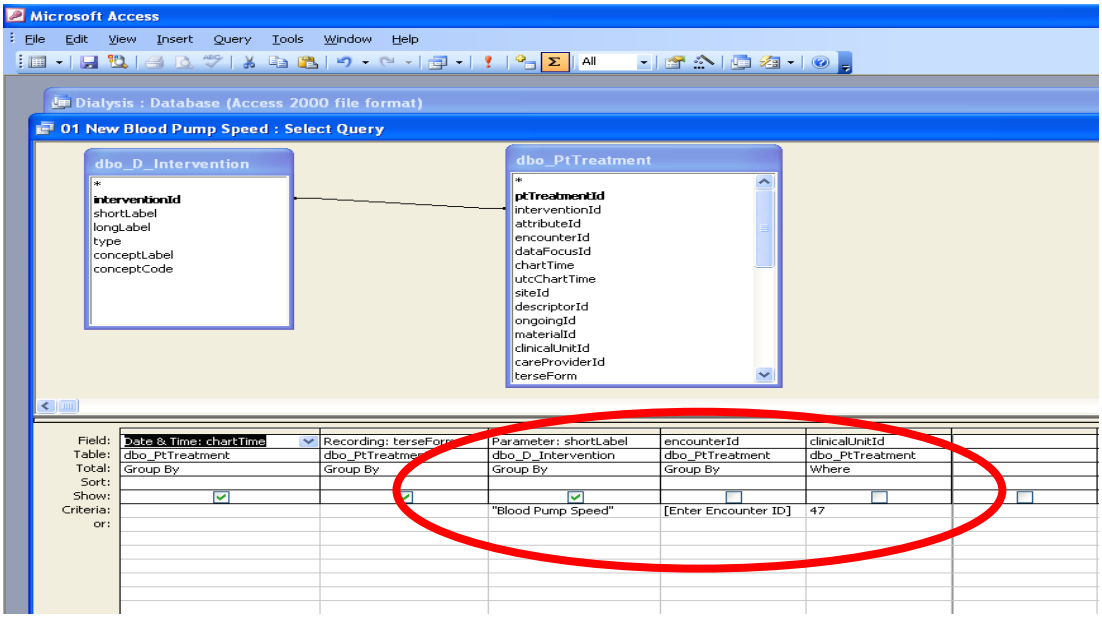
Data from the CareVue system is automatically manipulated by a programme within the system, and different aspects of the data are stored in relevant tables within the CareVue SQL database. SQL (Structured Query Language) is a programming language designed to manage data held within database management systems. These tables store data from all patients e.g. in Glasgow. Examples include an allergy table, an intervention table, and a treatment table. In other words, the allergy table within the SQL database holds information about many different patients' allergies. It is these tables which are interrogated by the CareVue administrators to extract data of interest. Figure 2-2 shows a screenshot from the treatment table within the SQL database. The circle shows some of the different treatments, e.g. cardiac ECHO and endotracheal intubation. Note that several different patients are contained within this particular table.

Figure 2-2: A screenshot from the treatment SQL database



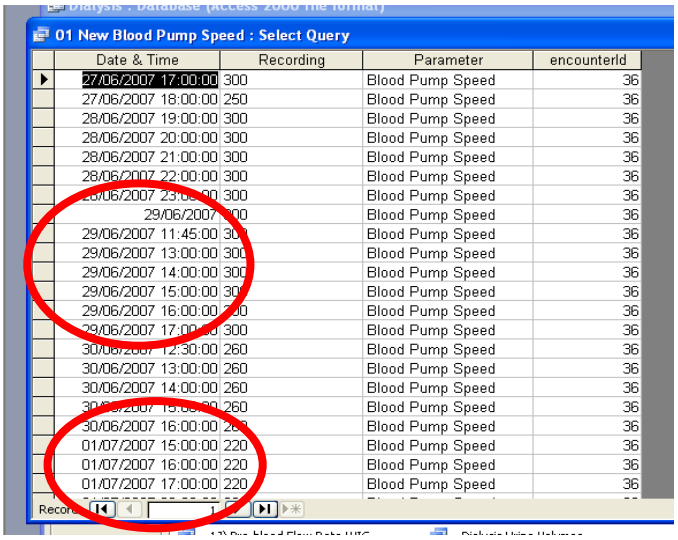
For this experiment, the CareVue administrators interrogated the CareVue SQL treatment table to look for instances where blood pump speed was recorded. This identifies when dialysis was occurring. Other tables were interrogated simultaneously to extract corresponding physiological parameters of interest, e.g. heart rate and mean arterial pressure. The queries were run using Microsoft Access software. Figure 2-3 shows an example of a query looking for blood pump speed (and by definition periods where haemodialysis is occurring). Note the number circled (47) in figure 2-2 is a unique identifying code for Glasgow Royal Infirmary, so that data for other patients held within the treatment database are not extracted.

Figure 2-3: A screenshot showing a query interrogating the treatment CareVue SQL database for blood pump speed



Once the query was run, the data were exported into a Microsoft Excel spreadsheet. Figure 2-4 gives an example of such a spreadsheet. As the query was for blood pump speed, time periods between dialysis sessions were not detected. These had to be detected manually by looking for breaks in the predominantly hourly time points in the far left hand column titled Date and Time. Examples of sessions are circled. The process is anonymised at this stage as data relating to a particular patient's admission now receive a unique encounter identification, shown in the far right column of figure 2-4. These can if necessary be de-anonymised by the CareVue administrators.

Figure 2-4: An example of the query data exported into a Microsoft Excel Spreadsheet

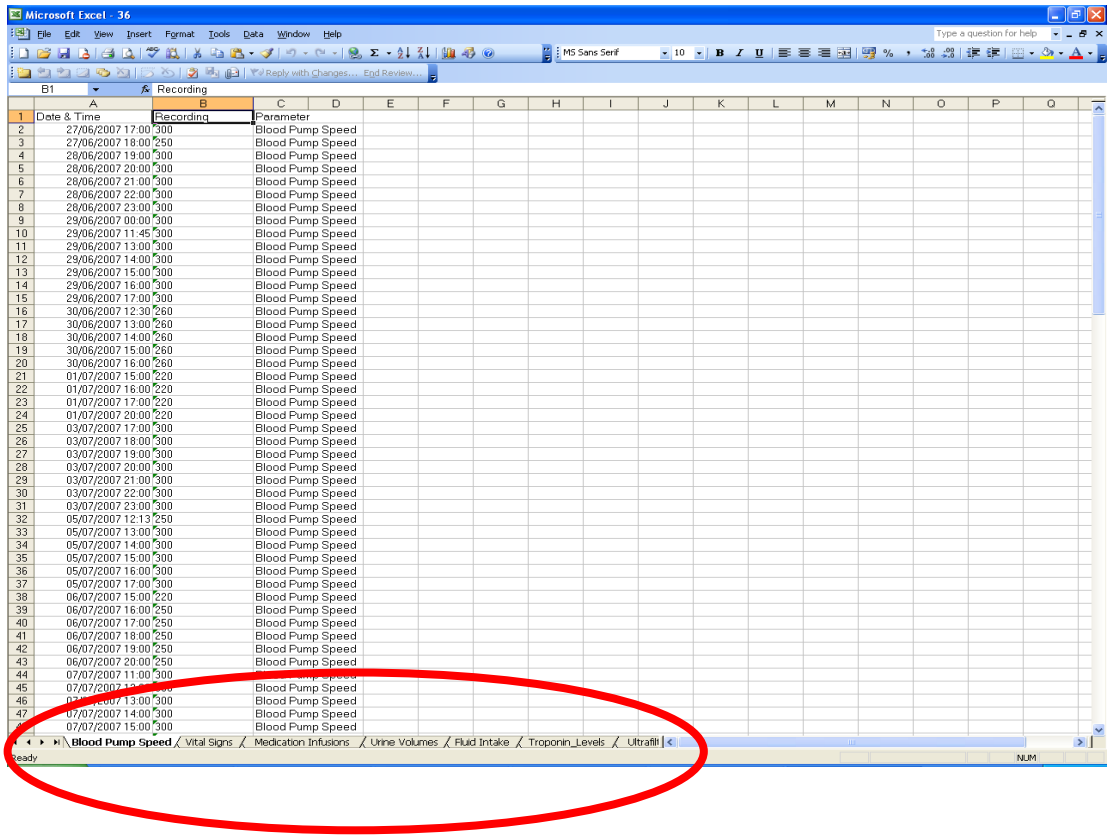


01 New Blood Pump Speed : Select Query

Date & Time	Recording	Parameter	encounterId
27/06/2007 17:00:00	300	Blood Pump Speed	36
27/06/2007 18:00:00	250	Blood Pump Speed	36
28/06/2007 19:00:00	300	Blood Pump Speed	36
28/06/2007 20:00:00	300	Blood Pump Speed	36
28/06/2007 21:00:00	300	Blood Pump Speed	36
28/06/2007 22:00:00	300	Blood Pump Speed	36
28/06/2007 23:00:00	300	Blood Pump Speed	36
29/06/2007	300	Blood Pump Speed	36
29/06/2007 11:45:00	300	Blood Pump Speed	36
29/06/2007 13:00:00	300	Blood Pump Speed	36
29/06/2007 14:00:00	300	Blood Pump Speed	36
29/06/2007 15:00:00	300	Blood Pump Speed	36
29/06/2007 16:00:00	300	Blood Pump Speed	36
29/06/2007 17:00:00	300	Blood Pump Speed	36
30/06/2007 12:30:00	260	Blood Pump Speed	36
30/06/2007 13:00:00	260	Blood Pump Speed	36
30/06/2007 14:00:00	260	Blood Pump Speed	36
30/06/2007 15:00:00	260	Blood Pump Speed	36
30/06/2007 16:00:00	260	Blood Pump Speed	36
01/07/2007 15:00:00	220	Blood Pump Speed	36
01/07/2007 16:00:00	220	Blood Pump Speed	36
01/07/2007 17:00:00	220	Blood Pump Speed	36

Finally, the relevant physiological data corresponding to the blood pump speed data were extracted by a number of queries, and combined to make a final Excel spreadsheet for analysis by our collaborating computing scientist colleagues. An example of the final spreadsheet is shown in figure 2-5. Different physiological parameters can be accessed by clicking on the tabs at the bottom of the screenshot (circled). A proportion of data (ranging between 10 and 20%) is always cross checked with the live patient records to ensure the reliability and validity of the data extracted.

Figure 2-5: Final Excel spreadsheet prepared by the CareVue administrators for analysis by computing science colleagues



The process of downloading the physiological and pharmacological information from CareVue by the administrators to get the "raw data" into basic Excel spreadsheets is described above A further example with a range of different parameters is shown in figure 2-6.

Figure 2-6: Final Excel spreadsheet with a range of parameters for analysis by computing science colleagues

08/12/2006 18:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 18:00	Noradrenaline 24mg/240ml(mg/hr),	4.8 mg/hr
08/12/2006 20:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 20:00	Noradrenaline 24mg/240ml(mg/hr),	4.8 mg/hr
08/12/2006 21:00	Noradrenaline 24mg/240ml(mg/hr),	3 mg/hr
08/12/2006 21:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 21:43	Alfentanil 25000mcg(mg/hr), 2500	1 mg/hr

As can be seen the data in its raw state comprises lines of information in an Excel spreadsheet which were not ordered in a manner by which they could be interpreted. It was appreciated that when larger data sets were starting to be analysed that there would need to

be a mechanism to view and annotate the data in an intelligible way and be able to apply rules to extrapolate where data were parameters were missing at different time points. Our computing science colleagues devised a web based tool ACHE (an Architecture for Clinical Hypothesis Examination) [88] to deal with these specific problems. It has two main components. ACHE "annotate" which allows the raw data to be viewed in an ordered fashion within the excel spread sheets and ACHE "pre-process" allows for any extrapolation rules to be applied to the data at this point. This was essential, as with the number of data points in some of the spreadsheets manual pre-processing of this time series data would have been impossible.

In this chapter where a small proof of concept study examining stability during dialysis is described, the only pre-processing which took place was a transformation of the raw data within the excel spread sheets. In this study the stability was assessed by examining two parameters, heart rate and mean arterial pressure. The average of these parameters was calculated before, during and after dialysis. A change of greater than 20% of baseline was the definition taken as representing "instability." Therefore as averages were being calculated, missing time points were less important as the average would be calculated for the available number of time points. It is unusual for nursing staff not to record these key parameters

Focusing on two key cardiovascular parameters, heart rate and blood pressure, I selected a change in these parameters of greater than 20% in either direction as representing cardiovascular instability. This was based on an accepted definition in the literature [85]. I rejected the wider definition of Vinsonneau [71], as I felt that a narrower 20% is still of clinical significance. The dialysis sessions were then analysed in this manner using the ACHE architecture. I thought it important to examine the cardiovascular state of the patient before and after the session, as this could have a bearing on the stability during the dialysis. Therefore, the blood pressure and heart rate were examined for 3 hours before the dialysis was commenced, and 4 hours afterwards.

Potential cardiovascular instability was also examined in a different manner. Selecting a "normal" mean arterial pressure in ICU of 70-109, the sessions were examined to ascertain what percentage of patients changed their mean arterial pressure within this range, from a low mean arterial pressure into the normal range, from a normal range to a low mean arterial pressure, from a normal range to a high mean arterial pressure, and from a high

mean arterial pressure into the normal range. This range was selected as it is one of the ranges from Knaus's original APACHE II paper [21].

Throughout the thesis, when a data set was analysed, all available time points were analysed with the exception in the later chapters of first 6 hours. This is because immediately after admission the patient is often unstable and nursing attention is focused more on admitting and stabilising the patient. Omissions in data entry were highest during this period. Otherwise when the scores were tested on data sets, every available time point was analysed.

The frequency of data entry at Glasgow Royal Infirmary ICU is hourly. Occasionally there are additional time points in between. These are sometimes associated with a clinical event e.g. starting haemodialysis.

To illustrate this, here is an example of a longer data set from patient 708 (one of the data sets on which the quantitative score described in chapter 3 was tested on). This patient was in Intensive Care for 3 days (19.45 on day one to 09.00 on day 3). 40 time points worth of data were recorded. These were all hourly except for the first one (19.45) and in day 2 where an extra time point is recorded at 13.37.

In other words, aberrant time points tended to be in addition to the hourly ones. It was unusual for there to be more than 2 extra time points in a 24 hour period i.e. no more than 26.

2.4. Results

Using a 20% change in mean arterial pressure or heart rate as a definition of instability, 65% of dialysis sessions were stable and 35% unstable.

Table 2-1 shows the percentage of changes in mean arterial pressure within an arbitrary normal range, from a low mean arterial pressure into the normal range, from a normal range to a low mean arterial pressure, from a normal range to a high mean arterial pressure and from a high mean arterial pressure into the normal range.

Table 2-1: Change in mean arterial pressure around an arbitrary normal range

	Change from low to normal mean arterial pressure	Change within an arbitrary normal mean arterial pressure range	Change from normal to low mean arterial pressure	Change from normal to high mean arterial pressure	Change from high to normal mean arterial pressure
Percentage of significant changes in mean arterial pressure	40	50	10	0	0

2.5. Discussion

The results of this small study suggest that haemodialysis is not a cardiovascularly unstable therapy. Specifically examining mean arterial pressure, most patients do not move out of what is regarded by many clinicians as a normal blood pressure. Further, it appears that mean arterial pressure may improve form low to normal in a significant proportion of cases. The mechanism may be due to removal of pro-inflammatory cytokines such as tumour necrosis factor- alpha, interleukin-1 beta and interleukin-6 [89]. The first dialysis session at the Royal Infirmary ICU is limited to 2 hours, and this is the session most often associated with instability. Risk factors which may contribute to this are patients who are already hypovolaemic, have valvular heart disease, poor left ventricular systolic function, patients greater than 65 years of age and patients with diabetic autonomic neuropathy.

As described previously, many of the studies embracing the stability of continuous techniques have small numbers of patients, and are crossover in nature. It is difficult under these circumstances to standardise other factors which may have an effect on

cardiovascular stability. These include the dialysate buffer, calcium, temperature, patient position, and the degree of purity of the water used [90, 91].

This small study had a number of weaknesses. Although it did appear that haemodialysis was not on the whole, a cardiovascularly unstable therapy, an arbitrary 20% change in heart rate or blood pressure is a very crude measure. It does not take into account the amount of physiological support the patient is receiving at the time. This might include inotropes, vasoconstrictors, or fluid boluses. On the other hand, when apparently unstable, the patient could be receiving a bolus of anaesthetic, or analgesic agent, or be undergoing ultrafiltration. None of these was standardised in this experiment.

To investigate this clinical problem further and to improve on this simple experiment, I concluded that I would have to design a more comprehensive score which took into account the amount of physiological or pharmacological support the patient was receiving. This was the aim of the work described in the rest of this thesis.

This work is being taken forward by another MD student who is planning a much larger study looking at both haemodialysis and haemofiltration and outcome. This modality is now possible at Glasgow Royal Infirmary ICU following a recent merger with a unit predominantly using this therapy.

2.6. Conclusion

From a simple experiment focusing on the measurement of two key cardiovascular parameters, there is a signal that it may not be an unstable therapy. This will need to be confirmed or refuted by repeating the experiment with a much larger number of data sets. Further, a scoring system which takes into account the amount of physiological support the patient is receiving at the time would have to be derived.

2.7. Acknowledgements

I co-conceived the idea for this study, undertook a review of the literature, and decided on both the cardiovascular parameters to be recorded and the ranges to be used. I assisted in the selection of appropriate patients to be analysed. The appropriate data were extracted and anonymised by the CareVue administrators. The data were pre-processed and then interrogated in terms of the parameters described by our computing scientist colleagues. The analysis of the results and the need for a new score were undertaken jointly by Prof. Kinsella and myself.

Chapter 3: Development of the Quantitative Score

3.1. Abstract

3.1.1. Background

In a simple experiment to quantify cardiovascular stability, preliminary results suggested that haemodialysis is not as cardiovascularly unstable a therapy as previously imagined. However, the experiment did not take into account the amount of physiological or pharmacological support the patient was receiving. To answer this question properly, a quantitative score which took these factors into account would have to be developed.

3.1.2. Methods

Physiological parameters were separated into those recorded at regular intervals and those recorded intermittently. After an extensive literature search, ranges were defined for each parameter, the more points being scored, the greater the derangement. Two parameters (mean arterial pressure and oxygen saturation) were then weighted against a range of pharmacological and physiological variables.

3.1.3. Results

Adding the weighted score for mean arterial pressure and oxygen saturation to the other physiological parameters recorded at regular intervals gives an overall score of cardiovascular instability. This is therefore weighted and influenced by the amount of pharmacological and physiological support the patients is receiving. The score was run in some hypothetical clinical scenarios, resulting in some of the weightings being altered to reflect clinical experience more closely. The outcome was the final quantitative score.

3.1.4. Conclusion

I had developed a novel quantitative score summarising the cardiovascular state of a patient. Unlike many currently available scoring systems, it takes into account the amount of pharmacological and physiological support the patient is receiving.

3.2. Introduction

3.2.1. Need for a quantitative score

As illustrated by the attempt to answer the question of whether haemodialysis is a cardiovascularly unstable therapy, I had concluded that I would have to design a new quantitative scoring system for cardiovascular instability. This was because, on reviewing the literature, one of the major problems of currently available scores is that they do not take into account, or only partially take into account, the level of physiological support the patient is receiving. For example, in a particular score a patient could have a "normal" blood pressure, but simultaneously be receiving large quantities of inotropes, vasoconstrictors and fluid boluses to maintain this apparent normality. The score would not therefore adequately quantify the severity of the underlying physiological disturbance. By taking into account the amount of pharmacological or other support the patient was receiving to maintain a range of physiological parameters at a particular level, the quantitative score was the first attempt to overcome the shortcomings of currently available scores.

3.3. Methods

3.3.1. Division of the parameters into ranges and their basic unweighted score

In the construct of a new model, Ridley states that it should "be based on a small number of explanatory variables that are routinely collected." [92]. On review of the electronic patient record, CareVue, the parameters regularly displayed at the bedside and (reliably) recorded were oxygen saturation, inspired oxygen concentration, heart rate, mean arterial pressure, propofol and alfentanil sedation, fluid administration, temperature, urine output and inotrope doses. Further, many of these parameters are those which form the physiological components of currently available scores e.g. the acute physiology score in APACHE II, multiple organ dysfunction score and SAPS II.

I therefore separated the parameters in my quantitative score into key ones which are recorded at regular intervals (above), and those recorded only intermittently in the Intensive Care Unit. The parameters only intermittently recorded are central venous pressure, cardiac output, cardiac index, stroke volume, stroke volume variance, systemic vascular resistance, systemic vascular resistance index and oxygen delivery.
Along the X-axis I divided each parameter into 3 broad ranges i.e. low abnormal range, normal, and high abnormal range. If appropriate, the low abnormal range and the high abnormal range were divided into 3 further subdivisions scoring +1, +2 and +3 points

respectively, if increasingly abnormal in the high range or increasingly abnormal in the low range. The normal range scored 0. Obviously, the ranges are subjective, but I undertook an extensive literature search to review other scores and the physiological limitations in an adult to inform my decisions about the upper and lower limits for each range in the score. This is represented in Table 3-1.

Table 3-1: Divisions of ranges

Low Abnormal Range			Normal	High Abnormal Range		
Subdivision	Subdivision	Subdivision		Subdivision	Subdivision	Subdivision
+3 Points	+2 Points	+1 Point	0 Points	+ 1 Point	+ 2 Points	+ 3 Points

3.3.2. Derivation of the ranges for the parameters recorded at regular intervals

I shall now describe the justification for the ranges for each parameter, dealing first with the parameters recorded at regular intervals (Tables 3-2 to 3-6). Where there is an explanation for the range, then it was based on ranges used in a previous score or is a standard physiological fact. Otherwise the ranges represent opinion based on my clinical experience. To avoid confusing numerical reference numbers with ranges in the score, the references in the table are designated a letter, explained in the narrative below each table and at that point given a corresponding numerical designator. All references can be viewed at the end of the thesis.

Table 3-2: Mean arterial pressure (mmHg)

Low Abnormal Range			Normal	High Abnormal Range		
<40	40-49	50-69 (A)	70-109 (A)	110-129 (A)	130-159 (A)	≥160 (A)
+3	+2	+1	0	+1	+2	+3

Low Abnormal Range

+1 Range is as per APACHE II score (A) [21]

Normal Range

Range is as per APACHE II score (A) [21]

High Abnormal Range

+1 Range is as per APACHE II score (A) [21]

+2 Range is as per APACHE II score (A) [21]

+3 Range is as per APACHE II score (A) [21]

Table 3-3: Heart Rate (Beats per minute)

Low Abnormal Range			Normal	High Abnormal Range		
≤ 35	36-39 (C)	40-49 (B)	50-59	91-140 (D)	141-179 (E)	≥ 180 (F)
+3	+2	+1	0	+1	+2	+3

Low Abnormal Range

+1 Heart rate of less than 50 is generally regarded as the lower limit of a sinus bradycardia (B) [93]

+2 Heart rate of 39 is generally regarded as the "normal" upper limit of escape rhythm in complete heart block (C) [94]

High Abnormal Range

+1 Heart rate >90 is one of the 4 SIRS criteria (D) **[95]**

+2 Heart rate>140 is the point where ventricular filling in early diastole becomes compromised (E) [96]

+3 Heart rate of 180 is approximately the maximum ventricular rate in man (F) [97]

Table 3-4: Oxygen Saturation, SpO2 (%)

Low Abnormal Range			Normal	High Abnormal Range		
<75 (I)	75-89 (H)	90-94 (G)	95-100	N/A	N/A	N/A
+3	+2	+1	0	-	-	-

Low Abnormal Range

+1 SpO2 of 95% is regarded as the lower limit of normal in health (also, the definition of hypoxaemia is a PaO2 of <80mmHg – approximately an SpO2 of 95% while breathing air) (G) [98]

+2 SpO2 below 90% (PaO2 of 60mmHg) is the definition of respiratory failure (H) [99].

+3 SpO2 of 75% is the mixed venous oxygen saturation (I) [100]

High Abnormal Range

There is no score, as saturations above 100% are physiologically impossible

Table 3-5: Urine Output (mls/h)

Low Abnormal Range			Normal	High Abnormal Range		
<2 (J)	2-19 (J)	20-35	>35	N/A	N/A	N/A
+3	+2	+1	0	-	-	-

Low Abnormal Range

+2 Oliguria is usually described as a urine output <20mls/h (J) [101]

+3 Anuria is usually described as a urine output of <50mls/day (approximately 2mls/h) (J) [101]

Table 3-6: Temperature (ºC)

Low Abnormal Range			Normal	High Abnormal Range		
<32 (L)	32-35 (K)	35-36 (D)	36-38	>38-40 (D)(L)	>40-42.1 (K)	>42.1 (L)
+3	+2	+1	0	+1	+2	+3

Low Abnormal Range

+1 Temperature <36(ºC) is one of the 4 SIRS criteria (D) [95]

+2 Temperature <35(ºC) is the definition of hypothermia (K) [102]

+3 Temperature <32(ºC) is the definition of moderate hypothermia (K) [102]

High Abnormal Range

+1 Temperature >38(ºC) is one of the 4 SIRS criteria (D) [95] (with standard definition of a fever being a temperature >38.3(ºC) (L) [103].

+2 Temperature >40(ºC) is the definition of hyperpyrexia (K) [102].

+3 Temperature >42.1(ºC) is the temperature at which cell damage may occur (L) [103].

3.3.3. Final unweighted score for the parameters recorded at regular intervals

The final unweighted score for the parameters recorded at regular intervals is shown in table 3-7.

Table 3-7: Final unweighted score for the parameters recorded at regular intervals

Mean arterial pressure (mmHg)						
Low Abnormal Range			Normal	High Abnormal Range		
<40	40-49	50-69	70-109	110-129	130-159	≥160
+3	+2	+1	0	+1	+2	+3

Heart rate (beats per minute)						
Low Abnormal Range			Normal	High Abnormal Range		
≤ 35	36-39	40-49	50-90	91-140	141-179	≥180
+3	+3	+1	0	+1	+1	+3

Oxygen Saturation, SpO2 (%)						
Low Abnormal Range			Normal	High Abnormal Range		
<75	75-89	90-94	95-100	N/A	N/A	N/A
+3	+2	+1	0	-	-	-

Urine Output (mls/h)						
Low Abnormal Range			Normal	High Abnormal Range		
<2	2-19	20-35	>35	N/A	N/A	N/A
+3	+2	+1	0	-	-	-

Temperature (ºC)						
Low Abnormal Range			Normal	High Abnormal Range		
<32	32-35	35-36	36-38	>38-40	>40-42.1	>42.1
+3	+2	+1	0	+1	+2	+3

To calculate the score, the value assigned to each of the 5 key parameters is summed to give the final score. By way of illustration, a patient with a mean arterial blood pressure of 55, a heart rate of 145, a normal oxygen saturation, a urine output of 18mls/h and a normal temperature would score 4 points in the score developed thus far.

3.3.4. Derivation of the ranges of parameters recorded intermittently

Similar methodology for deriving ranges for the parameters recorded at regular intervals, was applied to those recorded only intermittently (tables 3-8 to 3-15)

Table 3-8: Central Venous Pressure, CVP (cmH_2O)

Low Abnormal Range			Normal	High Abnormal Range		
-	-	0-2	3-10	11-18	19-24	>24
N/A	N/A	+1	0	+1	+2	+3

For this parameter, all the points were judgements based on my clinical experience.

Table 3-9: Cardiac Output, CO (l/min)

Low Abnormal Range			Normal	High Abnormal Range		
<2.5	2.5-3.4	3.5-4.9	5-6 (I)	6.1-15	15.1-29.9	≥30 (M)
+3	+3	+1	0	+1	+2	+3

Low Abnormal Range

+2 3.5l/min chosen as it would be the cardiac output generated with a heart rate of 50 (lowest generally accepted for a sinus bradycardia) and a stroke volume of 70mls.

Normal Range

5-6l/min is the normal cardiac output in a 70kg man (I) [100]

High Abnormal Range

+3 30l/min is the maximum cardiac output which can be achieved in a healthy adult under conditions of extreme exercise (M) [104]

Table 3-10: Cardiac Index, CI ($l/min/m^2$)

Low Abnormal Range			Normal	High Abnormal Range		
<1.5	1.5-2.0	2.1-2.9	3-3.5 (I)	3.6-8.8	8.9-16.9	≥17
+3	+2	+1	0	+1	+2	+3

Ranges were derived from those given for cardiac output, taking 1.7 m^2 as body surface area, e.g. 3-3.5 $l/min/m^2$ would be a normal cardiac output using this figure.

Table 3-11: Stroke Volume, SV (mls)

Low Abnormal Range			Normal	High Abnormal Range		
<40	40-54	55-69	70-80 (I)	81-90	91-100	>100 (I)
+3	+2	+1	0	+1	+2	+3

Normal Range

70-80 mls is regarded as a normal stroke volume in a 70kg man (I) [100]

High Abnormal Range

+3 Value of 100mls chosen as, under normal circumstances, this is at the start of the plateau of the Frank-Starling curve (I) [100]

Table 3-12: Stroke Volume Variance, SVV (%)

Low Abnormal Range			Normal	High Abnormal Range		
-	-	-	<10 (N)	10-13	13-18	18
N/A	N/A	N/A	0	+1	+2	+3

The values chosen were clinical except for normal value of <10% where patients are unlikely to be preload responsive (N) [105]

Table 3-13: Systemic Vascular Resistance, SVR (dynes x s/cm^5)

Low Abnormal Range			Normal	High Abnormal Range		
<450 (O)	451-700	701-899	900-1400	1401-1600	1601-1800	>1800
+3	+2	+1	0	+1	+2	+3

Low Abnormal Range

+3 Value of 450dynes.s/cm5 chosen, as it has been shown that below this level mortality is much greater, irrespective of aetiology (O) [106].

Table 3-14: Systemic Vascular Resistance Index, SVRI (dynes × s/cm^5/m^2)

Low Abnormal Range			Normal	High Abnormal Range		
<765	765-1189	1190-1529	1530-2380	2381-2720	2721-3060	>3060
+3	+2	+1	0	+1	+2	+3

These values were derived from Systemic Vascular Resistance (SVR) using 1.7 m^2 as the value for body surface area when calculating cardiac index

Table 3-15: Oxygen Delivery, DO2 (mls/min)

Low Abnormal Range			Normal	High Abnormal Range		
<300	300-699	700-999	1000-1200	1201-3000	3001-5980	>5980
+3	+2	+1	0	+1	+2	+3

These figures were derived from those given for Cardiac Output assuming the oxygen content of arterial blood to be 20mls/100mls

3.3.5. Final unweighted score for the parameters recorded intermittently

The final unweighted score for those parameters recorded intermittently is recorded in table 3-16.

Table 3-16: Final unweighted score for the parameters recorded intermittently

Central Venous Pressure, CVP (cmH_2O)						
Low Abnormal Range			Normal	High Abnormal Range		
-	-	0-2		11-18	19-24	>24
N/A	N/A	+1	0	+1	+2	+3

Cardiac Output, CO (l/min)						
Low Abnormal Range			Normal	High Abnormal Range		
<2.5	2.5-3.4	3.5-4.9	5-6	6.1-15	15.1-29.9	≥30
+3	+3	+1	0	+1	+2	+3

Cardiac Index, CI ($l/min/m^2$)						
Low Abnormal Range			Normal	High Abnormal Range		
<1.5	1.5-2.0	2.1-2.9	3-3.5	3.6-8.8	8.9-16.9	≥17
+3	+2	+1	0	+1	+2	+3

Stroke Volume, SV (mls)						
Low Abnormal Range			Normal	High Abnormal Range		
<40	40-54	55-69	70-80	81-90	91-100	>100
+3	+2	+1	0	+1	+2	+3

Stroke Volume Variance, SVV (%)						
Low Abnormal Range			Normal	High Abnormal Range		
-	-	-	<10	10-13	13-18	18
N/A	N/A	N/A	0	+1	+2	+3

Systemic Vascular Resistance, SVR (dynes x s/cm^5)						
Low Abnormal Range			Normal	High Abnormal Range		
<450	451-700	701-899	900-1400	1401-1600	1601-1800	>1800
+3	+2	+1	0	+1	+2	+3

Systemic Vascular Resistance Index, SVRI (dynes × $s/cm^5/m^2$)						
Low Abnormal Range			Normal	High Abnormal Range		
<765	765-1189	1190-1529	1530-2380	2381-2720	2721-3060	>3060
+3	+2	+1	0	+1	+2	+3

Oxygen Delivery, DO2 (mls/min)						
Low Abnormal Range			Normal	High Abnormal Range		
<300	300-699	700-999	1000-1200	1201-3000	3001-5980	>5980
+3	+2	+1	0	+1	+2	+3

3.3.6. Adjustment of the ranges of parameters to take into account the level of physiological or pharmacological support

In the next stage of development, I selected the two physiological parameters which I felt were not adequately captured by existing scores as they can be substantially affected by other physiological processes, or the degree of pharmacological support. These are mean arterial pressure and oxygen saturation. Table 3-17 shows the main factors which can be affected by other parameters.

Table 3-17: Factors that can either positively or negatively change mean arterial pressure or oxygen saturation

Parameter	**Mean Arterial Pressure**	**Oxygen Saturation**
Factors negatively (-ve) or positively (+ve) affecting the parameter	Adrenaline (+ve) Noradrenaline (+ve) Fluid Input (+ve) Propofol (-ve) Alfentanil (-ve)	Inspired Oxygen Fraction (+ve) Positive End Expiratory Pressure (+ve) Continuous Positive Airway Pressure (+ve)

For each factor I gave points based on my clinical judgement on how much that factor, at a particular level, would affect either the mean arterial pressure or the oxygen saturation (tables 3-18 and 3-19). The ranges for the drug doses in two instances were based on maximum recommended amounts of the drug, or doses used in previous research. Specifically 0.5mcg/kg/min of adrenaline or noradrenaline (2.1mg/h in a 70kg man) is the upper limit quoted for cardiovascular support in the critically ill (P) [107]. 4mls/kg/hour of propofol (or 280mg/h in a 70kg man) is the maximum rate recommended for sedation in ICU in the British National Formulary (Q) [108].

Table 3-18: Weighting of the various factors on Mean Arterial Pressure

		Low Abnormal Range			Normal	High Abnormal Range		
		Mean Arterial Pressure (MAP) (mmHg)						
		<40	40-49	50-69	70-109	110-129	130-159	≥160
Unweighted Score		3	2	1	0	1	2	3
Adrenaline (mg/h)	<0.2	0	0	0	0	0	0	0
	0.2-1.0	4	3	2	1	1	1	1
	1.1-2.0	5	4	3	2	2	2	2
	>2.0 (P)	6	5	4	3	3	3	3
Noradrenaline (mg/h)	<0.1	0	0	0	0	0	0	0
	0.1-1.0	4	3	2	1	1	1	1
	1.1-2.1	5	4	3	2	2	2	2
	>2.1 (P)	6	5	4	3	3	3	3
Fluid Input (mls/h)	0-125	0	0	0	0	0	0	0
	126-250	4	3	2	1	1	1	1
	251-500	5	4	3	2	2	2	2
	>500	6	5	4	3	3	3	3
Propofol (mg/h)	<10	0	0	0	0	0	0	0
	10-100	-1	-1	-1	-1	-2	-3	-4
	110-280	-2	-2	-2	-2	-3	-4	-5
	>280 (Q)	-3	-3	-3	-3	-4	-5	-6
Alfentanil (mg/h)	<0.5	0	0	0	0	0	0	0
	0.5-2.5	-1	-1	-1	-1	-2	-3	-4
	3-4	-2	-2	-2	-2	-3	-4	-4
	>4	-3	-3	-3	-3	-4	-5	-6

Table 3-19: Weighting of the various factors on Oxygen Saturation

		Oxygen Saturation (Sp02) (%)						
		<75	75-89	90-94	95-100	-	-	-
Unweighted Score	Air	3	2	1	0	-	-	-
Inspired Oxygen Fraction (FiO2)	0.22-0.49	4	3	2	1	-	-	-
	0.5-0.79	5	4	3	2	-	-	-
	≥0.80	6	5	4	3	-	-	-
PEEP / CPAP (cmH_20)	0-5	0	0	0	0	-	-	-
	6-8	4	3	2	1	-	-	-
	9-11	5	4	3	2	-	-	-
	≥12	6	5	4	3	-	-	-

3.4. Results

3.4.1. First version of completed quantitative score

The parameters continuously recorded (with some now weighted to take into account the level of physiological or pharmacological support) were combined to produce the first version of the completed quantitative score of cardiovascular instability as shown in figure 3-1.

Figure 3-1: First version of completed quantitative score

Parameters recorded at regular intervals

		Low Abnormal Range			Normal	High Abnormal Range		
		Mean Arterial Pressure (MAP) (mmHg)						
		<40	40-49	50-69	70-109	110-129	130-159	≥160
Unweighted Score		3	2	1	0	1	2	3
Adrenaline (mg/h)	<0.2	0	0	0	0	0	0	0
	0.2-1.0	4	3	2	1	1	1	1
	1.1-2.0	5	4	3	2	2	2	2
	>2.0	6	5	4	3	3	3	3
Noradrenaline (mg/h)	<0.1	0	0	0	0	0	0	0
	0.1-1.0	4	3	2	1	1	1	1
	1.1-2.1	5	4	3	2	2	2	2
	>2.1	6	5	4	3	3	3	3
Fluid Input (mls/h)	0-125	0	0	0	0	0	0	0
	126-250	4	3	2	1	1	1	1
	251-500	5	4	3	2	2	2	2
	>500	6	5	4	3	3	3	3
Propofol (mg/h)	<10	0	0	0	0	0	0	0
	10-100	-1	-1	-1	-1	-2	-3	-4
	110-280	-2	-2	-2	-2	-3	-4	-5
	>280	-3	-3	-3	-3	-4	-5	-6
Alfentanil (mg/h)	<0.5	0	0	0	0	0	0	0
	0.5-2.5	-1	-1	-1	-1	-2	-3	-4
	3-4	-2	-2	-2	-2	-3	-4	-4
	>4	-3	-3	-3	-3	-4	-5	-6

Heart Rate (HR) (Beats/min)						
<35	35-39	40-49	60-90	91-140	141-179	≥160

		Oxygen Saturation (Sp02) (%)						
		<75	75-89	90-94	95-100	-	-	-
Unweighted Score	Air	3	2	1	0	-	-	-
Inspired Oxygen Fraction (FiO2)	0.22-0.49	4	3	2	1	-	-	-
	0.5-0.79	5	4	3	2	-	-	-
	≥0.80	6	5	4	3	-	-	-
PEEP / CPAP (cmH_20)	0-5	0	0	0	0	-	-	-
	6-8	4	3	2	1	-	-	-
	9-11	5	4	3	2	-	-	-
	≥12	6	5	4	3	-	-	-

Urine Output (mls/h)						
<2	2-19	20-35	>35	-	-	-

Temperature (ºC)						
<32	32-35	35.1-35.9	36-38	38.1-40	40.1-42.1	>42.1

Parameters Intermittently Recorded

	Low Abnormal Range			Normal	High Abnormal Range		
Score	3	2	1	0	1	2	3
Central Venous Pressure (CVP) (cmH_20)	<-6	-5 to -3	0-2	2-10	11-18	19-24	>24
Cardiac Output (CO) (l/min)	<2.5	2.5-3.4	3.5-4.9	5-6	6.1-15	15.1-29.9	≥30
Cardiac Index (CI) ($l/min/m^2$)	<1.5	1.5-2.0	2.1-2.9	3-3.5	3.1-8.8	8.9-16.9	≥17
Stroke volume (SV) (mls)	<40	40-54	55-69	70-80	81-90	91-100	>100
Stroke Volume Variance (%)	-	-	-	<10	10-13	13-18	18
Systemic Vascular Resistance (SVR) (dynes × s/cm^5)	<450	451-700	701-899	900-1400	1401-1600	1601-1800	>1800
Systemic Vascular Resistance Index (SVRI) (dynes × $s/cm^5/m^2$)	<765	765-1189	1190-1529	1530-2380	2381-2720	2721-3060	>3060
Oxygen Delivery (D02) (mls/min)	<300	300-699	700-999	1000-1200	1201-3000	3001-5980	>5980

3.4.2. An example of calculating a score

To give an example of how the final score works, the first points comprise the parameters recorded at regular intervals.

- Take the unweighted score of mean arterial pressure and add or subtract points depending on the amount of adrenaline, noradrenaline, fluid, propofol or alfentanil the patient is receiving.
- Add the score for heart rate.
- Add the score for oxygen saturation, ALREADY weighted by points for level of inspired oxygen concentration, to the weighting for the amount of PEEP or CPAP.
- Add the score for temperature.
- Add the score for urine output.

This gives a total score of -3 to 48 for parameters recorded at regular intervals. It may be the case that this is all that is being recorded in the patient, in which case the score terminates here. If any of the parameters recorded intermittently are present, they can then be added to this score but, for the score to be reliable over time, they have to be present each time the score is recalculated. An example is shown in table 3-20.

Table 3-20: An example of calculating a patient's score

Parameter	**Value**	**Score**
Unweighted Mean Arterial Pressure	53	+1
Adrenaline dose (When MAP 53)	1.8mg/h	+3
Noradrenaline dose (When MAP 53)	1.0mg/h	+2
Fluid rate (When MAP 53)	250mls/h	+2
Propofol dose (When MAP 53)	130mg/h	-2
Alfentanil dose	3mg/h	-2
Heart Rate	130 beats/min	+1
Oxygen Saturation already weighted by inspired oxygen concentration	83% Saturated with FiO2 of 0.9	+5
Oxygen Saturation weighted by amount of PEEP/ CPAP	83% Saturated with PEEP of 12	+5
Urine Output	8mls/hour	+2
Temperature	39.3 (ºC)	+1

The total score for this patient at that moment is this sum of all of the above scores and, is 18.

3.4.3. Clinical reflection and alteration of parameters

Some adjustments were made to the first version of the completed score to produce a final version, for the following reasons. I ran various virtual clinical scenarios to ascertain if the new score gave clinically credible results. By using this approach, now with the means to utilise the score as a whole, it emerged that some parameters needed to make a greater or lesser contribution to the score.

For example, patients with a mean arterial pressure of 41, with no inotropic or vasoconstrictor support, would score 2 points. Further, if their urine output was 19 mls/hour they would also score two points. In a score capturing cardiovascular instability, borderline urine output, while noteworthy and requiring action, is not as acutely important or potentially as dangerous. Therefore I changed the unweighted score of mean arterial pressure to give more importance to extreme hypotension.

Similarly, if patients were hypertensive, yet receiving adrenaline, noradrenaline or large volumes of fluid, they would gain points. This is an unlikely clinical scenario and could possibly represent excessive dosing, or patient recovery. To represent this in the score, I now deducted points if hypertensive whilst simultaneously receiving adrenaline or noradrenaline, and made the score point neutral if hypertensive and being administered large amounts of fluid.

Until this point, if patients were hypertensive despite large doses of propofol or alfentanil, then they had points deducted. Again, after reflection, it was more clinically credible to gain points, since to be hypertensive despite, e.g., the vasodilating effect of propofol in large amounts, then this represents marked cardiovascular instability. I also took the opportunity to decrease the number of dosing bands of alfentanil from four to three, as I felt that, clinically, alfentanil induces less cardiovascular instability than propofol.

Finally, and for similar reasons to the unweighted mean arterial pressure scores being changed, I also changed the heart rate scores to reflect the importance of extreme bradycardia and tachycardia.

These changes are highlighted in table 21. Changes from the previous version are marked in bold. Note that no changes were made to parameters collected only intermittently.

Table 3-21: Changes to the scores for mean arterial pressure and heart rate

Parameters recorded at regular intervals

		Low Abnormal Range			Normal	High Abnormal Range		
		Mean Arterial Pressure (MAP) (mmHg)						
		<40	40-49	50-69	70-109	110-129	130-159	≥160
Unweighted Score		**7**	**5**	**3**	0	1	2	3
Adrenaline (mg/h)	<0.2	0	0	0	0	**-1**	**-2**	**-3**
	0.2-1.0	4	3	2	1	**-1**	**-2**	**-3**
	1.1-2.0	5	4	3	2	**-1**	**-2**	**-3**
	>2.0	6	5	4	3	**-1**	**-2**	**-3**
Noradrenaline (mg/h)	<0.1	0	0	0	0	**-1**	**-2**	**-3**
	0.1-1.0	4	3	2	1	**-1**	**-2**	**-3**
	1.1-2.1	5	4	3	2	**-1**	**-2**	**-3**
	>2.1	6	5	4	3	**-1**	**-2**	**-3**
Fluid Input (mls/h)	0-125	0	0	0	0	0	0	0
	126-250	4	3	2	1	**0**	**0**	**0**
	251-500	5	4	3	2	**0**	**0**	**0**
	>500	6	5	4	3	**0**	**0**	**0**
Propofol (mg/h)	<10	0	0	0	0	0	0	0
	10-100	-1	-1	**0**	**0**	**1**	**1**	**-1**
	110-280	-2	-2	**-1**	**0**	**1**	**1**	**2**
	>280	-3	-3	**-2**	**0**	**1**	**2**	**2**
Alfentanil (mg/h)	<0.5	0	0	0	0	0	0	0
	0.5-2.1 (R)	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	>2.1	**-1**	**-1**	**-1**	**0**	**1**	**1**	**1**

Heart Rate (HR) (Beats/min)						
<35	35-39	40-49	60-90	91-140	141-179	≥160
5	**3**	1	0	1	**3**	**5**

3.4.4. The final quantitative score

The alterations to the parameters recorded regularly, based upon clinical reflection, could now be couple to the parameters which were only recorded intermittently to give the final quantitative score. This is shown in figure 3-2. The score was now ready to be tested against data from actual Intensive Care Patients. This is the subject for discussion in the next chapter.

Figure 3-2: The final quantitative score

Parameters recorded at regular intervals

		Low Abnormal Range			Normal	High Abnormal Range		
		Mean Arterial Pressure (MAP) (mmHg)						
		<40	40-49	50-69	70-109	110-129	130-159	≥160
Unweighted Score		7	5	3	0	1	2	3
Adrenaline (mg/h)	<0.2	0	0	0	0	-1	-2	-3
	0.2-1.0	4	3	2	1	-1	-2	-3
	1.1-2.0	5	4	3	2	-1	-2	-3
	>2.0	6	5	4	3	-1	-2	-3
Noradrenaline (mg/h)	<0.1	0	0	0	0	-1	-2	-3
	0.1-1.0	4	3	2	1	-1	-2	-3
	1.1-2.1	5	4	3	2	-1	-2	-3
	>2.1	6	5	4	3	-1	-2	-3
Fluid Input (mls/h)	0-125	0	0	0	0	0	0	0
	126-250	4	3	2	1	0	0	0
	251-500	5	4	3	2	0	0	0
	>500	6	5	4	3	0	0	0
Propofol (mg/h)	<10	0	0	0	0	0	0	0
	10-100	-1	-1	0	0	1	1	-1
	110-280	-2	-2	-1	0	1	1	2
	>280	-3	-3	-2	0	1	2	2
Alfentanil (mg/h)	<0.5	0	0	0	0	0	0	0
	0.5-2.1	0	0	0	0	0	0	0
	>2.1	-1	-1	-1	0	1	1	1

Heart Rate (HR) (Beats/min)						
<35	35-39	40-49	60-90	91-140	141-179	≥160
5	3	1	0	1	3	5

		Oxygen Saturation (Sp02) (%)						
		<75	75-89	90-94	95-100	-	-	-
Unweighted Score	Air	3	2	1	0	-	-	-
Inspired Oxygen Fraction (FiO2)	0.22-0.49	4	3	2	1	-	-	-
	0.5-0.79	5	4	3	2	-	-	-
	≥0.80	6	5	4	3	-	-	-
PEEP / CPAP (cmH_20)	0-5	0	0	0	0	-	-	-
	6-8	4	3	2	1	-	-	-
	9-11	5	4	3	2	-	-	-
	≥12	6	5	4	3	-	-	-

Urine Output (mls/h)						
<2	2-19	20-35	>35	-	-	-

Temperature (ºC)						
<32	32-35	35.1-35.9	36-38	38.1-40	40.1-42.1	>42.1

Parameters Intermittently Recorded

	Low Abnormal Range			Normal	High Abnormal Range		
Score	3	2	1	0	1	2	3
Central Venous Pressure (CVP) (cmH_20)	<-6	-5 to -3	0-2	2-10	11-18	19-24	>24
Cardiac Output (CO) (l/min)	<2.5	2.5-3.4	3.5-4.9	5-6	6.1-15	15.1-29.9	≥30
Cardiac Index (CI) ($l/min/m^2$)	<1.5	1.5-2.0	2.1-2.9	3-3.5	3.1-8.8	8.9-16.9	≥17
Stroke volume (SV) (mls)	<40	40-54	55-69	70-80	81-90	91-100	>100
Stroke Volume Variance (%)	-	-	-	<10	10-13	13-18	18
Systemic Vascular Resistance (SVR) (dynes × s/cm^5)	<450	451-700	701-899	900-1400	1401-1600	1601-1800	>1800
Systemic Vascular Resistance Index (SVRI) (dynes × $s/cm^5/m^2$)	<765	765-1189	1190-1529	1530-2380	2381-2720	2721-3060	>3060
Oxygen Delivery (D02) (mls/min)	<300	300-699	700-999	1000-1200	1201-3000	3001-5980	>5980

3.5. Discussion

In this chapter the design of a new quantitative scoring system of instability in Intensive Care patients is described, which takes into account the amount of physiological and pharmacological support the patient is receiving. There were several advantages to this approach. The score was diagnosis independent (unlike the APACHE II system), could be repeated at regular intervals to give a trend of improvement or deterioration and was simple to calculate at the bedside. This score comprised ranges of physiological and pharmacological parameters, some of which were weighted against each other. There are various ways that this could have been undertaken. There were no similar scores available for comparison to design this new construct. As a pragmatic starting point the literature of currently available scores which use similar (unweighted) parameters was reviewed. Ranges were used from some of these scores and others devised from either common physiological facts (e.g. maximum ventricular rate before filling is impaired) and my own clinical opinion. In other words the process started with a combination of single "expert" opinion, common physiological facts and ranges from previous physiologically based scores (albeit devised for a different purpose).

There are other techniques that could have been employed to make this process more sophisticated. Brain storming was described by the advertising writer Alex Osborn as a means of using the brains to "storm a problem." [110]. The premise of brainstorming is that members of a group generate as many ideas on a topic that they can. Osborne defined 4 rules for a session. Members of a session should generate as many ideas as they can (and not worry about quality), further ideas should be generated from thoughts from other participants, judgement on ideas should be deferred and there should be no criticism of other's thoughts. Related to brainstorming sessions are group interviews in so-called focus groups. These are more structured in nature. The Delphi method was designed mainly by Dalkey and Helmer in the 1960s [111]. In this, experts' opinion to a problem is sought in two or sometimes more rounds. After each round a researcher provides a summary of the answers given in the previous round with reasons. In the next round the experts are asked to revise their answers in the light of the other opinions. Over time the range of opinion is decreased and consensus reached. This has the advantage over single surveys, whereby a researcher has to summarise opinion and there is more chance of bias. The anonymity achieved in the Delphi exercise is useful. In brainstorming sessions there is a danger of participants not speaking freely if there is a dominant member of the group. Individual interviews (which can be structured "qualitative" or unstructured) are another option. The

risk with this technique is that the researcher leading the sessions can unintentionally introduce bias by the direction in which the questioning follows. Structured interviews are preferred over unstructured as they are more reproducible, systematic, transparent and reliable.

The quantitative score developed has two components, one based on regularly recorded parameters and a possible score for parameters recorded intermittently. The score is tested in the next chapter using the first part of the score only. The second part is not used. Before this testing occurred, various problems were considered. If a parameter is only recorded intermittently then the score could change because this parameter is or is not being recorded, rather than the patient improving or deteriorating. A possible solution is to ascertain which of the intermittent parameters are being recorded, if the is occurring regularly, and at what frequency. These scores could be calculated, the maximum score for these intermittent parameters calculated, and added to the basic score out of -3 to + 48. The final score would be out of a value dependent on the parameters being recorded in a particular patient. This enhanced score could be tested against the basic score in clinical scenarios to ascertain if it better fits with clinical opinion. The reality however is that different intermittent parameters are likely to be recorded at different times e.g. a nurse calibrates the cardiac output monitor and takes a reading at one time point but measures CVP at another.

A related problem that applies to the first part of the score is unmeasured variables. It is unlikely, but not impossible that an Intensive Care patient would have one or more of the key parameters measured e.g. heart rate, blood pressure, oxygen saturation and so on (and this was one of the reasons they were chosen for the first part of the score). The solution is an algorithm that detects which parameters are present prior to analysis. If there is a missing parameter the only solution is that the final score is out of a value of less than 48. This would have to be made clear as it could affect how well the model fits real patient data in future testing. Another difficulty is the mathematically related intermittent variables. These are cardiac output / cardiac index and systemic vascular resistance / systemic vascular resistance index. There are three possibilities, namely the basic variable is being recorded, the mathematically derived index of the variable is being recorded, or both. If the latter then an algorithm could be designed to ignore the basic variable and only use the index. Further, in the score as it stands just now, for these two parameters (and their related index) the points scored are the same at different levels of derangement, so it is perhaps less important if there is inconsistency as to which is recorded.

To develop the score further to include parameters recorded intermittently would have required more algorithm development. Although there are possible solutions, to ultimately have a score that produces different value dependent on what parameters are recorded is unsatisfactory. These are the reasons this part of the score was not developed further at this time.

Various issues would have to be overcome prior to testing the first part of the quantitative score (parameters recorded at regular intervals) on real patient data sets. Data points are predominantly recorded at hourly intervals at the Royal Infirmary. There are occasional data points in between the hour if something of clinical importance has occurred. A greater frequency of recordings would have been useful but was not practical. A particular disadvantage of this approach is evident in a later chapter where I describe the association and prediction of troponin positive events in Intensive Care. A weakness of this model is that with only hourly data an extreme physiological event could occur within that period and be missed by the model. Greater frequency of data recording may be possible in the future as the CareVue system has an auto chart function and it may be possible to record this data for research purposes.

The potential for missing data affecting the score was a concern. That is to say the score increasing or decreasing not because the patient's clinical condition had changed but because a parameter was omitted. In the next chapter I describe the algorithm employed to overcome the missing data, the effect the algorithm had and possible further solutions to improve on this problem. Another difficulty was a parameter that had not just been recorded on an hourly time point, but one which had been omitted altogether. This was one of the reasons that such key parameters were chosen for the first part of the score. It is highly unlikely that a nurse at the bed space would consistently not record heart rate or mean arterial pressure. In this event the score would still be able to show trends but be out of a total of less than 48 (as long as the parameter was consistently missing).
The weighting of the quantitative score is difficult as there is no reference standard to model against. Parameters were selected and weighted after a literature review i.e. a single "expert" based approach. As described earlier in the chapter there are a number of techniques that could have been used at an earlier stage e.g. brainstorming sessions or Delphi exercises rather than rely on one clinician. This is one of the weaknesses of the score. Had a greater number of clinicians been involved, using one of the above techniques, the methodology would have been similar (though not on the same scale or degree of sophistication) used by Knaus to develop the APACHE score [22]. In this, he led a

panel of clinicians who, after a literature review, weighted 34 variables. They were able to simplify the number and change some of the weightings in APACHE II [21] using a process of multivariate analysis. This was possible because they were using survival prediction as an endpoint. A further example is the sequential organ failure assessment (SOFA) score [38]. This was developed in 1994 by Vincent et al. in October 1994 during a consensus meeting of the European Society of Intensive Care Medicine meeting in Paris, and revised in 1996. Internationally renowned critical care doctors from over Europe were invited. The aim of this score was to describe in a quantitative manner the degree of organ dysfunction over time in patients. Interestingly it was not designed to compete with other severity indices and predict outcome, rather it was to describe morbidity secondary to critical illness. The clinicians, after a literature review, limited the number of organ systems used in the score to six. Each organ system could score between 0 and 4 points (most abnormal) and ranges for the single parameter used as a marker for each failing organ system were decided by the clinicians. Although not designed to predict mortality, a prospective analysis of 1449 patients in 40 Intensive Care Units in 20 countries did show a correlation between raised SOFA scores and poor outcome across the 6 organ systems.

The starting point for the two scores above was consensus opinion. An alternative statistical approach was used in the development of the Logistic Organ Dysfunction Score (LODS). This was designed by Le Gall et al. in 1996 [112]. Rather than start from consensus opinion they compiled a database comprising 13152 admissions from 12 countries and 137 Intensive Care Units. Using multiple logistical regression, 12 variables were identified and weightings determined based on prognostic significance to capture the function of 6 different organ systems. The authors argued that organ dysfunction was being determined objectively by this methodology rather than by expert opinion. The score was subsequently validated on 2605 patients from the database. There was a strong correlation with mortality.

In selecting a methodology, there did not exist a large database of Intensive Care Admissions. Using statistical logistic regression was not practical. Further, the weightings in the LODS score were still based against prognostic significance. Further, the developed score was one of cardiovascular stability, rather than a predictor of outcome. An expert based approach is valid. The weakness of this score was not using more clinicians in the initial design of the construct. Also, in the absence of a reference standard, the score would either have had to be justified as a predictor of, for example, outcomes or tested against expert opinion. That is to say in clinical scenarios where a patient deteriorates or improves,

is the score and the weightings of the parameters credible to a group of clinicians. On reflection greater involvement of more clinicians would have undoubtedly improved the score. Nevertheless, the focus moved to develop a qualitative score. Rather than design an arbitrary quantitative score, and then justify it clinically, an alternative approach worth exploring was to base the actual score on clinical expertise.

In the quantitative score developed thus far a single researcher opinion was used to select the parameters, define ranges and add weightings. Although as can be seen in the next chapter the score does, on the face of it, show patient improvements and deteriorations, the single researcher approach is its major weakness and could be improved. As I have discussed, the design of a new score in the absence of a standard for comparison is extremely challenging.

Future work to refine the score will be as follows:

Senior clinicians not previously involved in this work will be invited to a brainstorming session. Using the regularly recorded parameters selected as a starting point, consensus will be sought from the clinicians whether they thought these were reasonable. Additions or deletions could be made at this point. In the next stage going through each parameter in turn, possibly in the context of clinical scenarios consensus would be sought on parameter ranges. These parameters and ranges would be collated in a document and sent out to a further group of clinicians for comment and potential further refinement.

In the absence of a reference standard the consensus score would initially have to be tested against expert opinion. That is to say, in a variety of scenarios, is the score credible to more clinicians not involved in its development, in particular the emphasis given to particular parameters. (As an aside, as the score was not designed to predict mortality, statistical regression techniques applied to a large prospectively collected database would be less useful.)

3.6. Conclusion

I had developed a quantitative score which took into account the amount of physiological and pharmacological support the patient was receiving. This was through the interaction of common physiological and pharmacological factors on two key physiological parameters, viz. mean arterial pressure and oxygen saturation. This now required to be tested on real patient data.

3.7. Acknowledgements

The work in this chapter was entirely my own.

Chapter 4: Handling missing data for large volume analysis, testing of the Quantitative Score and need for a Qualitative Score

4.1. Abstract

4.1.1. Background

The quantitative score described in the last chapter was now ready to be tested against real patient data sets. Before this could occur, the raw data would have to be cleaned, to take account of missing values and inconsistencies in the way the data were recorded.

4.1.2. Methods

The relevant physiological and pharmacological data which were required for quantitative scoring were extracted from the CareVue system for 3 patients. The raw data were examined and rules created to deal with missing values, or inconsistencies in the manner in which a parameter was recorded. The data were pre-processed in this way to be ready for analysis in the score.

4.1.3. Results

The data sets were analysed by the scoring system pre- and post-extrapolation, and a score of cardiovascular stability was displayed, either as a raw number or in graphical form over time.

4.1.4. Conclusion

With appropriate pre-processing of data, the quantitative score could give a read-out of cardiovascular stability over time from real patients in Intensive Care.

4.2. Introduction

The quantitative score described in the previous chapter now needed to be tested against data obtained from patients in Intensive Care, to ascertain if it gave clinically credible results. That is to say, the score increased as the patient deteriorated and decreased as the patient improved, i.e. it had a high degree of predictive validity. This task would have been extremely difficult to achieve if the scores had had to be calculated manually, and would have been prone to error. The interrogation and extraction of data from CareVue is already described. I now describe how any inconsistencies in the data or missing data were overcome. This was an important task prior to testing the quantitative score on real patient data. The testing was semi-automated thanks to our group's collaboration with Prof. Derek Sleeman of the Computing Science Department at Aberdeen University.

4.3. Methods

4.3.1. Data Collection

An approach was made to the Scotland A Research Ethics Committee rather than a local ethics committee as the aim was to analyse data from patients who are in Intensive Care, many of whom will have had an adults with incapacity form completed. Their view was that if the analysis was only of routinely collected physiological, pharmacological and biochemical data, this work did not require a formal ethics application as what was proposed was within the scope of service development and audit. It was also stated that the data would be downloaded, anonymised by the CareVue administrators and stored on NHS servers. There would be no patient identifiers but each patient data set would be given a unique identifying number. Any data would be sent from an NHS server to colleagues at Aberdeen University. Only the CareVue administrators could (theoretically) de-anonymise the data sets. Although not asked to by the ethics committee, we placed notices in the foyers of the Intensive Care unit that data was collected for routine analysis.

The quantitative score has two components, a score for parameters collected at regular intervals and a score for parameters collected intermittently. To make interpretation meaningful and compare the score over time, the same parameters need to be recorded. Therefore the CareVue administrators were asked to extract at least the core parameters recorded at regular intervals. These are mean arterial pressure, heart rate, oxygen saturation, urine output and temperature. The regularly recorded parameters involved in the weightings were similarly extracted, namely quantities of adrenaline, noradrenaline, fluid, propofol, alfentanil and the amount of PEEP. Other parameters extracted were some of

those intermittently recorded, and included cardiac output and cardiac index. In the rest of this chapter, when referring to the testing of the quantitative score, it is only on the component using regularly recorded parameters.

The criteria for selecting the patients whose data was to be downloaded by the CareVue administrators was that they had received haemodialysis, as this was the initial question which led to the need for developing a new score which took into account the amount of physiological and pharmacological support was receiving. (Although no medical information about the patients was available the fact they were receiving haemodialysis could be deduced by looking for "blood pump speed" in the query. In the testing of the quantitative score, it was run against datasets of patients who had received haemodialysis

As the patients had dialysis-dependent renal failure they were likely to be more ill, possibly more cardiovascularly unstable, and have greater amounts of cardiovascular monitoring e.g. Lithium Dilution Cardiac Output (LiDCO) Monitoring. The LiDCO is the Royal Infirmary Intensive Care Unit's cardiac output monitoring device of choice which uses lithium chloride dilution and the Stewart-Hamilton principle[113].

In the testing of the quantitative score, 3 data sets numbered 708, 728, and 733 were used. At the beginning of this work the number of data sets available for analysis was small and these particular data sets were amongst the first to be downloaded by the CareVue administrators.

Description of the three patient data sets.

In the following description of the 3 data sets used, the first and last "days" do not mean that the patient is in Intensive Care for the whole 24 hours.

Patient 708:
This patient was in Intensive Care for 3 days (19.45 on day one to 09.00 on day 3).
40 time points worth of data were recorded. These were all hourly except for the first one (19.45) and in day 2 where an extra time point is recorded at 13.37.
The patient had one session of haemodialysis on day 2 between 18.00 and 20.00

Patient 728:

This patient was in Intensive Care for 12 days (17.00 on day 1 to 14.00 on day 12).
278 time points were recorded. These were all hourly except for extra time points as follows: Day 1 at 17.51/20.07, Day 2 at 09.30/19.22, Day 3 at 10.12, Day 4 at 05.05/12.08, Day 5 at 08.48/09.16, Day 6 at 11.24, Day 7 at 09.27/11.32, Day 8 at 10.15/11.41, Day 9 at 10.56/12.26, Day 10 at 09.39, Day 11 at 14.59.
The patient had the following dialysis sessions: Day 2 (19.21-21.00), Day 3 (13.00-18.00), Day 4 (10.00-16.00), Day 5 (09.00-14.00), Day 6 (11.24-16.00), Day 7 (12.00-20.00), Day 8 (16.00-02.00) and Day 10 (12.00-14.00).

Patient 733:

This patient was in Intensive Care for 16 days (07.00 on day 1 to 17.00 on day 12).
395 time points were recorded. These were all hourly except for extra time points as follows: Day 1 at 07.42/10.23, Day 2 at 10.35/13.31/16.01, Day 3 at 11.09, day 4 at 05.20/12.54, Day 5 at 0844/1156/1256, Day 6 at 1003/2333, Day 7 at 11.09/12.32/16.29/18.04, Day 8 at 1509, Day 9 at 10.30, Day 10 at 15.33, Day 12 at 11.17, Day 13 at 10.12, Day 14 at 06.30/06.50/06.55/11.24, Day 15 at 1152 and Day 16 at 12.57.
The patient had the following dialysis sessions: Day 2 (16.01-18.00), Day 3 (11.00-16.00), Day 4 (05.20-10.00), Day 5 (18.00-00.00), Day 6 (17.00-19.00) and (22.00-02.00), Day 7 (12.32-19.00), Day 8 (06.00-15.00), Day 9 (13.00-17.00), Day 10 (11.00-17.00), Day 11 (12.00-18.00), Day 12 (12.00-18.00), Day 13 (06.00-12.00), Day 14 (06.30-13.00), Day 15 (06.00-12.00) and Day 16 (11.00-16.00).

The complete data set for patient 708 is recorded in *Appendix I* (the complete data sets for patients 728 and 733 are not included due to their size). Only parameters which are recorded at regular intervals, for the purpose of analysis with the quantitative scoring system, are shown.

4.3.2. Presentation of data

An example of the ”raw data” as it is produced by the CareVue administrators is shown below. The process of extracting the data up until this point is described in detail in Chapter 2. The data extracted at this point is already in a basic form within an excel spreadsheet of which an example is shown in table 4-1.

Table 4-1: Example of the presentation of the raw data

Date & Time	Infusion	Dose units
08/12/2006 18:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 18:00	Noradrenaline 24mg/240ml(mg/hr),	4.8 mg/hr
08/12/2006 20:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 20:00	Noradrenaline 24mg/240ml(mg/hr),	4.8 mg/hr
08/12/2006 21:00	Noradrenaline 24mg/240ml(mg/hr),	3 mg/hr
08/12/2006 21:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 21:43	Alfentanil 25000mcg(mg/hr), 2500	1 mg/hr
08/12/2006 22:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr
08/12/2006 22:00	Noradrenaline 24mg/240ml(mg/hr),	3.5 mg/hr
08/12/2006 22:00	Alfentanil 25000mcg(mg/hr), 2500	1 mg/hr
08/12/2006 23:00	Noradrenaline 24mg/240ml(mg/hr),	3.5 mg/hr
08/12/2006 23:00	Propofol 2%(mg/hr), 1000 mg/50 m	100 mg/hr

The data in its raw state comprises lines of information in an Excel spreadsheet which are not ordered in a manner by which they could be interpreted. There was an appreciation that when larger data sets required to be analysed that it would be important to be able to view and annotate the data in an intelligible way and also to be able to apply rules to extrapolate where data were parameters were missing at different time points. The Royal Infirmary ICU has 20 beds, and if we take an average of 10 parameters collected per hour for a patient at each bed space, which is open 365 days per year, then the number of pieces of data collected each year is 20 beds x 10 parameters x 24 hours x 365 days which equals 1.75 million pieces of data. This is theoretical, hypothetical and probably an underestimation. The CareVue system is capable of collecting many times more data than this. It illustrates that manual pre-processing and analysis would be an impossible task on that scale.

Our computing science colleagues devised a web based tool ACHE (an Architecture for Clinical Hypothesis Examination) [88] to deal with these specific problems. It has two main components. ACHE “annotate” which allows the raw data to be viewed in an ordered fashion within the excel spread sheets and ACHE “pre-process” allows for any extrapolation rules to be applied to the data at this point. This was essential, as with the

number of data points in some of the spreadsheets manual pre-processing of this time series data would have been impossible.

An example of formatted data as it would appear in ACHE is given in table 4-2.

Table 4-2: Example of formatted data (as it would be presented in ACHE)

Time	Noradrenaline	Fluids	Propofol	Alfentanil	HR	SpO2	FiO2	Urine	Temp	MAP
Day 1										
05/01/2007 17:00:00		1500.0	100.0	1.0	96.0			100.0		
05/01/2007 17:51:01					96.0	99.0			37.1	50.0
05/01/2007 18:00:00	0.5	2000.0	100.0	1.0	95.0	99.0	100.0	0.0	37.1	50.0
05/01/2007 19:00:00	1.0	500.0	100.0	1.0	106.0	100.0	80.0	0.0		57.0
05/01/2007 20:00:00	1.2		40.0	1.0	102.0	100.0	70.0		37.0	54.0
05/01/2007 20:07:38					106.0	100.0	80.0			57.0
05/01/2007 21:00:00	1.2		40.0	1.0	100.0	97.0				55.0
05/01/2007 22:00:00	1.4		40.0	1.0	98.0	96.0	70.0	5.0		58.0
05/01/2007 23:00:00	1.4	500.0	40.0	1.0	98.0	95.0	70.0		36.6	62.0

4.3.3. Types of entered data errors

As can be seen from the extract from ACHE above, although the data are more organised for semi-automated analysis, there are missing figures which could lead to meaningless values being generated by the quantitative score. Although the data can be auto-charted into the CareVue system and then verified, current practice at the Royal Infirmary ICU is that data are entered manually, at hourly time points, by the nursing staff. On further analysis three sources of error were identified. These are missing data points, a data point entered incorrectly and inconsistency in the manner in which the data point is recorded (e.g. mls/h or mg of a drug).

4.3.4. Mechanism for handling different types of missing data

The outcome of a discussion with the computing scientists was that missing values were to be dealt with in a pragmatic clinical manner rather than using a complex calculation. The quantitative score has two components, a score calculated using parameters recorded regularly and a score calculated using parameters recorded intermittently. In the data sets analysed, the quantitative score was calculated using the parameters recorded regularly. An analysis of the data sets used in this chapter showed that for the parameters recorded regularly it was unusual for nursing staff (as you might expect) to have omitted recording one of these key parameters for more than one hour (with the exception of urine output if only being averaged every few hours). The discussion therefore focussed on dealing with a single missing value.

There was more than one option. The most pragmatic (and simplest do achieve) was to use the previously recorded value. There were two exceptions. For urine output, nursing staff sometimes record a much larger volume every few hours, rather than measure every hour. In the absence of concurrent medical information (to perhaps show what the nursing staff were doing) the decision was taken that if there was a missing value and the previous value was less than 100mls then we would use the previous value. If there was a missing value and a previous value of 100mls or over the algorithm looked forward to the next available time point and calculated an average. For fluids it was assumed that one (or more) missing value was a conscious decision as fluids can be started and stopped at points throughout the day, and so took no action. A consideration was made to taking an average of the value (or values) before and after a single missing time point, but on reflection it was thought the other approach simpler (with less complicated algorithms). A summary of the actions taken is shown in the table 4-3.

Table 4-3: Mechanism for handling missing values

Parameter	**Action for a Single Missing Value**
Adrenaline	Use previously recorded value
Noradrenaline	Use previously recorded value
Propofol	Use previously recorded value
Alfentanil	Use previously recorded value
Heart Rate	Use previously recorded value
SpO2	Use previously recorded value
FiO2	Use previously recorded value
Urine	If previous value less than 100 replace missing value with previous value. Otherwise look forward in the data to the next value, take the average of it over the missing values and then replace missing values with that average.
Temperature	Use previously recorded value
MAP	Use previously recorded value
Fluids	Do nothing

No consideration was made for multiple missing values for parameters recorded at regular intervals (other than urine and a conscious decision to take “missing fluids” at face value). The following analysis shows the results of this strategy in the 3 patients analysed (using the part of the quantitative score with parameters recorded at regular intervals).
In the tables below each column represents parameters recorded at regular intervals with the number of times the algorithm for dealing with single missing values does not work when there is more than one time point missing for a parameter. This figure is shown in the

bottom row and the day (D) within the ICU stay and the relevant missing time points shown above. (Note fluids and urine output are not included). A gap in the column showing the time points separates different periods of multiple missing data. An analysis is given for each patient (tables 4-4 to 4-6). This forms the basis of a more informed approach as to how we might deal with these in the future (discussed below).

Table 4-4: Analysis of rules for handling missing values on patient 708

PATIENT 708 (40 time points)

Mean arterial Pressure	Heart Rate	Adren.	Noradr	Propofol	Alfentanil	FIO2	SpO2	Temp.
nil	nil	N/A	nil	nil	nil	nil	nil	D1 2300 D2 0000 D2 1300 D2 1337 D2 1900 D2 2000 D2 2200 D2 2300 D3 0100 D3 0200 D3 0300
0	**0**	**N/A**	**0**	**0**	**0**	**0**	**0**	**5**

In this patient the algorithms would have dealt with all missing parameters except for temperature. However, all of the missing temperature data falls within scoring no points or one point, this only making a very slight inaccuracy with the final score out of 48 for the parameters recorded at regular intervals.

Table 4-5: Analysis of rules for handling missing values on patient 728

PATIENT 728 (278 time points)

MAP	HR	Adren.	Norad.	Propofol	Alfent.	FIO2	SpO2	Temp.	Temp. (Cont)	Temp. (Cont)
nil	nil	N/A	nil	nil	D8 1200 D8 1226 D8 1300 D8 1400 D8 1500 D8 1600 D8 1700 D8 1800 D8 1900	D11 1800 D11 1900 D12 0300 D12 0400	D12 0300 D12 0400	D1 2007 D12100 D1 2200 D2 0400 D2 0500 D2 1300 D2 1400 D2 1800 D2 1900 D2 2100 D2 2200 D3 0000 D3 0100 D3 0300 D3 0400 D3 0500 D3 1012 D3 1100 D4 0300 D4 0400 D4 0505 D4 0600 D4 1200 D4 1208 D4 2100 D4 2200 D5 0600 D5 0700 D5 0916 D5 1000 D5 1800 D5 1900 D5 2100 D5 2200 D5 2300 D6 0000	D6 0000 D6 0100 D6 1100 D6 1156 D6 1500 D6 1600 D6 1700 D6 2100 D6 2200 D7 0927 D7 1000 D7 1100 D7 1132 D7 1300 D7 1400 D7 1800 D7 1900 D8 0600 D8 0700 D8 0800 D8 1000 D8 1015 D8 2200 D8 2300 D9 1050 D9 1100 D9 1226 D9 1300 D10 0700 D10 0800 D10 0900 D10 0939 D10 1100 D10 1200 D10 1300	D10 1500 D10 1600 D10 1800 D10 1900 D10 2100 D10 2200 D10 2300 D11 0000 D11 0300 D11 0400 D11 0500 D11 0700 D11 0800 D11 1200 D11 1300 D11 1459 D11 1500 D11 1600 D11 2300 D12 0000 D12 0100 D12 0200 D12 0300 D12 0400 D12 0500 D12 0600 D12 0800 D12 0900 D12 1100 D12 1200 D12 1300
0	**0**	**N/A**	**0**	**0**	**1**	**2**	**1**	**40**		

In this patient, the algorithms would have dealt with missing parameters for mean arterial pressure, heart rate, noradrenaline and propofol. An interesting issue is identified with missing data for alfentanil. The actual rate prior to the missing data was 0.5 mg/h. Without access to the patient's history, this probably represents a patient who no longer requires it. The infusion is started again at the very low rate of 0.5mg/h several hours later. It may be inappropriate to extrapolate under these circumstances. One solution would be an algorithm which detects a tapering dose to very low threshold and does not extrapolate on the assumption that this is a considered clinical decision. There were two episodes of two hours of missing FiO2 data. These were in the context of a high oxygen saturations and low FiO2 before and after which would have accrued no score in this case. Again it is

difficult to say if the patient who was near to discharge was not on any oxygen. There is a single episode of missing oxygen saturation, again near the end of the patient's stay. This is on day 12 between 0300 and 0400, the same time period as the second episode of missing FiO2. This raises the question of whether in the future (when rules which will have to be put in place to deal with more than one missing data point), if the score should be calculated if there are multiple extrapolated parameters. Finally, it is obvious from the table that there are multiple missing temperature time points. In only a single instance in the actual data was the temperature out with 36-38 ^{o}C (which scores zero points).

Table 4-6: Analysis of rules for handling missing values on patient 733

PATIENT 733 (395 time points)

Mean arterial Pressure	**Heart Rate**	**Adren.**	**Norad.**	**Propofol**	**Alfent.**	**FIO2**	**SpO2**	**Temp.**	**Temp. Cont.**	**Temp Cont.**
D14 0650 D14 0655	D14 0630 D14 0650 D14 0655	N/A	nil	D14 0650 D14 0655	D14 0630 D14 0650 D14 0655	D1 0742 D1 0800	D3 1400 D3 1500	D1 0900 D1 1000 D1 1023 D4 1000 D4 1100 D4 1254 D4 1300 D4 1500 D4 1600 D4 1800 D4 1900 D4 2000 D4 2100 D4 2200 D5 0000 D5 0100 D5 0600 D5 0700 D5 0800 D5 0844 D5 1156 D5 1200 D5 1256 D5 1300 D5 2000 D5 2100 D5 2200 D5 2300 D6 1000 D6 1003 D6 1100 D6 1300 D6 1400 D6 1600 D6 1700 D6 1900 D6 2000	D6 2000 D6 2300 D7 0100 D7 0200 D7 0400 D7 0500 D7 1109 D7 1200 D7 1600 D7 1629 D7 1700 D7 1804 D7 1900 D 7 2000 D7 2200 D7 2300 D8 0600 D8 0700 D8 2200 D8 2300 D9 0200 D9 0300 D9 0400 D9 0600 D9 0700 D9 0800 D9 2100 D9 2200 D11 0200 D11 0300 D11 0400 D11 0500 D11 1800 D11 1900 D11 2100 D11 2200	D12 0000 D12 0100 D12 0200 D12 0400 D12 0500 D12 2100 D12 2200 D12 2300 D13 0100 D13 0200 D13 0300 D14 0400 D13 0600 D13 0700 D14 0600 D14 0630 D14 0650 D14 0655 D14 1124 D 141200 D14 1400 D14 1500 D14 1600 D14 1700 D15 1000 D15 1152 D15 1200 D15 1300 D15 1900 D15 2000 D16 0800 D16 0900 D16 1200 D16 1257
1	**1**	**N/A**	**0**	**1**	**1**	**1**	**1**	**41**		

There is one episode of several missing data points for mean arterial pressure, heart rate, propofol and alfentanil. Interestingly they all occur when extra time points are recorded in between the routine hourly time points. Again this raises the question of what to do when there are multiple missing parameters potentially being extrapolated. The single episode of missing FiO2 data occurs just after the patient's admission, again where there is an extra time point between two hourly time points. This raises the question of extrapolation with little information before to compare with. Again the dominant source of missing data is with temperature measurement. As with the previous example all of the temperature time

points within which the missing examples occur would score 0 points (with the exception of 2 instances).

These analysis shows that apart from temperature (where almost all values fell in a range that would not alter the calculated score) there were very few examples of missing data with more than a single time point for all the other parameters.

Nevertheless, to take this work forward these must be addressed. This analysis identifies some associations for more than one missing time point and identifies some further challenges for consideration of how to handle the data where there is more than one missing time point.

Possible associations:

- An event has occurred and the nurse at the bed space has recorded an additional time point with only a few parameters between the routine recorded hourly time points. A drug (e.g. noradrenaline, propofol) has been at a low level as the patient's condition improves but then has to be started i.e. there is no "missing" data.
- The patient has just been admitted and is very unstable and the focus of clinical care is on admitting and stabilising rather than recording data.
- The patient is about to be discharged and some "missing" parameters have in fact been stopped.
- Temperature appears to be a low priority parameter to record and is often omitted by nursing staff.

Additional challenges:

- What should occur where there is sparse data for a parameter before or after a period of missing data?
- If at an hourly time point, how many extrapolated parameters is it reasonable to calculate the score with?
- If there is a period of stability before and a period of instability after a series of missing values, how can you determine where the true value is likely to lie?

It is likely that a series of hierarchical algorithms will have to be developed to account for these different situations. The first of these to be applied could be one which takes a mean of 3 time points before and after the missing values (and substitutes this value within the missing time points.) If this is not possible then an algorithm should examine the value

immediately before and after the missing time points. Missing data closest to the "before value" should be substituted with that value and ditto for missing values closest to the "after value." If there is an odd number then arbitrarily one extra missing value to align itself with the nearest "after value". Alternatively (but much more complicated to achieve) the rest of the parameters could be analysed to ascertain in which direction they are changing to determine whether the values which are missing should more closely align with those before or after. Further, if a parameter has been slowly reducing, low level thresholds will have to be determined whereby the score is not extrapolated i.e. there is no "missing" data because the parameter in questions has been appropriately stopped. If the score is developed into a clinical tool, it would be important for the score calculated at a particular time point to be qualified by the number of extrapolated parameters. For example, a score of 18 with no extrapolated parameters might be displayed 18(0) and with 3 extrapolated time points 18 (3) and so on.

4.4. Results

4.4.1 Application of the quantitative score to the unextrapolated data sets

The quantitative score was applied to the three data sets (patients 708, 728 and 733). Initially this was to the data before the extrapolation rules were applied. An extract of patient 708 is shown in table 4-7. The complete data set for patient 708, with the scores for each time point, is recorded in *Appendix I* (the complete data sets for patients 728 and 733 are not included due to their size).

Table 4-7: Testing of the score on unextrapolated data (extract patient 708)

Time	Adrenaline	Noradrenaline	Fluids	Propofol	Alfentanil	HR	SpO2	FiO2	Urine	Temp	MAP	Dialysis	Score
19/12/2006 19:45:37						114.0	90.0						**2**
19/12/2006 20:00:00			500.0			111.0	92.0	100.0		37.7	62.0		**11**
19/12/2006 21:00:00			500.0			116.0	79.0	100.0			68.0		**12**
19/12/2006 22:00:00	1.0	2.0	500.0	60.0	1.5	99.0	69.0	100.0	80.0	38.1	62.0		**19**
19/12/2006 23:00:00	2.0	2.0	500.0	60.0	1.5	108.0	83.0	100.0	10.0		62.0		**20**
20/12/2006 00:00:00	2.8	4.0	250.0	60.0	1.5	110.0	100.0	100.0	15.0		59.0		**19**
20/12/2006 01:00:00	2.8	4.0	350.0	60.0	1.5	112.0	83.0	100.0	10.0	38.3	59.0		**23**
20/12/2006 02:00:00	2.8	4.0	100.0	60.0	1.5	109.0	83.0	100.0	25.0		63.0		**18**
20/12/2006 03:00:00	2.8	4.0	100.0	60.0	1.0	107.0	75.0	100.0	15.0	38.8	67.0		**20**
20/12/2006 04:00:00	2.8	4.0	100.0	60.0	1.0	112.0	76.0	100.0	0.0		80.0		**15**
20/12/2006 05:00:00	2.8	4.0	50.0	60.0	1.0	118.0	80.0	100.0		38.9	82.0		**13**
20/12/2006 06:00:00	2.8	4.0	100.0	60.0	1.0	121.0	82.0	100.0	35.0		83.0		**13**
20/12/2006 07:00:00	2.8	4.0	100.0	60.0	1.0	124.0	82.0	100.0	15.0	38.9	64.0		**20**
20/12/2006 08:00:00	2.8	4.0	50.0	60.0	1.0	126.0	85.0	100.0	10.0	39.8	66.0		**20**

At this point, on further analysis, there was a realisation that there was still an inconsistency in the way in which inspired oxygen concentration was being recorded. For example an inspired oxygen concentration of 0.8 was being recorded as 0.8, .8, 0.80, 8 (assumed to be a transcription error) and 80% (not a fraction). It was decided to convert all forms of fractions to percentages. This is, of course, merely a change in nomenclature, does not affect the values awarded in the score, and is reflected (in bold) in table 4-8.

Table 4-8: Extract from quantitative score showing changes in nomenclature for inspired oxygen concentration

		Oxygen Saturation (Sp02) (%)						
		<75	75-89	90-94	95-100	-	-	-
Unweighted Score	Air	3	2	1	0	-	-	-
Inspired Oxygen Fraction (FiO2)	**22-49**	4	3	2	1	-	-	-
	50-79	5	4	3	2	-	-	-
	≥80	6	5	4	3	-	-	-
PEEP / CPAP (cmH_20)	0-5	0	0	0	0	-	-	-
	6-8	4	3	2	1	-	-	-
	9-11	5	4	3	2	-	-	-
	≥12	6	5	4	3	-	-	-

4.4.2 .Application of the score to extrapolated data sets

The quantitative score, with extrapolation rules, was now applied to the data, with inspired oxygen concentration given consistently as a percentage. The quantitative score applied to extrapolated data set 708 is shown in *Appendix II.*

4.4.3. Comparison of the application of the score to unextrapolated and extrapolated data

To ascertain the effect of the extrapolation of data, the quantitative scores calculated pre- and post-implementation of the rules for handling missing data points, were compared. As patient 708's data set only contains 40 data points, it is shown in its entirety in table 4-9.

Table 4-9: Calculated score from patient 708 pre- and post-extrapolation of data

Differences are shown in bold. As this data set only contains 40 parameters it is shown in its entirety.

Time	Score pre extrapolation	Score post extrapolation
Day 1		
19/12/2006 19:45:37	2	2
19/12/2006 20:00:00	11	11
19/12/2006 21:00:00	12	12
19/12/2006 22:00:00	19	19
19/12/2006 23:00:00	20	20
Day 2		
20/12/2006 00:00:00	19	19
20/12/2006 01:00:00	23	23
20/12/2006 02:00:00	18	**19**
20/12/2006 03:00:00	20	20
20/12/2006 04:00:00	15	**16**
20/12/2006 05:00:00	13	**16**
20/12/2006 06:00:00	13	**14**
20/12/2006 07:00:00	20	20
20/12/2006 08:00:00	20	20
20/12/2006 09:00:00	17	**20**
20/12/2006 10:00:00	20	20
20/12/2006 11:00:00	16	**19**
20/12/2006 12:00:00	18	18
20/12/2006 13:00:00	20	20
20/12/2006 13:37:00	9	**20**
20/12/2006 14:00:00	18	**21**
20/12/2006 15:00:00	19	**20**
20/12/2006 16:00:00	20	20
20/12/2006 17:00:00	17	**21**
20/12/2006 18:00:00	18	18
20/12/2006 19:00:00	16	**17**
20/12/2006 20:00:00	4	**17**
20/12/2006 21:00:00	20	20
20/12/2006 22:00:00	19	19
20/12/2006 23:00:00	17	**20**
Day 3		
21/12/2006 00:00:00	21	21
21/12/2006 01:00:00	20	20
21/12/2006 02:00:00	19	**22**
21/12/2006 03:00:00	19	**22**
21/12/2006 04:00:00	20	**23**
21/12/2006 05:00:00	23	**24**
21/12/2006 06:00:00	18	**25**
21/12/2006 07:00:00	23	**25**
21/12/2006 08:00:00	22	**25**
21/12/2006 09:00:00	11	**14**

As can be seen, extrapolation of the rules to handle missing single data points makes a difference to the score at several of the time points. These are perhaps better illustrated in graphical form. Figure 4-1 shows the results for patient 708 and figure 4-2 for patient 728 (which has 278 time points). The extrapolated scores have fewer peaks and troughs than the original scores, seen better in patient 708, as fewer time points are being displayed in a single graph.

Figure 4-1: Quantitative score over time for patient 708

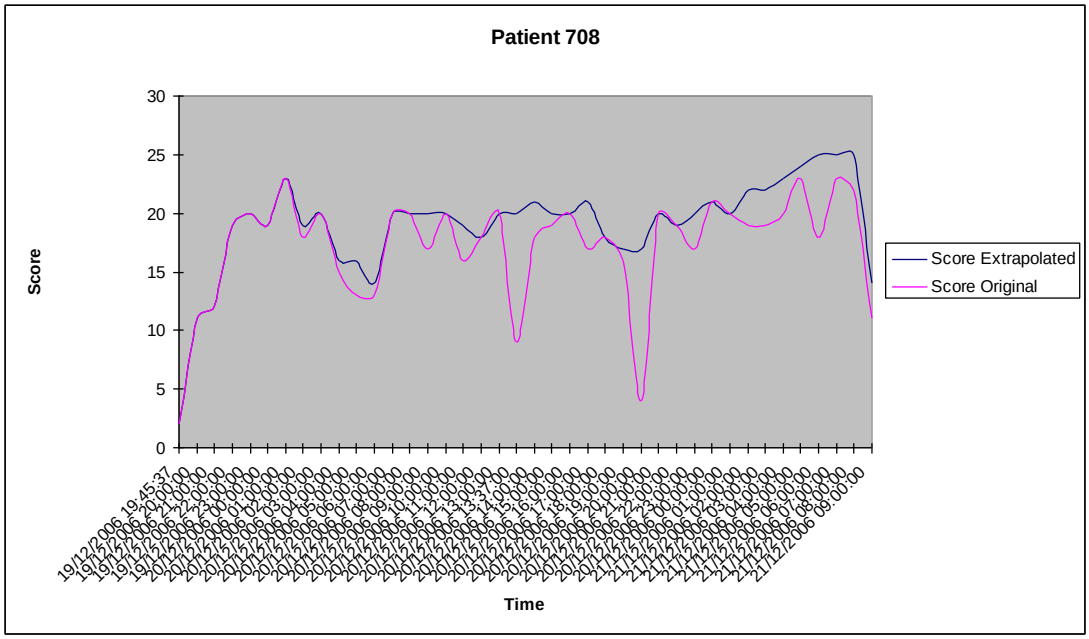
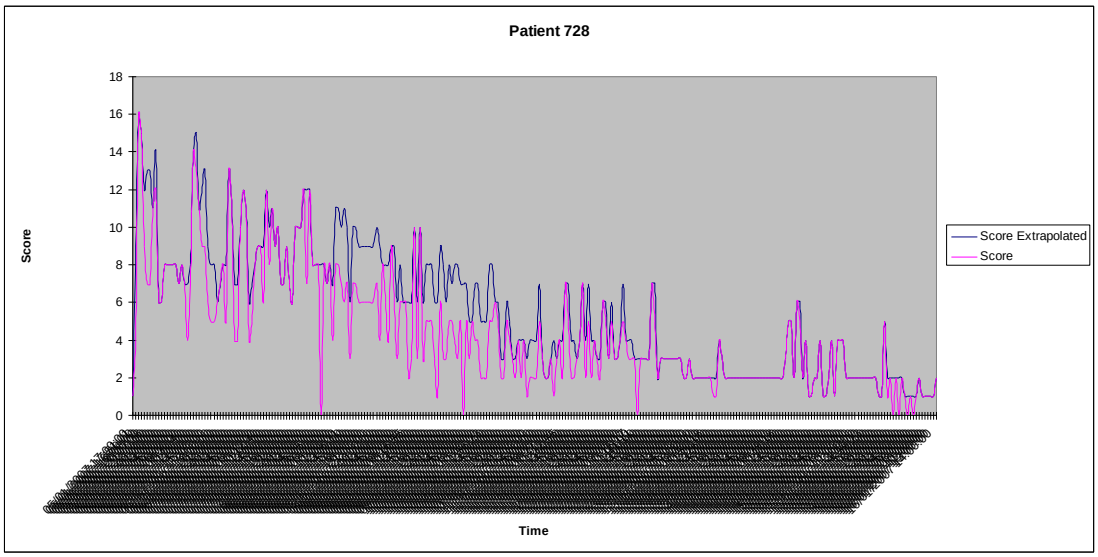


Figure 4-2: Quantitative score over time for patient 728



4.4.4. Hameodialysis events as represented by the score

As the initial idea for a score of instability arose from an interest in cardiovascular instability during haemodialysis, the following analysis shows how the score changed during renal replacement therapy. Arrows in the graphs below show where haemodialysis occurred. Note that in patient 728 and 733 the first arrow is smaller than the others. This is because at Glasgow Royal Infirmary the first session of haemodialysis is characteristically 2 hours in duration.

The qualitative score over time is shown for the 3 patient's data sets described earlier (708, 728 and 733). Figure 4-3 shows the complete Intensive Care stay for patient 708, figure 4-4 approximately the first third of patient 728 and figure 4-5 the first fifth of patient 733. Patients 728 and 733 contain too many data points to compress into the one graph.

Figure 4-3: Haemodialysis events as represented by the quantitative score of patient 708

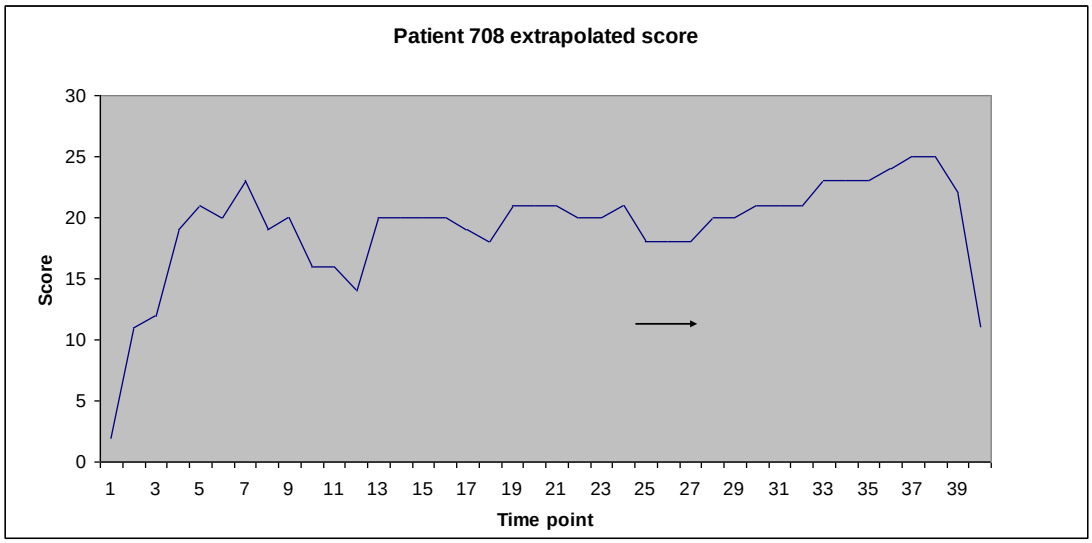


Figure 4-4: Haemodialysis events as represented by the quantitative score on patient 728

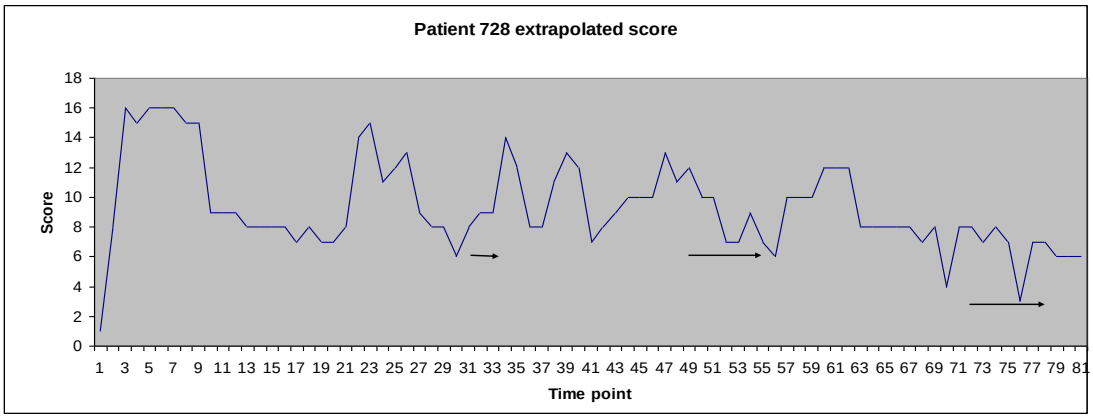
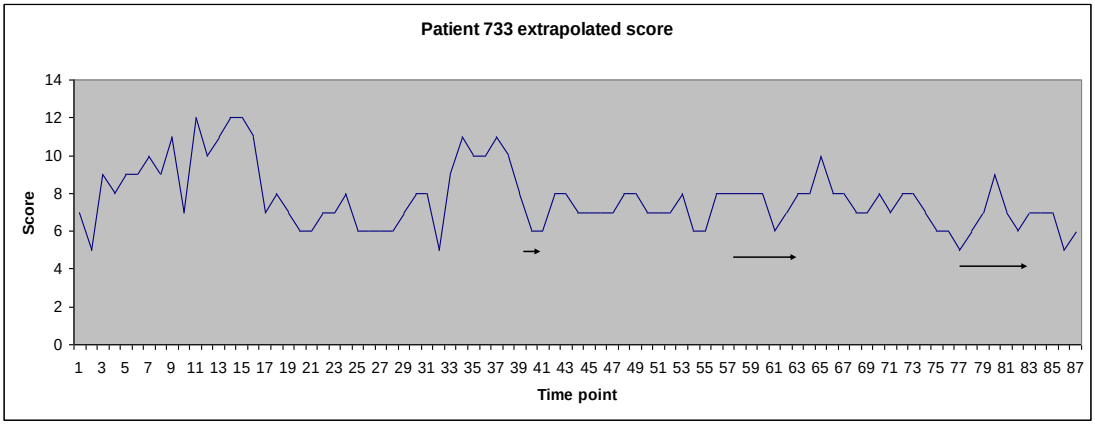


Figure 4-5: Haemodialysis events as represented by the quantitative score on patient 733



4.5. Discussion

For the first time there now exists a score which takes into account the level of physiological and pharmacological support which the patient is receiving. The score could now be applied to real patient data sets and give an indication of overall cardiovascular improvement or deterioration. It could theoretically be applied in real time at the patient's bedside.

Using anonymised data as per the agreement with the ethics committee resulted in certain limitations. There was no access to any clinical information for the patient's whose data sets were being analysed. However, certain clinical events could be deduced e.g. blood pump speed recorded means that the patient was undergoing a period of renal replacement therapy. Being able to place lines of data in a clinical context would have been useful in certain situations. For example, in Chapter 6 a two stage discrimination experiment is described as part early work to validate the qualitative Intensive Care Unit - Patient Scoring System developed in Chapter 5. In this, consultants were shown lines of data and asked whether they thought the patient was improving or deteriorating. It would have been useful (but not essential) to have been able to place this data within a clinical context. By definition there was reliance upon the CareVue administrators, who had other onerous commitments to download the data for us. Although they were extremely helpful in facilitating the studies, this could at times be a rate-limiting step.

Work on the parameters recorded intermittently was not taken further at this stage because of the various problems identified in the last chapter e.g. the score changing because the parameter is or is not being recorded rather than the patient improving or deteriorating. These additional parameters could only be of potential interest if they are recorded at regular intervals throughout the patient stay. For example, in a highly unstable patient systemic vascular resistance (SVR) might be measured if they have cardiac output monitoring in place. In this case the score would be out of 51, not 48 (as a maximum of 3 points are given for changes in SVR.

Algorithms were applied to deal with a single missing value (with the exception of fluid inputs and urine output). In this chapter it was applied to 3 patient data sets at every time point accepting that on occasion the score would not be accurate because of more than one missing data point for a parameter. However, there was an assumption that nursing staff were unlikely to miss recording essential parameters for more than one time point in a row.

An analysis of the data sets shows that with the exception of temperature this assumption is correct.

However, to take the quantitative score further, for it to be a potentially useful bed side tool, possible methods for dealing with multiple missing time points have been suggested. The alternative is for there to be no score recorded when there is missing data. Clearly this would be a major drawback were the score to be used as a clinical tool.

The algorithms suggested could be tested by taking a data set and deliberately removing blocks of time points for different parameters. This could be done at different periods during the patient's stay. The different types of algorithm could be applied and compared to the score calculated on the complete data to see which gives the most accurate reflection of the true score. This will form part of future work.

The quantitative score does appear to have some merit. From the graphs above it shows improvements and deteriorations over time in the overall physiological state of the patient. However, to make the score more meaningful it would require further refinement to overcome its disadvantages. Although some of the ranges are based upon those in previous scores and common physiological facts, many are based on the experience of a single clinician. It could be refined by including a greater number of clinicians using some of the techniques described in chapter 3, e.g. a Delphi process or brain storming sessions to achieve a consensus about the parameters for inclusion and the ranges. As discussed, in the absence of a reference standard, the weighting of the score would be difficult. We could test it in a number of clinical scenarios to establish if the score attributed to various parameter ranges is credible to other experts. The relative weightings of parameters could then be altered to give the best fit. Therefore, although the score could have been improved, in the absence of a reference standard it would ultimately be justified by testing against expert opinion.

Reflection on this problem led to the idea of examining it from a different angle. Rather than define ranges of parameters to give an arbitrary score between -3 and + 48 and then justify if clinically credible, an alternative approach was to capture and base the actual score on clinical expertise in the first place. This was the premise behind the Intensive Care Unit Patient Scoring System described in Chapter 5. It was hoped that it might better capture clinical interpretation of a situation than the quantitative score. For example, a

patient who is cardiovascularly stable in every respect except for requiring 100% would score moderately highly in the quantitative score, but in reality is clinically extremely ill.

4.6. Conclusion

Having overcome difficulties with missing or inconsistent data, the quantitative score of cardiovascular stability could be successfully applied to real patient data sets and quantify stability over time. Although this approach had merit it was judged that a more clinically credible, and alternative approach was to capture and base the actual score on clinical expertise in the first place. This was the premise of the Intensive Care Unit – Patient Scoring System described in Chapter 5.

4.7. Acknowledgements

I decided on the nomenclature and units used for recording the chosen parameters. The protocol for handling missing values was decided as a result of a clinical discussion between myself and Prof. Kinsella. The scoring system being tested was designed entirely by myself.

The 3 data sets analysed were extracted from CareVue by the CareVue administrators. The data were pre-processed and my quantitative score was run against the cleaned data by our Computing Science Collaborators, to produce the graphical representations of quantitative score over time.

Chapter 5: Development of the Intensive Care Unit Patient Scoring System (ICU-PSS)

Abstract

5.1.1. Background

The quantitative score developed could produce a measure of cardiovascular stability over time. However, it produced a number on a scale which was difficult to relate to clinically, and which relied upon one clinician's experience. The aim was now to develop a qualitative 5 point scale underpinned by a sophisticated physiological rule base.

5.1.2. Methods

In the initial phase, two clinicians annotated real patient datasets and marked broad time periods on a 5 point scale of stability. In the next more detailed phase, one clinician annotated 10 data sets with 2761 predominantly hourly time points, and simultaneously described physiological rules to justify his annotations. These rules were used by a computer programme to annotate unseen datasets. The unseen datasets were also annotated by the clinician, and the computer prediction based on his rules, and his actual annotations, compared in a confusion matrix. Points of disagreement were analysed and the rule base refined.

5.1.3. Results

As a result of comparing his annotations with the computer prediction the clinician was able to produce a rule base which captured his clinical expertise. This process was repeated with two other senior Intensive Care Clinicians to produce a sophisticated set of physiological rules (The ICU-PSS), which underpinned the 5 point qualitative scale.

5.1.4. Conclusions

Through a process of gradual refinement, a complex series of physiological rules was developed which captured the clinical expertise of 3 senior Intensive Care Clinicians. This formed the Intensive Care Patient Scoring System (The ICU-PSS)

5.2. Introduction

I had devised a quantitative scoring system for cardiovascular instability in Intensive Care Patients which took into account the level of support the patient required. However, it produced a number between -3 and +48 (for the parameters recorded at regular intervals). As with many clinical measurements, a numerical score has relatively little usefulness at the bedside. An obvious comparison could be made with visual analogue scores for pain, which are of less use than simple verbal rating scores. Rather than define ranges of parameters to give an arbitrary score between -3 and + 48 and then justify if clinically credible, an alternative approach was to capture and base the actual score on clinical expertise in the first place. It was felt that it would be more clinically relevant if the patient state could be summarised in a 5 level qualitative score. This chapter describes the design and refinement of such a 5 level qualitative score, with each level based on detailed and clinically reasoned quantitative rules.

In 1946 De Groot described what is considered to be the first attempt to capture the performance of an expert [114]. He was a skilled chess player and determined that the ability to play chess is "best captured in the task of selecting the next move for a given chess position taken from the middle of the game between two chess masters." He postulated that if you give the same unfamiliar chess position to a number of different players of different skill levels then this should discriminate and capture innate expertise at playing the game. His group found that this expertise was not associated with looking ahead, rather with skilled pattern recognition that came from the storage in an expert's memory of many different patterns built up over time. In 1996 Ericsson and Lehmann postulated that experts have no innate ability in a domain, rather expertise arises from "extensive deliberate practice" [115]. However, Camerer and Johnson argue that despite this, experts are susceptible to "systematic errors, biases, and limitations of performance" [116]. This is described well in two related studies by Lewandowsky and Kirsner in 2000 [117]. In simulations 14 expert bush fire commanders with an average of 18.31 years of expertise were asked to predict the extent of a fire and the best method to bring it under control. Accuracy was high but large errors made when the two major predictors were in opposition. Secondly opposing predictions were made even when the conditions were kept the same. The authors postulated that experts may hold "*separate, and sometimes even mutually exclusive, components of knowledge*".

Ericsson argues in the acquisition of expertise that it is therefore better to observe an expert solving a task rather than ask them to describe what they are doing in the abstract[118].This

was illustrated in part by earlier work by Jonson [119] who observed a medical professor's explanation of his diagnostic process. When he accompanied the professor on his ward round he was struck by differences in his diagnostic technique in practice. When challenged the professor said "*Oh, I know that, but you see I don't know how I do diagnosis, and yet I need to teach things to students. I create what I think of as plausible means for doing tasks and hope students will be able to convert them into effective ones*". Therefore, to avoid inconsistency and bias, one of the fundamentals of knowledge acquisition to ask an expert to explain how they are solving a task in real time.

In any study to capture expertise it is worth trying to dissect expertise into a number of components to better understand what it is you are trying to capture. The American psychologist Gary Klein divides expertise into knowledge "*what you have to know*" and skills "*what you can do with that knowledge*" [120] .The two are interrelated. As an expert gains knowledge in an area the skill which follows is the ability to spot anomalies. With experience causal frameworks are constructed in the mind as the expert rationalises why things happen in a particular manner. The ensuing skill is the ability to assess new situations. With greater knowledge and experience many thought processes and tasks become routine with and ability to make perceptual discriminations. This results in the ability to task prioritise, make rapid decisions and detect problems early.

Knowledge elicitation (or cognitive task analysis in psychology) has no universally accepted classifications of methodology. Hoffman in 1995 proposed that they be divided into unstructured interviews, structured interviews, analysis of unfamiliar tasks and analysis of contrived tasks [121]. A more comprehensive suggested classification is provided by Klein in 2001 [120]. He suggests the following categories: Interview methods, observation methods and modelling methods. To deal with each in turn:

Interview methods include structured, semi-structured and unstructured techniques. Concept maps are akin to a circuit diagram for an electrical appliance where the relationships between different ideas are explored. Critical decision analysis involves an expert recalling a specific incident. This is then analysed in an interview whose task is to elicit different strands of information from the expert [122].

Observation methods include direct observation questioning. Process tracing allows the expert to verbalise what they are thinking, a so-called concurrent verbal protocol. In psychological terms the subject's verbalisation reflects working memory and reflects the

cognitive processes associated with performing a task. These sessions are typically recorded so that transcripts can be made of the concurrent verbal protocols. They can be converted into "protorepresentations." This means that some modelling is already completed at the pre-modelling stage. High and low fidelity simulation as well as expert/novice comparisons are self-explanatory.

Modelling of the information gained during interviews or observational methods can be done in several ways. Sorting techniques are used to capture how experts order and compare concepts and can lead to knowledge about task prioritisation. Hierarchy-generation utilises "laddering" to build taxonomies (a hierarchical classification of concepts) or other hierarchical structures e.g. decision networks. In Matrix-based techniques, grids are constructed where problems encountered are placed against possible solutions. This was the modelling technique used in the development of the Intensive Care Unit – Patient Scoring System described in this chapter).

There is no correct way in which to undertake knowledge elicitation given that there are an infinite numbers of types of knowledge in existence. However, there are some guiding principles. Hoffman [121] argues that argues that a single technique should not be relied upon as it may potentially yield only partial information. Further knowledge, which is acquired early in the process, should "constrain subsequent knowledge elicitation." Various authors, for example Brule and Blount in 1989 [123] suggest possible stages in the knowledge elicitation process which are shown in the table 5-1.

Table 5-1: Suggested stages for the knowledge acquisition process

STAGE	METHOD	PURPOSES
1 / 2 (identical stages)	Unstructured Interviewing Observation of tasks familiar to the expert	Researcher becomes familiar with the "domain" under investigation Generates a "first pass" knowledge base
3/4	Structured interviews "Think aloud" problem solving (forming concurrent verbal protocols) Contrived tasks	Provides some validity to stages 1/2 Extends knowledge base Refines knowledge base
Once these stages are complete the refined knowledge base can be modelled using the techniques described above, in the case of the ICU-PSS in confusion matrices.		

5.3. Methods

5.3.1. Development of broad classifications of instability

The patient data sets on which the quantitative score was tested (708,723 and 728) were reviewed independently by two clinicians (myself and Prof. John Kinsella) for the entire patient stay. We each identified time periods during which the patient was deteriorating, stable, or improving, and gave reasons for our opinion, e.g. inotropes decreasing, inspired oxygen falling etc. This was done in the form of an interview with the computing scientists, which was recorded, and transcripts made. As this information was recorded by computing scientists either the medical terms used were explained, or lay language was used.

In knowledge elicitation methodology, although there are no specific protocols, it is generally accepted that the first step in the process should be an interview (either unstructured or structured) between the expert and the researcher [123].These have several important functions. In knowledge elicitation parlance "bootstrapping" is the process whereby the researcher develops a "conceptual model of the domain." [124] Neale goes on to suggest that this should be to the level of an apprentice. This is defined as "a student undergoing a programme of instruction beyond the introductory level" and someone who "assists someone at a higher level." [121]. This poses some challenges, particularly in this project. The researchers are computing scientists who although involved in a number of medically related studies previously have little medical knowledge, especially in Intensive Care. These interviews led to lengthy discussions between myself and the computing scientists explaining the relevant aspects of the "domain." What was quickly appreciated was that completely routine terms and concepts to myself were completely alien to them. An example of a summary document I prepared and sent to them, as a reference of terms is included in the *Appendix III* of the thesis.

Initial Interviews are important for the "experts" as they allow them to get used to interacting with the researcher and starting to examine the data. Further, structured interviews constrain the expert response and are more likely to result in systematic coverage of the domain [125]. Specifically relating to this work, in these interviews myself and Prof. Kinsella examined data comprising the whole stay of three patients in Intensive Care. The aim was not to describe clinical situations (no clinical information was made available because of anonymisation), rather it was to describe whether the patient in our opinion from the physiological and pharmacological data in front of us was clinically

stable, improving or deteriorating. (Although no clinical history was available "blood pump speed" would indicate the patient undergoing a period of haemodialysis).
The data reviewed was predominantly hourly but it was decided to comment on it in different blocks of time (usually 1 to 5 hour periods). The specific time periods were less important than the commentary given as to why the patient was very unstable, stable and so on. There were no prescribed time periods for analysis or comparison within the whole patient stay as the purpose of this exercise was to generate what is defined in knowledge language as a "first pass knowledge base." [121]. This knowledge base was a very basic description of 5 levels of stability of a patient from A (stable) to E (highly unstable). Note Prof. Kinsella and myself undertook the interviews at different times. There was no direct exploration of discordance or concordance at this stage. As an example the transcript with my opinions for patient 708 is shown in figure 5-1. All three complete transcripts are shown in *Appendix IV*.

Figure 5-1: Transcript of descriptions of stability for patient 708

Time & Date	Comments	Condition of Patient
19/12/06 19:45:37 – 19/12/06 21:00:00	Heart rate high, oxygen saturation low. Aggressive fluids are given in response to high heart rate. Test patient's response to fluids. Despite being given the fluids, the blood pressure is low. Incubated at 8pm. Very sick patient. Despite intubation, oxygen saturation only went up to 92.	Very bad, getting worse
19/12/06 22:00:00	Central line put in. Blood pressure not responding to Adrenaline and Noradrenaline. FiO2 and SpO2 are grim. Most likely septic as a lot of adrenaline and nor adrenaline given.	Worse - Decreasing
20/12/06 02:00:00 - 20/12/06 04:00:00	100% oxygen given but saturation decreasing. Increase in adrenaline and nor adrenaline but blood pressure still low.	Worse - Decreasing
20/12/06 05:00:00 - 20/12/06 10:00:00	Getting worse. Blood pressure not moving. Heart rate increasing. Urine output tailing off. Oxygen saturation dire.	Worse - Decreasing
20/12/06 13:00:00 - 20/12/06 15:00:00	High amounts of adrenaline and Noradrenaline. Blood pressure grim. No change, very unwell.	Stable – No worse
20/12/06 16:00:00 - 20/12/06 17:00:00	Oxygen worse. Noradrenaline increased. Not enough blood pressure for urine.	Worse
20/12/06 18:00:00 - 20/12/06 20:00:00	**DIALYSIS** Gets worse on dialysis. Blood pressure even lower. Oxygen saturation is slightly higher. Could be because fluid had built up in the lungs and has now been removed.	On balance, stable
Rest of session	Gradual deterioration. Oxygen saturation and blood pressure continue to decrease. Becoming more and more septic. Patient in the end dies, probably from a cardiac arrest/deciding not to increase drugs further.	Much worse

The transcripts of these interviews were reviewed by myself for factual accuracy and corrected accordingly. Further misconceptions were discussed with the computing scientists. They then analysed the data and extracted the knowledge captured by identifying periods where we had said the patient was stable or unstable, improving or deteriorating and by grouping similar types of comments together generated a suggested scheme for descriptions of the 5 levels of stability. The classification is shown in table 5-2.

Table 5-2: Suggested classification of instability by the computing scientists based upon our annotations

Level	**Clinical Summary**
1 (A)	Patient's CVS stabilized, with low or no AD/NADR, and reduced oxygen; Urine production often essentially normal.
2 (B)	Patient CVS stabilized, and probably needing less of AD / NADR, and reducing levels of sedatives & Oxygen.
3 (C)	Patient CVS system is effectively stabilised; probably on decreased dosage of AD / NADR
4 (D)	Patient's CVS is beginning to stabilize but requires high doses of AD / NADR and / or fluid to retain stability.
5 (E)	Patient's CVS is very unstable (which is usually true in early phases of resuscitation); low BP or rapidly changing AD / NADR, and large fluid inputs.
6	Dead

At this stage Prof. Kinsella and myself discussed this suggested classification and both agreed that it was a good reflection and ordering of our commentaries in the first set of interviews. However it was felt that at level C the wording "probably on decreased dosage of adrenaline / noradrenaline" should be changed to "probably on low dosage of adrenaline/noradrenaline" and to add in a comment about oxygen. It was also thought better to stress the relative levels of drug dosage than to have a commentary about rate of change. Therefore, discussion about concordance / discordance was at this stage and not at the interview stage. This broad classification was a useful anchor for future more detailed descriptions of instability classes. The revised classification is shown in table 5-3.

Table 5-3: Suggested classification of instability after review by Dr. Sim and Prof Kinsella

Level	**Clinical Summary**
1 (A)	Patient's cardiovascular system stabilised, with low or no adrenaline / noradrenaline, and no or low levels of oxygen; urine production often essentially normal
2 (B)	Patient cardiovascular system stabilised, and probably needs low levels of adrenaline / noradrenaline, and low levels of sedatives and oxygen
3 (C)	Patient cardiovascular system is effectively stabilised; probably on low dosage of adrenaline / noradrenaline and oxygen.
4 (D)	Patient's CVS is beginning to stabilise but requires high doses of adrenaline / noradrenaline and / or fluid to retain stability.
5 (E)	Patient's CVS is very unstable (which is usually true in early phases of resuscitation) with low BP and high HR or rapidly changing adrenaline / noradrenaline dosage, and requires substantial fluid inputs.
6	Dead

This classification was a capturing and refinement of what 2 clinicians thought the 5 broad stability levels represented.

5.3.2 Using the broad classifications to assign levels of stability to datasets

In knowledge elicitation methodology it is common practice to have more than one round of interviews between the researchers and expert (stage 1/2 in Brule and Blount's suggested strategy for acquiring knowledge [121]. On the basis of the first stage of interviews a 5 point scale giving a broad descriptions of patients in the different classes of stability had been derived from our comments (and then refined by myself and Prof. Kinsella). These second stage interviews took this process a step further and were a form of *forward scenario simulation* [125] whereby an expert is taken through a problem (in this case the data from patients in intensive care) and create some basic "if-then" rules. That is to say the entire patient stays from the 3 patients were reviewed by us both in the first set of interviews. For each day of the stay (in its entirety) an overall classification on the A to E scale was given and some reasons why a patient was placed in a particular category. The broad classifications from the first set of interviews were useful as a suggested structure (this is important as experts can be very inconsistent if asked to analyse a task completely in the abstract) [118]. This served several purposes. Firstly there was opportunity to get used to looking at classifying data on an A to E scale and secondly I was able to start forming "rules" underpinning my classifications which would be required in the next very detailed experiment. This unstructured (at this stage) information became organised into my first

rule base that was used to detect inconsistencies in my subsequent more detailed annotations. Lastly the detailed transcripts were again reviewed by myself for factual inaccuracies and this was another opportunity to clarify more misconceptions held by the computing scientists i.e. the on going process of them becoming more familiar with the "domain."

The complete annotations for patients 708 and 728 with the explanations are given in the *Appendix V*. An extract is shown in figure 5-2.

Figure 5-2: Assigning stability levels and starting to formulate ranges for parameters to justify a stability level

Patient 708

Day	Comments - Prof. Kinsella	Level	Comments – Dr. Malcolm Sim	Level
Day 1 19/12/06	First thing, given patient fluid in a large volume and ADR & NORAD and it has taken several hours to get on top of the situation. Low BP, High HR and a lot of treatment used to get those values. MAP, adequate value is between 60 -100. HR >100 abnormal. Fluid > 1litre – high amount Fluid > 700ml – Starting to worry Supporting evidence, Urine and FiO2. Main points: 1) Total fluids initially high	E	100% oxygen but saturation only up to 90%. Heart Rate is very high	E
Day 2 20/12/09	Still unstable despite need for fluids decreasing. NORAD increases and ADR decreased. Still high HR & BP not impressive. IF BP in the 50s – losing the battle. Looking at trends in particular, the running average for MAP & HR.	E	Low oxygen saturation despite still at 100% Oxygen. High Heart Rate and hypotensive despite both adrenaline and noradrenaline. 1) Oxygen 2) High HR despite ADR & NORAD 3) Blood Pressure 4) Urine Output Oxygen Saturation: 96-94: Not bad Below 94: Bad 90-84: Very Bad	E
Day 3 21/12/09	Remains unstable/dying. Increase in NORAD & ADR, BP decreasing and there is a fast HR. Average ADR & NORAD. Looking at trend of MAP.	6	Patient stays bad 1) Oxygen saturation bad and 100% oxygen 2) Blood Pressure & HR 3) Urine Heart Rate HR > 140: Bad as heart doesn't refill properly.	E

In the next stage of the experiment, the process of annotating data sets with an overall level of stability, and the formation of rules governing a level of stability assigned, was repeated on a much larger scale and in greater detail. This would lead to the formulation of a sophisticated rule base. In order to run this next experiment which was designed to show

and correct inconsistencies in my annotations of the stability of a patient I had to create a rule base which would be then be compared against these annotations. As above it has been shown that if you ask an "expert" to describe something they do subconsciously or intuitively in the abstract they are often significantly inconsistent [118,119]. The same principle applies to this task. De novo ranges could have been created of relevant parameters that I perceived represented level A through to E. However knowledge elicitation methodology suggests it is better to first extract a basic knowledge base as you are more likely to tease out knowledge that is ingrained. Having done this in the first set of interviews I then started to formulate some basic "if - then" rules (e.g. "if oxygen saturation 84% then level E") during the second set of interviews, having the framework of the 5 point classification to work form during the review of real patient data. These "if -then" rules were subsequently placed into the skeleton of the first rule base. This made the task of filling in the other ranges easier."

5.3.3. Detailed annotation of data sets and formulation of a rule base

10 patient data sets were prepared for annotation and analysis. They contained all of the commonly collected physiological and drug data obtained from the Electronic Patient Record. There were up to 41 parameters, depending upon the infusions the patient was receiving, or if there was cardiac output monitoring attached. Examples of the main parameters are heart rate, mean arterial pressure, diastolic and systolic blood pressure, inspired oxygen concentration, oxygen saturation, central venous pressure, temperature, urine output, fluid administration and doses of adrenaline, noradrenaline, propofol, midazolam and alfentanil. Clearly, not all of these parameters would be present at every time point. The data were presented at predominantly hourly intervals throughout the patient stay, as this is the interval at which the nursing staff record and verify the information in the CareVue system. The details of the patients and the number of time points are shown in table 5-4. The total number of time points is 2761. Note that patient number 708 was used in the pilot. However, the annotation in the pilot was an overall stability level for a 24-hour period and not the much more detailed annotation at this stage of development.

Table 5-4: Summary of the 10 annotated patient data sets

Patient Code	696	705	707	708	720	728	733	738	751	782
Number of time points (or instances)	129	576	475	40	188	281	396	110	493	73

I annotated the 10 data sets throughout the patients' complete stay in Intensive Care, 2761 points in total. Unlike the pilot annotations described above, I marked each recorded time point on a 5 point scale (A to E), using the descriptions in table 1 and my clinical experience as a starting point. The complete annotations for patient 720 are recorded in the *Appendix VI*. The others are not included due to their length. An extract from patient 705 is shown below in table 5-5.

As I annotated the data sets, I formulated a rule base for the key parameters regarding what should constitute their range e.g. an "A" mean arterial pressure or a "D" inspired oxygen saturation and so on. Figure 5-3 shows the first rule base I formulated capturing the clinical expertise underpinning my annotations.

Table 5-5: Extract from my annotations of patient 705

Time	HR	MAP	CVP	FiO2	SpO2	Norad.	Adren.	Prop.	Alf.	Hartmanns	A-E Score
Day 1											
15/12/2006 03:00											
15/12/2006 03:06	129	73	0		100	0.4	2.4				**D**
15/12/2006 03:08	115	70			98						**D**
15/12/2006 03:15				0.5	100						**D**
15/12/2006 03:24								100	1		**D**
15/12/2006 04:00	120	80	24	0.6	98	0.9	1.6	100	1	125	**D**
15/12/2006 05:00	113	66		0.6	99	0.9	1.6	100	1	125	**D**
15/12/2006 06:00	114	66	27	0.55	94	1.3	1.6	80	0.5	125	**C**
15/12/2006 07:00	104	60		0.6	97	1.4	1	20		125	**C**

Figure 5-3: First rule base underpinning my stability classifications

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an **A**	SpO2 is 97-100% FiO2 is 0.21-0.4 Heart Rate 60-80 Mean Arterial Pressure is 65-85
If **ANY** of the parameters fall within the ranges described then time point is a **B**	SpO2 is 94-96% FiO2 is 0.41-0.59 Heart Rate 50-59 Heart Rate 81-109 Mean Arterial Pressure is 60-64 Mean Arterial Pressure is 85-109 Adrenaline 0.1-1.3 mg/h Noradrenaline 0.1-1.3 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **C**	SpO2 is 91-93% FiO2 is 0.6-0.69 Heart Rate 45-49 Heart Rate 110-120 Mean Arterial Pressure is 55-59 Mean Arterial Pressure is 110-119 Adrenaline 1-1.4 mg/h Noradrenaline 1-1.4 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **D**	SpO2 is 89-91% FiO2 is 0.7-0.84 Heart Rate 121-140 Heart Rate 40-44 Mean Arterial Pressure is 50-54 Mean Arterial Pressure is 120-129 Adrenaline 1.5-1.9 mg/h Noradrenaline 1.5-1.9 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **E**	SpO2 is 0-88% FiO2 is 0.85-1.0 Heart Rate 0-39 Heart Rate 141-500 Mean Arterial Pressure is 0-49 Mean Arterial Pressure is 130-200 Adrenaline 2.0-10 mg/h Noradrenaline 2.0-10 mg/h

5.3.4. Resolving inconsistencies between Dr Sim's clinically based annotations and his rule base

I needed to ascertain how consistent I was being in the use of the rule base I had formulated during the annotation of time points. Was my clinical opinion consistent with the rules I was formulating, characterising each level of the 5 point score? If this were the case, the rule base incorporated into a computer programme which automatically scored the same data sets should agree with my scoring based on clinical experience.

In order to test this hypothesis I made use of the INSIGHT system created by our computing science colleagues. This system allows clinicians to explore and remove inconsistencies in their classification of data[126]. In this particular case the rule base I had formulated (figure 5-3) was incorporated into INSIGHT, which then automatically assigned an A to E score for the 10 data sets. The difference between my clinical annotations and the automatic annotations based on my rule base was displayed in a series of confusion matrices, as illustrated in figure 5-4

Figure 5-4: An example of a confusion matrix

	Expected: A	Expected: B	Expected: C	Expected: D	Expected: E	Expected: (none)
Observed: A	90 % 181 of 202	4 % 9 of 202	1 % 3 of 202	(none)	0 % 1 of 202	4 % 8 of 202
Observed: B	0 % 4 of 1053	96 % 1014 of 1053	2 % 22 of 1053	0 % 4 of 1053	0 % 2 of 1053	1 % 7 of 1053
Observed: C	(none)	4 % 22 of 533	87 % 462 of 533	6 % 34 of 533	0 % 1 of 533	3 % 14 of 533
Observed: D	(none)	2 % 6 of 360	8 % 29 of 360	83 % 297 of 360	2 % 6 of 360	6 % 22 of 360
Observed: E	(none)	4 % 20 of 540	2 % 11 of 540	12 % 67 of 540	78 % 423 of 540	4 % 19 of 540
Observed: (none)	(none)	22 % 16 of 73	4 % 3 of 73	8 % 6 of 73	1 % 1 of 73	64 % 47 of 73

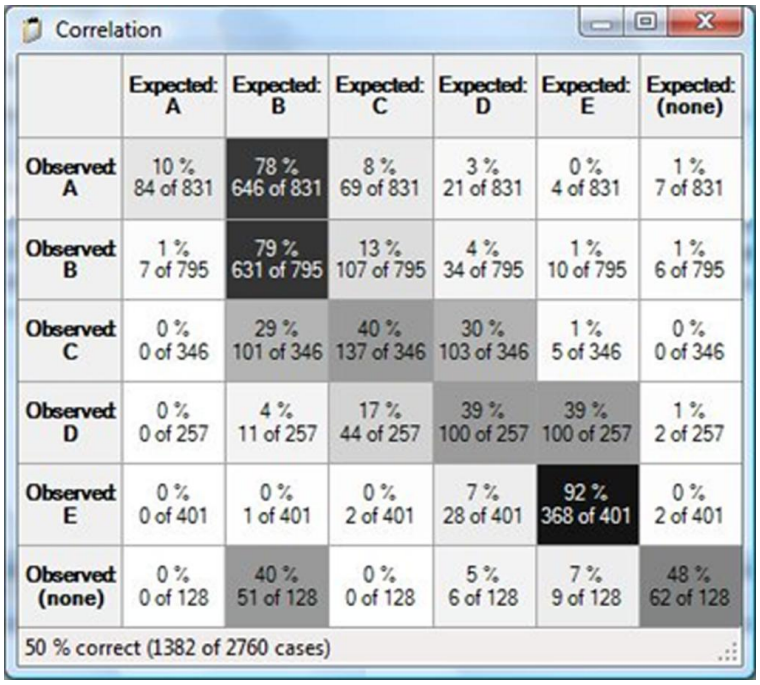
A confusion matrix is a pictorial representation of, in this case, the A to E levels I assigned to the time points within a patient's stay (observed) and the A to E levels the computer programme assigned to the same time points based on my rule base. A diagonal line from top left to bottom right represents a 100% agreement between my clinically based annotations and the computer's annotations, adhering to the rules. The further away a matrix box is from this diagonal, the greater the difference between the clinical and computer scores.

In the following methodology, each box away from the diagonal was examined to ascertain the reasons for the disparities between the clinically based annotation and the computer prediction. This led to successive alterations in the rule base which captured clinical reality more accurately with each revision.

5.3.5. Refinement of Dr. Sim's rule base

Refinement was achieved in two phases. My initial rule base was run against the 10 patient, 2760 time point data set. The resulting confusion matrix is shown in figure 5-5.

Figure 5-5: Confusion matrix of the initial rule base run against the 10 patient data set

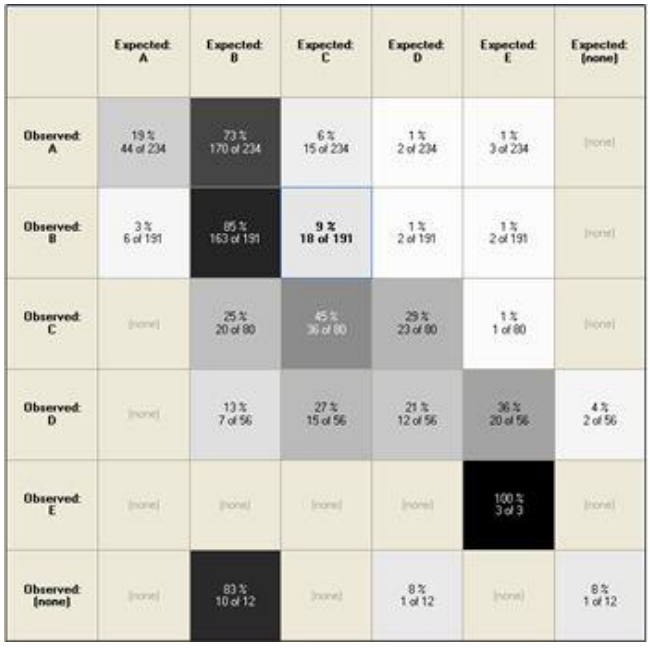


As can be seen the correlation between the initial rules and the data set is highest in the most severely unstable categories (92% agreement across the 10 data sets on observed E, expected E). It is worst for the most stable category (only 10% agreement in observed A, expected A across the 10 data sets). Many of the inconsistencies only differ by one category. For example expected C and observed C gives 40% agreement across the sets, but add in the observed C and expected D cell plus the observed C but expected B cell and this takes the agreement to 99%.

Due to the large number of instances in some of the cells (831 in category A) the refinement was divided into two phases. In the first session I chose initially to concentrate

on patient 705 as this patient stay had the greatest number of time points (576) out of the 10 contained within the data set. The initial rule base I articulated as I made the annotations (figure 3) was run against my clinical annotations of this patient's data and presented as a confusion matrix in INSIGHT. As above if I was consistent with myself, then there should have been 100% agreement between my rule base and my clinical annotations. The confusion matrix for the initial rule set run against patient 705 is shown in figure 5-6.

Figure 5-6: Confusion matrix of the initial rule base run against patient 705 (576 data points)



This process of refinement for each cell within the matrix was undertaken in the following order (given as observed / expected): A/E, B/E, C/E, B/D, D/B, A/C, A/B, B/A, D/E, D/C, B/C, D/C, B/D, C/B and C/D.

With my computing science colleagues, I considered each cell in the confusion matrix for that patient in turn, where there was disagreement between what I had annotated (observed) and the result my rule base produced (expected). There were no instances where I had annotated level "E" and my rule base had annotated anything other than "E". Within every other cell of the matrix there was disagreement. I looked at cells where there the disagreement was gross, i.e. I had observed A and the rule base had annotated an E. I examined each time point within the cell to try to understand the nature of the

disagreement. To make this task easier, I predominantly focused on 6 key parameters, namely heart rate, mean arterial pressure, oxygen saturation, inspired oxygen concentration, dose of adrenaline and noradrenaline.

Figure 5-7 shows the transcript for the refinement of patient 705 in the exact order in which it occurred. This leads to the reasons for discordance and actions discussed after the transcript, and a refined rule base which is also shown after the transcript.

Figure 5-7: Transcript for the refinement of patient 705

TRANSCRIPT OF SESSION 1, patient 705

Annotation 'A' – Rules 'E'

Row ID	Parameter	Action	Comment/Rule Change
457	HR	Heart rate of 372 removed	
494	HR	Heart rate of 7 removed	
544	HR	Heart rate of 3 removed	

Annotation 'B' – Rules 'E'

Row ID	Parameter	Action	Comment/Rule Change
577	SpO2	Changed annotation to E because saturation was low	
681	Heart Rate	Heart Rate value of 16 removed	

Annotation 'C' – Rules 'E'

Row ID	Parameter	Action	Comment/Rule Change
272	SpO2	Changed annotation to E as saturation had fallen and hence patient unstable	

Annotation 'A' – Rules 'D'

Row ID	Parameter	Action	Comment/Rule Change
586	MAP	Mean pressure too low for an 'A'. Rules were changed for a category 'D' and agreed. Annotation changed to 'D'.	
641	MAP	Mean pressure too low for an 'A'. Rules were changed for a category 'D' and agreed. Annotation changed to 'D'.	

Annotation 'B' – Rules 'D'

Row ID	Parameter	Action	Comment/Rule Change
172	FiO2	Annotated as a 'B' because FiO2 was getting better over time. Made the annotation a 'C' because the amount of Noradrenaline not as important.	Examine relationship of Mean and Noradrenaline
643	MAP	Annotation changed to D because of the MAP.	-

Annotation 'D' – Rules 'B'

Row ID	Parameter	Action	Comment/Rule Change
134	-	Although there are some key parameters there, still missing data for some parameters. Annotation still most likely but changed to 'Unclassified' due to missing values.	-
135	-	Limited data, probably gave 'D' based on averaging previous parameters. Annotation changed to unclassified.	-
252	-	Would make the annotation a C as D was a bit harsh.	-
253	FiO2	Changed to C as FiO2 at 55%	-
254	FiO2	Changed to C as FiO2 at 55%	
256	FiO2	Changed to C as FiO2 at 55%	
259	FiO2	Changed to C as FiO2 at 55%	May need to look at FiO2

Also altered:

Row ID	Parameter	Action	Comment/Rule Change
233		Annotation changed to E	-
234		Annotation changed to E	-
236 & 237		Annotation changed to E	-
242 – 249		Annotation changed to C	-
250 – 251		Annotation changed to unclassified	-

Annotation 'A' – Rules 'C'

Row ID	Parameter	Action	Comment/Rule Change
322		Changed annotation to D	-
330	MAP	Changed annotation to B because the Mean was low	-
359	MAP	Changed annotation to B because the Mean was low	-
381	Heart Rate	Annotation changed to C because Heart Rate was 114.	-
478	MAP	Annotation changed to B because Mean was 58	
486	SpO2	Annotation changed to C because saturation was 93	
491	MAP	Annotation changed to B because Mean was 57	
528	MAP	Annotation changed to B because Mean was 57	
529	MAP	Annotation changed to B because Mean was 57	
571	SpO2	Annotation changed to C because the saturation was 93	
323		Annotation changed to C	
324		Annotation changed to C	
325		Annotation changed to C	
326		Annotation changed to C	
587	MAP	Annotation changed to B	
606	MAP	Annotation changed to B	
613	MAP	Annotation changed to B	
615	MAP	Annotation changed to B	
640	MAP	Annotation changed to B	

Annotation 'A' – Rules 'B'

Row ID	Parameter	Action	Comment/Rule Change
314	HR	Annotation changed to B as heart rate was high	
316	HR	Annotation changed to B as heart rate and mean high	
317	HR	Annotation changed to B as heart rate was high	
318	HR	Annotation changed to B as heart rate was high	
319	HR	Annotation changed to B as heart rate was high	
320	HR	Annotation changed to B as heart rate was high	
321	HR	Annotation changed to B as heart rate was high	
327	HR	Annotation changed to B as heart rate was high	
329	HR	Annotation changed to B as heart rate was high	
331	HR	Annotation changed to B as heart rate was high	
332	HR	Annotation changed to B as heart rate was high	
			RULE CHANGE Rule 'A' , HR changed to (60-83) instead of (60-80) Rule 'B', HR changed to (84-109)
333	HR	Annotation changed to B as heart rate was high	
334	HR	Annotation changed to B as heart rate was high	
335	HR	Annotation changed to B as heart rate was high	
336	HR	Annotation changed to B as heart rate was high	
337	HR	Annotation changed to B as heart rate was high	
338 - 349	HR	Annotation changed to B as heart rate was high	
350	HR	Annotation changed to C because HR was 100	Might need a rule change for 'C', HR (110 – 100)
351	HR	Annotation changed to B as heart rate was high	
352	HR	Small amount of values but still enough to change the annotation to B	
353		Annotation changed to B	
354		Annotation changed to B	
355 - 368	HR	Annotation changed to B as heart rate was high	
363	HR & MAP	Annotation changed to B because of both the HR and MAP	
369	HR & MAP	Annotation changed to B because of both the HR and MAP	
370 - 377	HR	Annotation changed to B because of the HR	
383	HR & MAP	Annotation changed to B because of both the HR and MAP	
384	HR	Annotation changed to B because of the heart rate	
453	HR & SpO2	Annotation changed to B because of the HR and an SpO2 of 95	
454	HR & SpO2	Annotation changed to B because of the HR and the SpO2	
455	HR	Annotation changed to B	
456	HR	Annotation changed to B	
457		Annotation changed to B because of some of the other values. However some values missing and trending used.	
458	HR	Annotation changed to B because of the heart rate	
459 - 469	HR	Annotation changed to B because of the heart rate	
461 & 466	HR & MAP	Annotation changed to B because of the heart rate and MAP	
469	HR, MAP and SpO2	Annotation changed to B because of the values for Heart Rate, MAP and SpO2	
470	HR	Annotation changed to C because of the value for heart rate	
479	HR	Annotation changed to B because of the value for heart rate	
480	SpO2	Annotation changed to B because the SpO2 was 96	Should the rule for 'A' be changed ? SpO2 96 or above.
482 – 484	SpO2	Annotation changed to B because of the value for SpO2	
485	HR & SpO2	Annotation changed to B because of the value for SpO2and the value for HR	
487 - 489	SpO2	Annotation changed to B because of the value for SpO2	
			RULE CHANGE 'B' SpO2 changed to (94- 95) 'A' SpO2 changed to (96 - 100) 'B' HR changed to (84 - 99) 'C' HR changed to (100 - 120)
466	HR & MAP	Annotation changed to B because of the value for HR and MAP	
490	FiO2	Annotation changed to B because of the value for FiO2	
492	FiO2	Annotation changed to B because of the value for FiO2	
493	FiO2	Annotation changed to B because of the value for FiO2	
494 – 507	FiO2	Annotation changed to B because of the value for FiO2	
508	SpO2 & FiO2	Annotation changed to B because of the values for SpO2 and FiO2.	
509 - 523	FiO2	Annotation changed to B because of the value for FiO2	
524	SpO2	Annotation changed to B because of the values for SpO2.	
525 - 532	FiO2	Annotation changed to B because of the value for FiO2	
533	FiO2 & MAP	Annotation changed to B because of the values for FiO2 and MAP	
534 - 537	FiO2	Annotation changed to B because of the value for FiO2	
538	FiO2 & HR	Annotation changed to B because of the values for FiO2 and HR	-
539	FiO2	Annotation changed to B because of the value for FiO2	-
540 - 543	FiO2 & HR	Annotation changed to B because of the values for FiO2 and HR	-
544- 546	FiO2	Annotation changed to B because of the value for FiO2	-
548	HR	Annotation changed to B because of the value for Heart Rate	-
550 - 551	HR	Annotation changed to B because of the value for Heart Rate	-
552	FiO2	Annotation changed to B because of the value for FiO2	-
553	FiO2, HR, MAP, SpO2	Annotation changed to B because of the values for FiO2, HR, MAP and SpO2	-
555	SpO2	Annotation changed to B because of the value for SpO2	-
557			Should be A – Something wrong with the rules?
564			Should be A
598	SpO2	Annotation changed to B because of the value for SpO2	-
600	SpO2	Annotation changed to B because of the value for SpO2	-
617	HR	Annotation changed to B because of the value for HR	-

623	MAP	Annotation changed to B because of the value for MAP	-
657	HR & MAP	Annotation changed to B because of the values for Heart Rate and MAP.	-
706	FiO2	Annotation changed to B because of the value for FiO2	-
			RULE CHANGE 'A' , MAP changed to (60 - 84) 'B' MAP changed to (57 - 59) 'C' MAP changed to (55-58)
636	MAP	Annotation changed to B because of the value for MAP	-

Annotation 'B' – Rules 'A'

Row ID	Parameter	Action	Comment/Rule Change
314	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
318	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
319	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
480	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
482	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
623	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
548	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
644	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
645	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
646	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
647	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
648	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
649	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-
650	HR & SpO2	Annotation changed to A as HR and SpO2 had increased in rule changes	-

Annotation D – Rules E

Row ID	Parameter	Action	Comment/Rule Change
284 - 303	FiO2	Misclassified as hadn't noticed that the FiO2 was 1. Annotation changed to E	
207	FiO2	Misclassified as hadn't noticed that the FiO2 was 1. Annotation changed to E	
			RULE CHANGE 'D' Noradrenaline changed to (1.5 – 2.4) 'E' Noradrenaline changed to (> 2.4) 'D' Adrenaline changed to (1.5 – 2.4) 'E' Adrenaline changed to (> 2.4)
283	HR	Heart rate value removed as value was '8'.	

Annotation 'D' – Rules 'C'

Row ID	Parameter	Action	Comment/Rule Change
133		Annotation changed to unclassified as suggested that extrapolation was being used	
275		Annotation changed to C	
276		Annotation changed to C	
277		Annotation changed to C	
			RULE CHANGE 'C' HR changed to (100 - 110) 'D' HR changed to (111 - 140)
281 – 283		Annotation changed to C	
302		Annotation changed to C. D was originally given as influenced by parameters further on in the dataset	
322		Annotation changed to C	
130		Annotation changed to E	
132		Annotation changed to unclassified	

Annotation 'B' – Rules 'C'

Row ID	Parameter	Action	Comment/Rule Change
			RULE CHANGE 'B' MAP changed to (55 - 59) 'C' MAP changed to (52 - 54) 'D' MAP changed to (50 – 51)

Rule change created following errors:
Annotation 'D' – Rules 'C'

Row ID	Parameter	Action	Comment/Rule Change
641		Annotation changed to C	
643		Annotation changed to C	
586		Annotation changed to C	

Annotation 'B' – Rules 'D'

Row ID	Parameter	Action	Comment/Rule Change
472		Annotation changed to C	

Back to original list…

Row ID	Parameter	Action	Comment/Rule Change
225		Annotation changed to unclassified because trending used as SpO2 low and FiO2 hasn't changed	
310	SpO2	Annotation changed to C as didn't notice the FiO2 value	
378	HR	Annotation changed to C as the heart rate is in the 100s	
380	HR	Annotation changed to C as the heart rate is in the 100s	
387	HR	Annotation changed to C as the heart rate is in the 100s	
389	HR	Annotation changed to C as the heart rate is in the 100s	
407	HR	Annotation changed to C as the heart rate is in the 100s	
471	HR	Annotation changed to C as the heart rate is in the 100s	
578	SpO2 & FiO2	Annotation changed to C as didn't notice the SpO2 and the FiO2	
580	FiO2	Annotation changed to C as didn't notice the FiO2	
612	FiO2	Annotation changed to C as didn't notice the FiO2	
658	FiO2	Annotation changed to C as didn't notice the FiO2	

Annotation None – Rules 'B'

Row ID	Parameter	Action	Comment/Rule Change
1304		Left as unclassified	

Annotation 'C' – Rules 'D'

Row ID	Parameter	Action	Comment/Rule Change
138		Annotation changed to 'D'	
140		No action – left as a disagreement	
			RULE CHANGE 'B', FiO2 changed to (0.41 – 0.54) 'C', FiO2 changed to (0.55 – 0.69)
156		Left to go back to	
157		Left to go back to	
164		Left to go back to	

Annotation 'C' – Rules 'B'

Row ID	Parameter	Action	Comment/Rule Change
198	FiO2	Annotation changed to B because of the FiO2 value	
199	FiO2	Annotation changed to B because of the FiO2 value	
200	FiO2	Annotation changed to B because of the FiO2 value	
201	FiO2	Annotation changed to B because of the FiO2 value	
202	FiO2	Annotation changed to B because of the FiO2 value	
203	FiO2	Annotation changed to B because of the FiO2 value	
204	FiO2	Annotation changed to B because of the FiO2 value	
205	FiO2	Annotation changed to B because of the FiO2 value	
206	FiO2	Annotation changed to B because of the FiO2 value	
208	FiO2	Annotation changed to B because of the FiO2 value	
323	FiO2	Annotation changed to B as the oxygen was low	
324		Annotation changed to B	

Annotation 'C' – Rules 'D'

Row ID	Parameter	Action	Comment/Rule Change
195		Annotation changed to D	
290		Annotation changed to D	
291		Annotation changed to D	
292		Annotation changed to D	
293		Annotation changed to D	
294		Annotation changed to D	
295		Annotation changed to D	
301		Annotation changed to D	
305		Annotation changed to D	
309		Annotation changed to D	
381	HR	Annotation changed to D on the basis of the Heart Rate value	
472	HR	Annotation changed to D on the basis of the Heart Rate value	
140	HR & Noradrenaline	Annotation changed to D on the basis of the Heart Rate value and the Noradrenaline amount.	
			RULE CHANGE 'C' Adrenaline changed to (1-1.7) 'C' Noradrenaline changed to (1-1.7) 'D' Adrenaline changed to (1.8-2.4) 'D' Noradrenaline changed to (1.8-2.4)
164	HR	Annotation changed to E as heart rate was very low	May need to make change to HR rules as 42 is very low
197		Annotation changed to D as FiO2 low	
			RULE CHANGE 'D', HR changed to (43-45) 'C', HR changed to (46-49) 'E', HR changed to (>43)

In the refinement of the confusion matrices the reasons for discordance were identified as:

- Inadmissible readings: These are physiological impossibilities e.g. a heart rate of 372 or a heart rate of 3. The processing of the data is described later in this chapter. At this stage there were rules in place to deal with missing values for certain parameters but not extreme values.

- Extrapolated data points (annotations changes to “unclassified”): In the rule base in the more stable categories the rules were “conjunctive”. That is to say all the parameters had to be present for the rule condition to be satisfied. For example for a time point to be in category “A” then all the parameters for oxygen saturation, inspired oxygen concentration, heart rate and mean arterial pressure had fall within prescribed ranges. It was decided that in some instances the expert had annotated a time point where there was missing information for some of the parameters that had not been dealt with by the extrapolation rules for the data. These time points were changed to “unclassified.”

- Significant values overlooked: The “expert” annotating a time point agreed that he had overlooked a significant physiological abnormality during his scoring of the data.

The INSIGHT tool was designed to demonstrate inconsistency between two perspectives on the same task i.e. a clinical annotation of a time point based on physiological data and a set of rules trying to articulate the clinical process. In the workings described later in the chapter there were few discrepancies between the far apart categories e.g. annotation B and rule set predicts E. However there was greater “inconsistency” in adjacent categories, in fact mainly between clinical annotation A but rule set predicts B. In my session with the computing scientists, if based on my clinical acumen I had consistently annotated e.g. level B but the rule set predicted C then a rule change was made to try and better capture clinical judgement. However when a rule change is made it then affects many of the other cells in the confusion matrix. Indeed it can increase agreement in the cell that is being refined at that moment, and either increase or decrease agreement in other cells. This leads to other possible discordance, namely:

- Clinical annotations disagree consistently (often between adjacent cells). This led to a rule change.

- The expert did not feel a rule change was merited and annotation changed to be consistent with the rule set. This was often very subtle where the ranges are very narrow between categories for a parameter (particularly category A-B). In other words although the computer prediction was classifying an instance differently from my original annotation, on reflection the expert perceived it to still be clinically credible.

- Where the expert was not prepared to change his annotation to "fit" the rules or where he did not feel a rule change was merited a small number of instances were labelled as "inconsistencies." In other words the rule base as it stood (or changing the rule base) and the clinical annotations could not be reconciled."

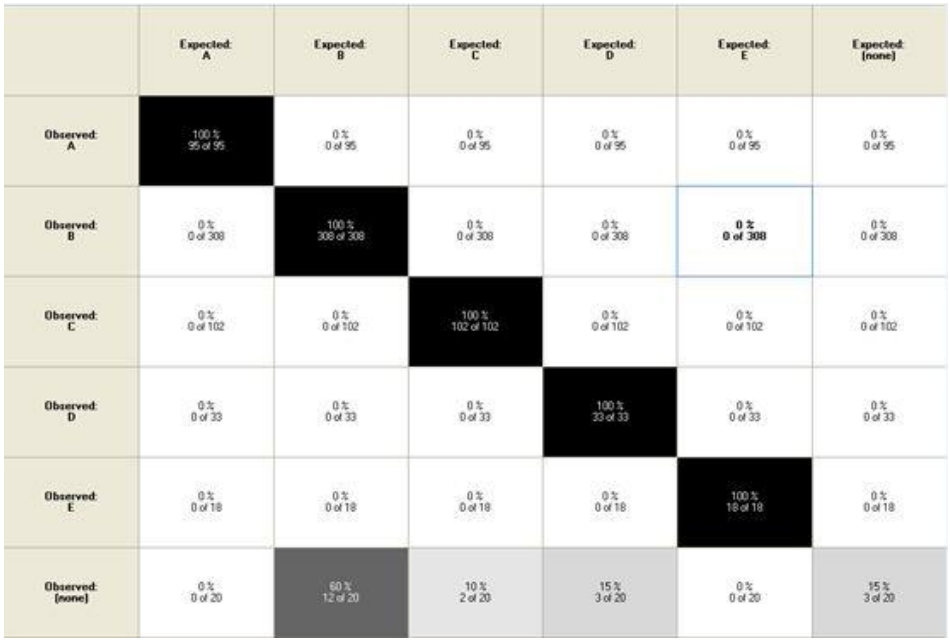
The INSIGHT algorithms cannot calculate the proportions of discordance described above but this has been done manually for patient 705 and is shown in table 5-6.

Table 5-6: Summary of types of discordance and actions taken for patient 705

Number of instances in set	576
Number of inadmissible values	5
Number of annotations changed to "unclassified"	7
Number on annotations that could not be reconciled i.e. "inconsistencies"	7
Number of annotations changed to another A-E level e.g. as significant piece of information overlooked	46
Number of annotations changed to be consistent with rules	242
Number of rule changes	10

It is worth noting that it would have been possible to create a confusion matrix after each adjustment in the transcript for patient 705. This was not done in practice. The methodology was to examine instances when the observed and expected outcomes showed a discrepancy. The aim was to identify inadmissible readings, unclassified examples and situations where values were overlooked. Rule changes were only made where in hindsight the ranges appeared inappropriate. Such rule changes were infrequently made due to the significant risk of creating new inconsistencies.
The confusion matrix after refinement of patient 705 is shown on figure 5-8.

Figure 5-8: Final confusion matrix after refinement of patient 705



The refinement of patient 705 led to an interim rule base that would be run against the remaining 9 patient datasets. The initial rule base is shown in summarised form below in table 5- 7 followed in figure 5-9 by the interim rule base after refinement of patient 705.

Table 5-7: Initial rule base prior to any refinement

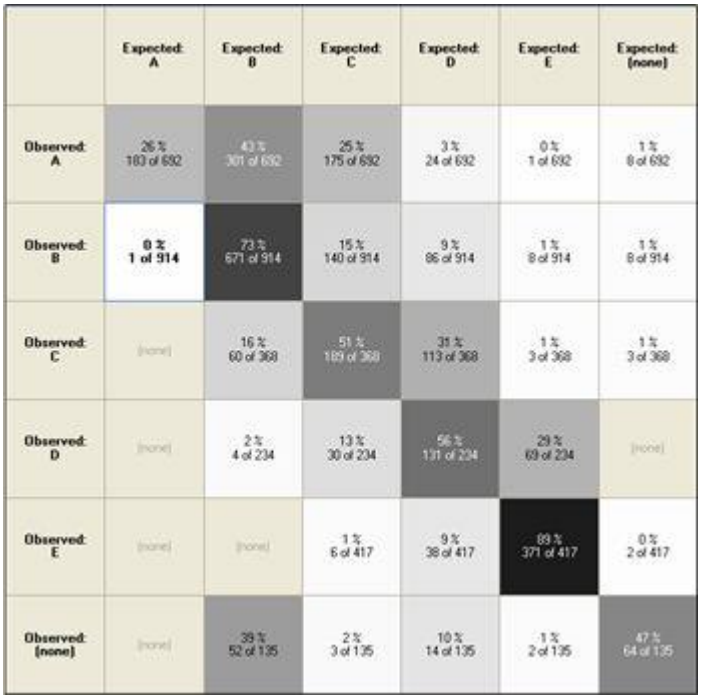
	SpO2	**FiO2**	**HR**	**MAP**	**ADR.**	**NORAD.**
A	97-100	0.21-0.4	60-80	65-85	0	0
B	94-96	0.41-0.59	50 -59 Or 81-109	60-64 Or 85-109	0.1 -1.3	0.1 – 1.3
C	91-93	0.6-0.69	45-49 Or 110-120	55-59 Or 110-119	1- 1.4	1 – 1.4
D	89 - 91	0.7- 0.84	121-140 Or 40-44	50-54 Or 120-129	1.5- 1.9	1.5–1.9
E	0-88	0.85 -1	141-500	0-49 Or 130-200	2.0-10	2.0-10

Figure 5-9: Interim rule base following refinement of patient 705 (changes from original are shown in bold)

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an **A**	**SpO2 is 96-100%** FiO2 is 0.21-0.4 **Heart Rate 60-83** **Mean Arterial Pressure is 60-84**
If **ANY** of the parameters fall within the ranges described then time point is a **B**	**SpO2 is 94-95%** **FiO2 is 0.41-0.54** Heart Rate 50-59 **Heart Rate 84-99** **Mean Arterial Pressure is 55-59** Mean Arterial Pressure is 85-109 **Adrenaline 0.1-0.9 mg/h** **Noradrenaline 0.1-0.9 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **C**	**SpO2 is 92-93%** **FiO2 is 0.55-0.69** **Heart Rate 46-49** **Heart Rate 100-110** **Mean Arterial Pressure is 52-54** Mean Arterial Pressure is 110-119 **Adrenaline 1-1.7 mg/h** **Noradrenaline 1-1.7 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **D**	SpO2 is 89-91% FiO2 is 0.7-0.84 **Heart Rate 110-140** **Heart Rate 43-45** **Mean Arterial Pressure is 50-51** Mean Arterial Pressure is 120-129 **Adrenaline 1.8-2.4 mg/h** **Noradrenaline 1.8-2.4 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **E**	SpO2 is 0-88% FiO2 is 0.85-1.0 **Heart Rate 0-42** Heart Rate 141-500 Mean Arterial Pressure is 0-49 Mean Arterial Pressure is 130-200 **Adrenaline 2.5-10 mg/h** **Noradrenaline 2.5-10 mg/h**

In the next stage of refinement the interim rule base was tested against the remaining 9 data sets. The resultant confusion matrix is shown in figure 5-10.

Figure 5-10: Confusion matrix produced when interim rule base is run against the 9 remaining data sets



In a similar process to the refinement of patient 705, the remaining 9 datasets were refined. At the start of this process the interim rule base after patient 705 was tested against the entire data set and gave a 58.3% agreement. Using the same methodology as before, the differences between my clinical annotations and the computer predictions using the refined rule base were resolved. Overall, 225 instances were viewed and there were 170 unclassifiable time points. There were 6 further rule base changes, including the addition of several conjunctive rules i.e. x and y both have to occur at the same time. A summary of the types of refinement is shown in table 5-8.

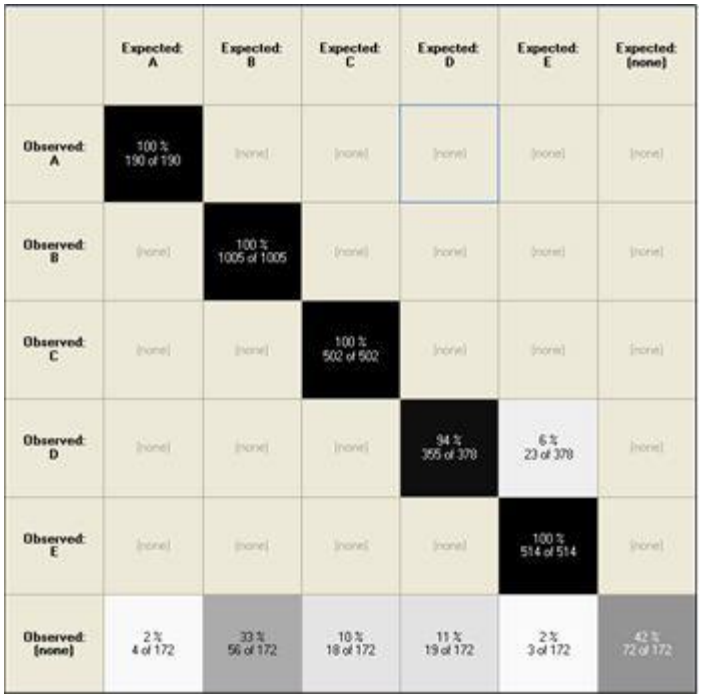
Table 5-8: Types of discordance and actions taken during the refinement of the remaining 9 patient data sets

Number of instances in set	2185
Number of inadmissible values	7
Number of annotations changed to "unclassified"	97
Number on annotations that could not be reconciled i.e. "inconsistencies"	16
Number of annotations changed to be consistent with rules	104
Number of rule changes	6

Note that in the second phase the number of "unclassified" annotations was higher. One possible explanation is that this data set contained data from 9 patients and therefore contained 9 admissions time periods where data are often missing around this time of high activity to a degree that it could not be dealt with by the extrapolation rules in place."

The transcript from this session is not included here due to its considerable length (there were many more instances to analyse) but is included in the appendix of the thesis. The final confusion matrix after refinement of the remaining 9 patient data sets is shown in figure 5-11.

Figure 5-11: Confusion matrix after refinement of the remaining 9 patient data sets

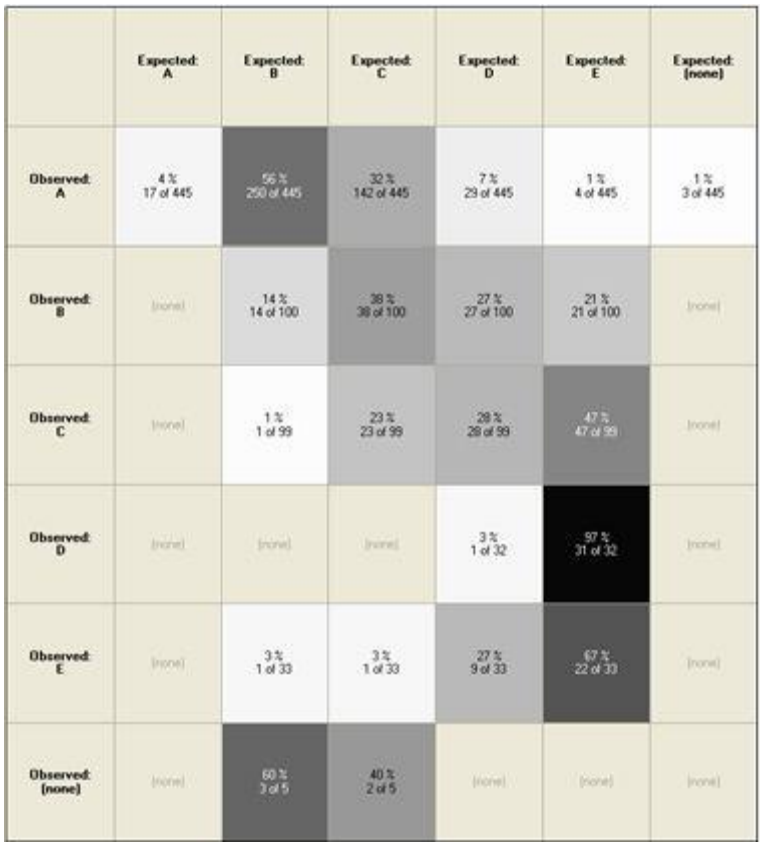


5.3.6. Refinement of the rule base by a second clinician

In the next stage of the process, a second clinician (Prof. Kinsella) annotated 3 data sets (708, 728 and 733). These contained 717 data points. The process of annotating the 10 data sets, previously done by myself, took many hours and was therefore unrealistic for other clinicians. Prof. Kinsella's 3 sets contained a maximum of 36 parameters, of which he chose to view 18.

My final rule base was tested against Prof. Kinsella's annotations. Figure 5-12 shows the resulting confusion matrix.

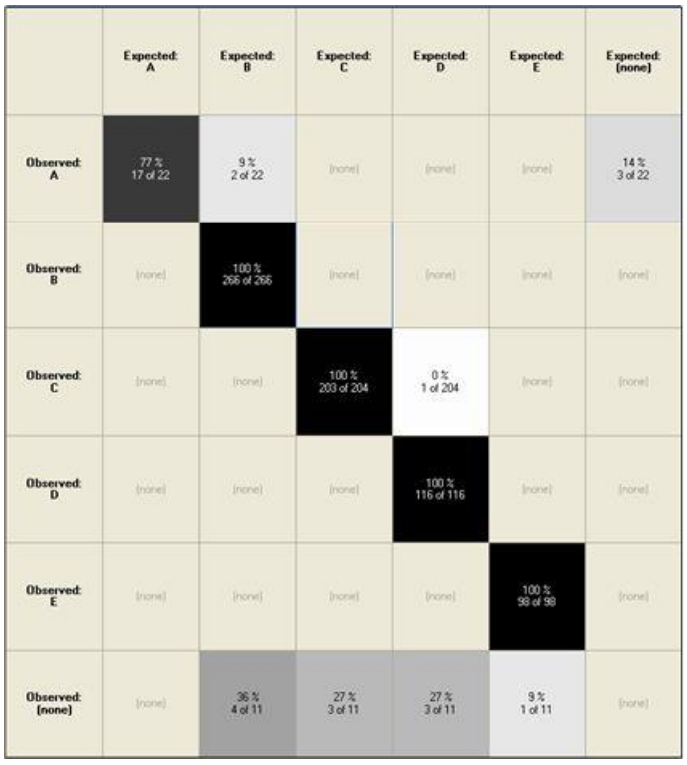
Figure 5-12: Confusion matrix produced when the final rule base of Dr. Sim is run against the annotations of 3 data sets by Prof. Kinsella



	Expected: A	Expected: B	Expected: C	Expected: D	Expected: E	Expected: (none)
Observed: A	4 % 17 of 445	56 % 250 of 445	32 % 142 of 445	7 % 29 of 445	1 % 4 of 445	1 % 3 of 445
Observed: B	(none)	14 % 14 of 100	38 % 38 of 100	27 % 27 of 100	21 % 21 of 100	(none)
Observed: C	(none)	1 % 1 of 99	23 % 23 of 99	28 % 28 of 99	47 % 47 of 99	(none)
Observed: D	(none)	(none)	(none)	3 % 1 of 32	97 % 31 of 32	(none)
Observed: E	(none)	3 % 1 of 33	3 % 1 of 33	27 % 9 of 33	67 % 22 of 33	(none)
Observed: (none)	(none)	60 % 3 of 5	40 % 2 of 5	(none)	(none)	(none)

The agreement with my final rule base run against the 3 patients annotated by Prof. Kinsella was only 10.7% (40% for patient 708, 10.7% for patient 728 and 8.1% for patient 733). In his analysis of the 3 patients, Prof Kinsella started with the most distant cells first, considering adjacent cells last. 14 annotations were changed to unclassified. 9 of these were due to missing data and 3 due to impossible extremes of physiology. Other refinements were similar to those described previously, namely a change of annotation because INSIGHT demonstrated that the expert had overlooked a significant value, or in adjacent cells either a rule change or a reclassification to fit the rues. The confusion matrix after refinement of the 3 patients is shown in figure 5-13.

Figure 5-13: Confusion matrix after the refinement of 3 patient data sets annotated by Prof. Kinsella



	Expected: A	Expected: B	Expected: C	Expected: D	Expected: E	Expected: (none)
Observed: A	77 % 17 of 22	9 % 2 of 22	(none)	(none)	(none)	14 % 3 of 22
Observed: B	(none)	100 % 266 of 266	(none)	(none)	(none)	(none)
Observed: C	(none)	(none)	100 % 203 of 204	0 % 1 of 204	(none)	(none)
Observed: D	(none)	(none)	(none)	100 % 116 of 116	(none)	(none)
Observed: E	(none)	(none)	(none)	(none)	100 % 98 of 98	(none)
Observed: (none)	(none)	36 % 4 of 11	27 % 3 of 11	27 % 3 of 11	9 % 1 of 11	(none)

This led to Prof. Kinsella's final rule base which is shown in the results section.

5.3.7. Refinement of the rule base by a third clinician

In the next stage of the process, a third clinician (Dr. Hughes) annotated 3 data sets (708, 728 and 733). Prof Kinsella's final rule base was tested against Dr. Hughes' annotations and the confusion matrix is shown in figure 5-14.

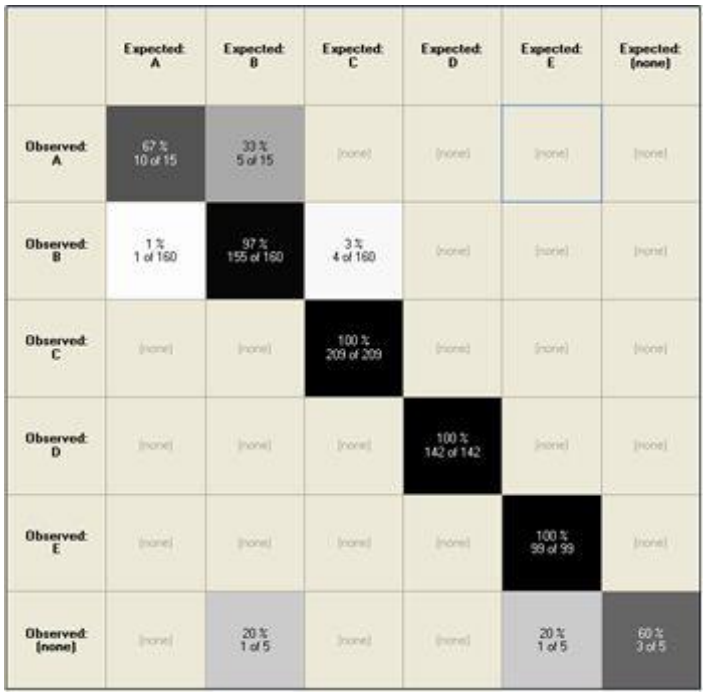
Figure 5-14: Confusion matrix produced when the final rule base of Prof. Kinsella is run against the annotations of 3 data sets by Dr. Hughes

	From Rules: A	From Rules: B	From Rules: C	From Rules: D	From Rules: E	From Rules: (none)
From Data: A	100 % 17 of 17	(none)	(none)	(none)	(none)	(none)
From Data: B	(none)	98 % 168 of 171	1 % 2 of 171	1 % 1 of 171	(none)	(none)
From Data: C	(none)	4 % 8 of 206	90 % 185 of 206	6 % 13 of 206	(none)	(none)
From Data: D	(none)	7 % 9 of 129	14 % 18 of 129	78 % 101 of 129	1 % 1 of 129	(none)
From Data: E	(none)	(none)	(none)	5 % 5 of 102	95 % 97 of 102	(none)
From Data: (none)	(none)	(none)	20 % 1 of 5	(none)	20 % 1 of 5	60 % 3 of 5

What is interesting about this confusion matrix is that the agreement is high between the confusion matrix of the first two clinicians and Dr. Hughes' annotations at 90.6% (571/630 instances). He annotated the same three patients as Prof. Kinsella (708, 728 and 733). Note that Prof. Kinsella annotated 717 time points and Dr. Hughes' agreement includes 630 time points for the same 3 patients. Unfortunately during the refinement process there was a deletion of a block of annotations. INSIGHT has subsequently been altered to avoid this happening in the future. It was felt unreasonable to go back to the start of the session as bias could have been introduced by doing the same task twice.

5 annotations were changed to "unclassified," 7 annotations were changed due to overlooking significant values and 130 other changes were made, predominantly to adjacent categories. 5 rules changes were made, all to the ranges for mean arterial pressure. The confusion matrix after Dr. Hughes' refinement is shown in Figure 5-15.

Figure 5-15: Confusion matrix after refinement of 3 patient data sets annotated by Dr. Hughes



	Expected: A	Expected: B	Expected: C	Expected: D	Expected: E	Expected: (none)
Observed: A	67 % 10 of 15	33 % 5 of 15	(none)	(none)	(none)	(none)
Observed: B	1 % 1 of 160	97 % 155 of 160	3 % 4 of 160	(none)	(none)	(none)
Observed: C	(none)	(none)	100 % 209 of 209	(none)	(none)	(none)
Observed: D	(none)	(none)	(none)	100 % 142 of 142	(none)	(none)
Observed: E	(none)	(none)	(none)	(none)	100 % 99 of 99	(none)
Observed: (none)	(none)	20 % 1 of 5	(none)	(none)	20 % 1 of 5	60 % 3 of 5

5.3.8. Final 3 clinician refinement of the rule base

Finally, all 3 clinicians analysed the extent to which the rule base captured their expertise and discussed the boundaries of certain parameters. This was done over an afternoon by focussed round table discussion. All the parameters were reviewed and final boundaries for ranges discussed based on the rule set produced after the refinement by Dr. Hughes i.e. the 3 clinician expertise. This led to the creation of the final rule base i.e. the Intensive Care Unit – Patient Scoring System.

5.4. Results

5.4.1. Final rule base of Dr. M Sim

Figure 5-16 shows my final rule base, produced after the annotation of 10 data sets. Changes from the earlier, intermediate refined rule base (after analysis of patient 705, figure 5-9) are shown in bold.

Figure 5-16: The final rule base of Dr. M. Sim

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an **A**	SpO2 is 96-100% FiO2 is 0.21-0.4 Heart Rate 60-83 Mean Arterial Pressure is 60-84
If **ANY** of the parameters fall within the ranges described then time point is a **B**	SpO2 is 94-95% FiO2 is 0.41-0.54 Heart Rate 50-59 Heart Rate 84-99 Mean Arterial Pressure is 55-59 Mean Arterial Pressure is 85-109 Adrenaline 0.1-0.9 mg/h Noradrenaline 0.1-0.9 mg/h **Dobutamine 0.1-10.5 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **C**	SpO2 is 92-93% FiO2 is 0.55-0.69 Heart Rate 46-49 Heart Rate 100-110 Mean Arterial Pressure is 52-54 Mean Arterial Pressure is 110-119 Adrenaline 1-1.7 mg/h Noradrenaline 1-1.7 mg/h **Dobutamine 10.6-25 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **D**	SpO2 is 89-91% **FiO2 is 0.7-0.83** Heart Rate 111-140 Heart Rate 43-45 **Mean Arterial Pressure is 49-51** Mean Arterial Pressure is 120-129 Adrenaline 1.8-2.4 mg/h Noradrenaline 1.8-2.4 mg/h **Dobutamine 25.1-42 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**Adrenaline 2.1-2.4 mg/h** **Noradrenaline 2.1-2.4 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.75-0.79** **Noradrenaline 2.1-2.4 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.75-0.79** **Dobutamine 33-42 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.75-0.79** **Adrenaline 2.1-2.4 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **E**	SpO2 is 0-88% **FiO2 is 0.84-1.0** Heart Rate 0-42 Heart Rate 141-500 **Mean Arterial Pressure is 0-48** Mean Arterial Pressure is 130-200 Adrenaline 2.5-10 mg/h **Noradrenaline 2.5-10 mg/h** **Dobutamine 42.1-200 mg/h**

5.4.2. Final rule base of Prof. J. Kinsella (two clinician expertise)

Figure 5-17 shows the final rule base of Prof. Kinsella, produced after the annotation of 3 data sets. Changes from my final rule base (figure 5-16) are in bold. This rule base reflects the expertise of two clinicians.

Figure 5-17: Final rule base of Prof. J. Kinsella (two clinician expertise).

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an A	SpO2 is 96-100% FiO2 is 0.21-0.4 Heart Rate 60-83 Mean Arterial Pressure is 60-84
If **ANY** of the parameters fall within the ranges described then time point is a B	SpO2 is 94-95% FiO2 is 0.41-0.54 Heart Rate 50-59 Heart Rate 84-99 Mean Arterial Pressure is 55-59 Mean Arterial Pressure is 85-109 Adrenaline 0.1-0.9 mg/h Noradrenaline 0.1-0.9 mg/h Dobutamine 0.1-10.5 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a C	SpO2 is 92-93% FiO2 is 0.55-0.69 Heart Rate 46-49 Heart Rate 100-110 Mean Arterial Pressure is 52-54 Mean Arterial Pressure is 110-119 Adrenaline 1-1.7 mg/h Noradrenaline 1-1.7 mg/h Dobutamine 10.6-25 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a D	SpO2 is 89-91% **FiO2 is 0.7-0.93** Heart Rate 111-140 Heart Rate 43-45 Mean Arterial Pressure is 49-51 Mean Arterial Pressure is 120-129 Adrenaline 1.8-2.4 mg/h Noradrenaline 1.8-2.4 mg/h Dobutamine 25.1-42 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a E	Adrenaline 2.1-2.4 mg/h Noradrenaline 2.1-2.4 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a E	FiO2 is 0.75-0.79 Noradrenaline 2.1-2.4 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a E	FiO2 is 0.75-0.79 Dobutamine 33-42 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a E	FiO2 is 0.75-0.79 Adrenaline 2.1-2.4 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a E	SpO2 is 0-88% **FiO2 is 0.94-1.0** Heart Rate 0-42 Heart Rate 141-500 Mean Arterial Pressure is 0-48 Mean Arterial Pressure is 130-200 Adrenaline 2.5-10 mg/h Noradrenaline 2.5-10 mg/h Dobutamine 42.1-200 mg/h

5.4.3. Final rule base of Dr. Hughes (three clinician expertise)

Figure 5-18 shows the final rule base of Dr. Hughes, produced after the annotation of 3 data sets. Changes from the final rule base of Prof Kinsella (capturing the expertise of two clinicians, figure 5-17) are shown in bold. This rule base benefits from the expertise of three clinicians.

Figure 5-18: The final rule base of Dr. M. Hughes (three clinician expertise)

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an **A**	SpO2 is 96-100% FiO2 is 0.21-0.4 Heart Rate 60-83 **Mean Arterial Pressure is 71-90**
If **ANY** of the parameters fall within the ranges described then time point is a **B**	SpO2 is 94-95% FiO2 is 0.41-0.54 Heart Rate 50-59 Heart Rate 84-99 **Mean Arterial Pressure is 66-70** **Mean Arterial Pressure is 91-99** Adrenaline 0.1-0.9 mg/h Noradrenaline 0.1-0.9 mg/h Dobutamine 0.1-10.5 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **C**	SpO2 is 92-93% FiO2 is 0.55-0.69 Heart Rate 46-49 Heart Rate 100-110 **Mean Arterial Pressure is 60-65** **Mean Arterial Pressure is 100-109** Adrenaline 1-1.7 mg/h Noradrenaline 1-1.7 mg/h Dobutamine 10.6-25 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **D**	SpO2 is 89-91% FiO2 is 0.7-0.93 Heart Rate 111-140 Heart Rate 43-45 **Mean Arterial Pressure is 51-59** **Mean Arterial Pressure is 110-129** Adrenaline 1.8-2.4 mg/h Noradrenaline 1.8-2.4 mg/h Dobutamine 25.1-42 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a **E**	Adrenaline 2.1-2.4 mg/h Noradrenaline 2.1-2.4 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a **E**	FiO2 is 0.75-0.79 Noradrenaline 2.1-2.4 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a **E**	FiO2 is 0.75-0.79 Dobutamine 33-42 mg/h
If **ALL** of the parameters fall within the ranges described then time point is a **E**	FiO2 is 0.75-0.79 Adrenaline 2.1-2.4 mg/h
If **ANY** of the parameters fall within the ranges described then time point is a **E**	SpO2 is 0-88% FiO2 is 0.94-1.0 Heart Rate 0-42 Heart Rate 141-500 **Mean Arterial Pressure is 0-50** Mean Arterial Pressure is 130-200 Adrenaline 2.5-10 mg/h Noradrenaline 2.5-10 mg/h Dobutamine 42.1-200 mg/h

5.4.4. Final rule base of the Intensive Care Unit Patient Scoring System (ICU-PSS)

Figure 5-19 shows the final rule base of 3 three clinicians, produced after round table discussion. Changes from the final rule base of Dr. Hughes (capturing the expertise of three clinicians, figure 5-18) are shown in bold. This is the Intensive Care Unit Patient Scoring System (ICU-PSS).

Figure 5-19: Final rule base after the 3 clinician discussion, the ICU-PSS

Conditions to be met to score a particular level	**Ranges of parameters**
If **ALL** of the parameters fall within the ranges described then time point is an **A**	SpO2 is 96-100% FiO2 is 0.21-0.4 **Heart Rate 56-89** Mean Arterial Pressure is 71-90
If **ANY** of the parameters fall within the ranges described then time point is a **B**	SpO2 is 94-95% FiO2 is 0.41-0.54 **Heart Rate 51-55** **Heart Rate 90-99** Mean Arterial Pressure is 66-70 **Mean Arterial Pressure is 100-109** **Adrenaline 0.05-0.2 mg/h** **Noradrenaline 0.1-0.4 mg/h** **Dobutamine 0.1-20 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **C**	SpO2 is 92-93% FiO2 is 0.55-0.69 **Heart Rate 46-49** Heart Rate 100-110 Mean Arterial Pressure is 60-65 **Mean Arterial Pressure is 110-119** **Adrenaline 0.3-0.4 mg/h** **Noradrenaline 0.5-0.9 mg/h** **Dobutamine 21-40 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **D**	SpO2 is 89-91% **FiO2 is 0.7-0.89** Heart Rate 111-140 **Heart Rate 41-45** Mean Arterial Pressure is 51-59 **Mean Arterial Pressure is 120-129** **Adrenaline 0.5-0.9 mg/h** **Noradrenaline 1-1.9 mg/h** **Dobutamine 41-60 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**Adrenaline 0.5-0.9 mg/h** **Dobutamine 41-60 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**Dobutamine 41-60 mg/h** **Noradrenaline 1-1.9 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**Adrenaline 0.5-0.9 mg/h** **Noradrenaline 1-1.9 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.7-0.89** **Dobutamine 41-60 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.7-0.89** **Noradrenaline 1-1.9 mg/h**
If **ALL** of the parameters fall within the ranges described then time point is a **E**	**FiO2 is 0.7-0.89** **Adrenaline 0.5-0.9 mg/h**
If **ANY** of the parameters fall within the ranges described then time point is a **E**	**SpO2 is 0-85%** **FiO2 is 0.9-1.0** **Heart Rate 0-40** **Heart Rate 141-300** Mean Arterial Pressure is 0-50 Mean Arterial Pressure is 130-200 **Adrenaline 1-10 mg/h** **Noradrenaline 2-10 mg/h** **Dobutamine 61-200 mg/h**

5.5. Discussion

The process of developing a new scoring system is challenging, not least because the aim was to devise a new score of instability in Intensive Care patients where no such previously validated score exists. (The difficulties of validating a new score under these circumstances are discussed in chapter 6). There are other methods by which we could have used to tackle this challenge. I will discuss the benefits and difficulties and potential weakness of the methodology used followed by the advantages and drawbacks of other strategies.

In a new score capturing cardiovascular instability, there are no hard endpoints to use as a surrogate (or judge against a hypothetical construct). In the APACHE II score[21] for example, it was a considerable task to undertake, but in this score used to predict mortality Knaus et al were able to improve upon the original APACHE score[22]. They did this by analysing databases containing thousands of patients and by multiple logistic regression were able to either weight differently or remove parameters in the acute physiology score which did not add to mortality prediction. It was not possible to employ this type of strategy in this research.

The fundamental problem overcome in this work is that **clinicians are inconsistent between what they say they do and what they actually do**. To devise the new score we had to capture knowledge and express it in the form of physiological rules. There was a reluctance to ask clinicians to describe "rules" de novo (with no reference standard) as this is one of the key ways which experts (including in the field of medicine) can be completely inconsistent between what they say they do in the abstract and what they actually do in reality. This is the considerable advantage of the new computing based INSIGHT system which demonstrates to a clinician where his description in the abstract differs from his clinical judgment (although in this work was still in an artificial setting).

A number of specific problems were encountered. The collaboration was with very experienced computer scientists. However, they had very little knowledge of Intensive Care. A lot of time was spent explaining basic physiology, critical care monitoring techniques, pathologies typically encountered and drugs used in the support of the critically ill (some documents prepared for them are included in the appendix for reference). Given the magnitude of the data used, this methodology would have been previously impossible to undertake manually This volume led to problems of standardising units, nonsense values entered and recorded in error by nursing staff and missing data.

Data was pre-processed to include algorithms to handle single missing data points. Future work will require the development of more sophisticated trending or INSIGHT itself to trend the score it is producing if enough similar parameters have not changed significantly. It will of course be easier to deal with missing data in the severely unstable categories where the rules are disjunctive. For example a patient's oxygen saturation is 75% this automatically scores an "E" even if mean arterial pressure, heart rate and inspired oxygen concentration are missing.

Hoffmann et al. [121] refer to the "bottleneck" of knowledge elicitation. They reckon that to transcribe 1 hour of an interview takes 24 hours subsequent work. This is particularly so when trying to describe the precise sequence of changes to annotations and reasons during the INSIGHT sessions (although it was possible to produce a detailed document for my two refinement sessions). To attempt to document every change made to an annotation in real time would make the sessions impracticable unless we had unlimited access to "expert" time. A possibility is that INSIGHT itself can be altered to allow the user to click a range of options as to why they are altering an annotation e.g. overlooked a significant value, missing data, changing his mind to be consistent with the rule set, making a new rule etc.

There are some other potential weaknesses. Although the system does allow the user to see how they are being inconsistent between clinical rules to define instability in the abstract and there annotations based on clinical experience, if this were to be repeated in the future it would be useful to include a clinical outline of what was happening to the patients at the same time as the lines of data being scored. The context of this work was still somewhat artificial.

The most contentious part of refinement was in general between adjacent categories, particularly "A" and "B." The bands here are very tight e.g. the sigmoid shape of the oxyhaemoglobin dissociation curve makes it particularly so for oxygen saturation. In refinement the clinician either made a rule change or changed his mind to be consistent with the data or where the two could not be reconciled changed the annotation to "inconsistency." Where a very small change in parameters can move the expert annotation up or down a category, it is difficult for the clinician to determine where the line should be drawn between a rule change to better model the data or an annotation change to be consistent with his abstract rule (when INSIGHT demonstrates inconsistency between what the abstract rules and clinical annotations.) However a subsequent analysis comparing

clinician 3's final rule set (3 "experts"), the percentage agreement is higher between clinician 3's final rule set run against clinician 2's final data set than when run against clinician 1's final data set. This would perhaps favour algorithm development. Further the initial agreement between clinician 1's final rule set and clinician 2's initial annotations was low (10.7%) but much higher between clinician 2's final rule base (2 experts) and clinician 3's initial annotations (90.6%). Although only in three clinicians these facts could suggest meaningful algorithm development.

The table below is from a subsequent analysis which shows the agreement when each clinician's final rule set is run against the final data set of the other clinicians.

Table 5-9: Percentage agreement between final rule set of clinician 3 against the final data set of all individual clinicians

Clinician 3's final rule (3 "experts") set run against clinician 1's final data set	Clinician 3's final rule set run against clinician 2's final data set	Clinician 3's final rule set run against clinician 3's final data set
84.4%	88.8%	98.1%

However, the counter argument is that clinician 1 introduces bias into clinician 2's refinement who introduces bias into clinician 3's refinement. This is because each clinician was starting with the rule base of the previous clinician (except for clinician 1). A potential way round this is for separate clinicians to make independent initial rule bases and use INSIGHT to show inconsistency and refine their rules.
The independent rule bases could then be compared and areas of disagreement resolved.
Another form of bias is the presence of the computing scientists sitting in the refinement sessions where the "expert" is perhaps more under pressure of time to make a decision about new rules or re-annotations. Since this work has been done the INSIGHT system is more "stable" and a detailed manual has been produced. The system can be installed on a PC and the user could now undertake an analysis independently from the computing scientists.
There are several other methodologies potentially have used in the absence of a gold standard:

- Panel of experts: A round table discussion could have been facilitated between "experts" in the field. We could have devised and refined instability rules this way. This method would have face and content validity but it would be more difficult to

test if it works or whether there is inconsistency between an abstract discussion and clinical reality.

- Survey: A postal survey could have been sent out and asked a much larger number of clinicians to describe what they mean by instability in a 5 point scale with suggested ranges and collate the responses. A large enough sample would decrease the chance of inconsistency.

- Delphi Process: Experts could answer detailed questionnaires on instability, giving suggested ranges, in two or more rounds. After each round, a clinical facilitator would provide an anonymous summary of the experts' opinion and their reasons for their judgments. With each round it is possible that there would be closer consensus for what experts would class instability as over say a 5 point scale. The anonymity of this helps remove bias.

- Clinical simulation: Experts could be shown mock scenarios of an accelerated patient stay in Intensive Care and asked to characterise why they are improving or deteriorating and try to characterise why. This would help eliminate the problem of describing instability in the abstract.

All of the above methods could be used to produce a score of instability. None is perfect, as is the methodology used in this research. However, it did give the clinician feedback on what they might do in reality versus what they might to in the abstract which is a novel way of tackling the problem of designing a new score where there is no reference gold standard.

In summary the process of repeated and gradual refinement led to a 5 point qualitative score which captured the clinical expertise of three senior Intensive Care clinicians. This score has the advantage that it can be calculated automatically and an infinite number of times during the patient stay. It provides an easy to understand 5 point scale summarising the overall clinical state of the patient.

What now needed to be established was whether the score would be applicable to other units with a different case mix. Glasgow Royal Infirmary Intensive Care Unit has the general case mix found in most adult units, but is also a tertiary referral centre for complex

pancreatic surgery and burns. It was possible that the rule base developed at this centre might not be applicable to other units with a different case mix. It was also possible that this process would have to be repeated to capture the expertise of clinicians in other centres, to mould a rule base to their patient population.

In the next chapter I shall describe the first stage in the validation of the score, and why I believe a problem of case mix does not actually apply.

5.6. Conclusion

Through a process refinement of physiological rules, a quantitative score of the stability of critically ill patients has been developed. The sophisticated physiological rule base underpinning the ranges, captures the expertise of 3 senior Intensive Care clinicians.

5.7. Acknowledgements

I co-conceived the idea of a qualitative score, underpinned by a sophisticated physiological rule base, as a superior method of quantifying cardiovascular instability. The data sets were extracted by the CareVue administrator and pre-processed by our computing science colleagues. Both Prof. Kinsella and myself independently annotated 3 data sets in the initial phase, to produce broad definitions of cardiovascular instability. In the next phase, producing detailed rules for the classifications, I undertook the majority of the annotations (10 datasets and 2761 time points). My data annotations were displayed and compared with the computer predictions in the INSIGHT programme developed by our Computing Science Colleagues. This allowed refinement of my rue base. I decided upon the ways erratic outcomes could be dealt with. Prof. Kinsella and Dr. M. Hughes annotated a smaller data set of 3 patients, and took part in a round table discussion with myself, to refine my rule base and produce the final Intensive Care Patient Scoring System. I acknowledge that the 9 Confusion Matrices reported in this chapter were created using the INSIGHT system and appropriate datasets and rule bases by Prof Derek Sleeman.

Chapter 6: First stage in the validation of the Intensive Care Unit - Patient scoring system

6.1. Abstract

6.1.1. Background

The 5 point qualitative score developed, which encapsulated the expertise of three senior Intensive Care clinicians at one centre, had to be validated to ascertain if it was clinically credible. In the absence of a previously validated gold standard with which to make a comparison, a number of validation tests would have to be applied. The first of a series of tests (discriminant validity) is described in this chapter.

6.1.2. Methods

Two separate discrimination experiments, involving 10 Intensive Care Consultants from two hospitals not involved in the development of the score, were conducted. In the first experiment, they were shown random examples of two lines of hourly data, representing different combinations of two steps of either improvement or deterioration in the qualitative score. In the second experiment, the process was repeated using a series of random one-category changes in the qualitative score. In both experiments their clinical impressions of improvement or deterioration were compared with the score's prediction.

6.1.3. Results

The 10 consultants, using their clinical acumen to score the examples, agreed with the scoring system's prediction in 92.9% of cases where there was a two category change between the two lines of hourly data, and 90.9% in the one-category change examples. Both results were highly statistically significant.

6.1.4. Conclusion

The successful tests of discriminant validity are a useful foundation to full validation of the score.

6.2. Introduction

In the previous chapter, I described the development of the Intensive Care Unit Patient Scoring System (ICU-PSS). The end product is a 5 level score which describes the overall physiological state of the patient, based on a sophisticated physiological rule base. It captures the clinical opinion and expertise of 3 clinicians in Intensive Care at a single centre. It was next necessary to validate the score. However, there was no previously validated gold standard with which to compare as this was a novel score. Therefore, in order to best validate this new score, more than one form of validation test would have to be applied. The different types of validation that could potentially be applied have been covered in the introductory chapter to the thesis. To recap, these are summarised in table 6-1.

Table 6-1: Summary of the different types of validity and their relevance to a novel scoring system

Type of validity	**Pragmatic definition in relation to a novel instability score**	**Can it be used to validate a new score of instability?**	**Comments**
Face validity	Score appears on the face of it to be a reasonable score of instability	Yes	Simple to do and useful. A weak form of validity as relies on expert opinion. They may all be incorrect.
Content validity	The score takes into account what most would regards are the key parameters of instability	Yes	Important to do but still relies upon professional consensual judgement.
Criterion validity	Score is correlated to a previously validated gold standard of instability	No	Cannot be done as there is no gold standard.
Convergent validity	When there is no change in level of instability in the score a clinician agrees	Yes	Useful in the absence of a gold standard.
Discriminant validity	When there is a change in the level of instability as judged by the score the clinician agrees	Yes	Useful in the absence of a gold standard.

In this chapter two tests of discriminant validity are described, the first of a series of validation experiments which will be required for full validation of the score. I hypothesised that, if clinicians were shown clinical cases with different combinations of improvement or deterioration in the patient state as scored by the ICU-PSS, they should, using their clinical expertise, identify the same improvements or deteriorations in more than 50% of cases (random chance). I now describe the methodology looking, firstly, at improvements or deteriorations where there is a two category change in the score and, secondly, with a one category change in the score.

6.3. Methods

6.3.1. Two step change experiment

A dataset was prepared containing an amalgamation of several patients with 6827 time points in total, an extract of which is shown in table 6-2.

Table 6-2: Extract from the 6827 time point dataset

Case	Time of Timepoint	Adren.	FiO2	HR	Mean	Norad.	SpO2	Hypothesis
1	18/09/2009 04:02			100	105		100	C-E
2	18/09/2009 06:00		1	85	145		100	Missing Value
3	18/09/2009 06:15							Missing Value
4	18/09/2009 07:00		0.6	92	113		100	C-D
5	18/09/2009 08:00		0.45	111	100		100	D-D
6	18/09/2009 09:00		0.45	120	81		97	D-B
7	18/09/2009 11:33		0.45					B-B
8	18/09/2009 13:00		0.45	92	79		95	B-B
9	18/09/2009 14:00		0.3	91	91		100	B-B

In the hypothesis column there are two ICU-PSS assignments. The first letter represents the ICU-PSS at that time point, and the second represents the ICU-PSS score at the next hourly time point. Certain time points have not been scored due to missing data**.** Although extrapolated data was used, the algorithms set up to do this only dealt with a single missing time point for a parameter. There will be missing data which the system (at present) cannot deal with. The dataset was of a size such that there should be examples of all types of two step change in the qualitative score.

Examples were extracted at random from the 6827 time point data set of all combinations of two step changes, namely A-C, B-D, C-E, C-A, D-B, E-C. A power calculation was undertaken to test if the prediction of positive change is better than 50 percent. This should show if clinicians are just guessing between the groups or if they can detect a real difference. For this, it was determined that a simple single proportion test with of a value greater than 50 percent was suitable. For an alpha of 0.05, a power of 0.9 and to detect a medium to small effect size (0.25 estimated) the sample size needed was to ask 5 consultants, N=27.4, so say 30 in each group (large positive change, large negative change). Therefore with 10 consultants it was overpowered. This translated into 4 examples of each type of two step change. Knowing that there was likely to be examples with missing data, more were selected than required i.e. 13 random examples of each type of two step change from the 6827 time point data set. The search however, only revealed 8 examples of an A to C change from the entire data set.

The random examples from each category were ordered in a smaller spreadsheet. The full spreadsheet can be seen in *Appendix VII.* An extract is shown in table 3. The columns of interest are highlighted in bold. The first column is the type of change. Note that there are only 8 examples of a C-A change. The next column in bold is the INSIGHT case number. This is the number where the first line of data in the pair falls in the larger 6827 time point data set, which allowed reference back to this original data set. The ability to reference was required as the examples ordered in the smaller spreadsheet contained no physiological data. In the far right column it can be seen that there is a number, "not used" or "missing data". Each example was manually checked in turn referring back to the original 6827 time point database until found 4 examples were found which contained the key core parameters (oxygen saturation, inspired oxygen concentration, heart rate, mean arterial pressure, noradrenaline and adrenaline). Within each category the filtering process was stopped when there were 4 examples with complete data and the other examples discarded. As above, the much smaller spread sheet showing 13 examples of each type of change (8 for A to C), where the process was stopped within a group of examples when 4 with complete data were identified, and examples not used because of incomplete data is shown in the appendix. Further there is a document showing every example rejected because of missing data.

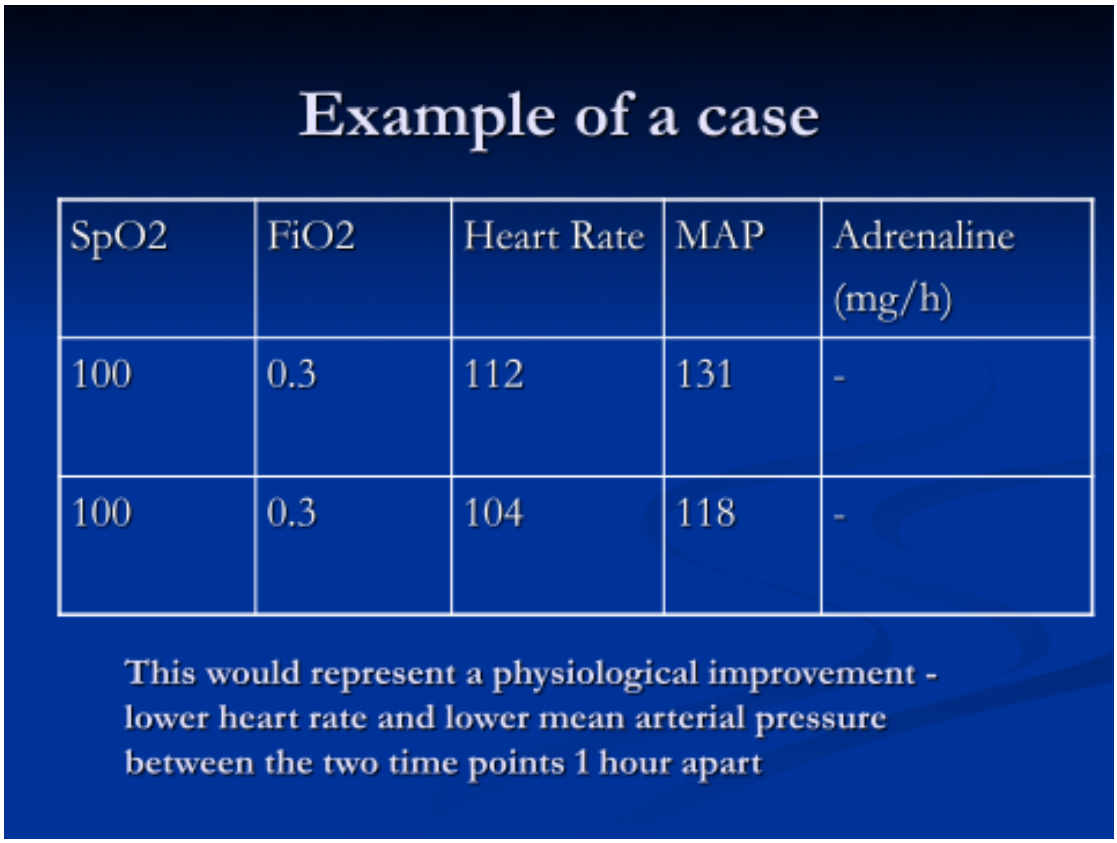
Therefore, in the extract from the smaller spreadsheet represented in table 6-3, out of the 8 C-A examples 1 to 4 were marked and the rest marked "not used." The first usable pair in the next category of change (D-B) started at 5 and so on. No randomisation was done of the usable examples, as they were already selected at random from the larger data base.

Table 6-3: Smaller spreadsheet containing from which examples of all types of 2 step change were chosen

Pair type	First of group	Last of group	# of pairs	Base-Line-in-Spreadsheet	Rand Num	(Ordered) Spreadsheet row number	INSIGHT Case Num	Example
CA	846	853	8	846	1	846	**594**	**1**
CA				846	2	847	**628**	**2**
CA				846	3	848	**4049**	**3**
CA				846	4	849	**4251**	**4**
CA				846	5	850	**5040**	**Not used**
CA				846	6	851	**6462**	**Not used**
CA				846	7	852	**6507**	**Not used**
CA				846	8	853	**6688**	**Not used**
DB	2043	2074	32	2043	17	2059	**2976**	**5**
DB				2043	9	2051	**867**	**6**
DB				2043	23	2065	**3984**	**7**
DB				2043	10	2052	**1380**	**Missing**
DB				2043	16	2058	**2059**	**Missing**
DB				2043	30	2072	**6257**	**8**
DB				2043	28	2070	**5637**	**Not used**
DB				2043	18	2060	**3464**	**Not used**
DB				2043	12	2054	**1428**	**Missing**

There were now 6 types of category of change, with 4 examples in each with complete data. Each of the examples with two lines of physiological data was prepared into an individual table as one of 24 power point slides. An example of a power point slide is shown in figure 6-1 below. The 24 slides were randomised using an online random number generator [127].

Figure 6- 1: Example of a power point slide



The 24 slides were shown to 5 consultants in ICU at the Western Infirmary Glasgow, and 5 at Crosshouse Hospital Ayrshire. These hospitals were chosen as the consultants there were not involved in the design or testing of the ICU-PSS. One is a city centre ICU and one is a district general ICU. The consultants were instructed not to confer with each other, and were shown two slides, one with introductory comments (figure 6-2), and one with an example for practise (figure 6-3). They were then shown each slide in random order and asked to mark each pair of data as improved or deteriorated as shown on the scoring sheet. An extract is shown in figure 6-4. The entire slide show can be viewed in *Appendix VIII* and the scoring sheet in *Appendix IX*. The consultants were given as much time as they needed for each example, and after they had completed the 24 examples they were marked using the scoring template (an extract of which is shown in figure 6-5. The whole sheet can be viewed in the *Appendix X*).

Figure 6-2: Introductory slide shown to the consultants

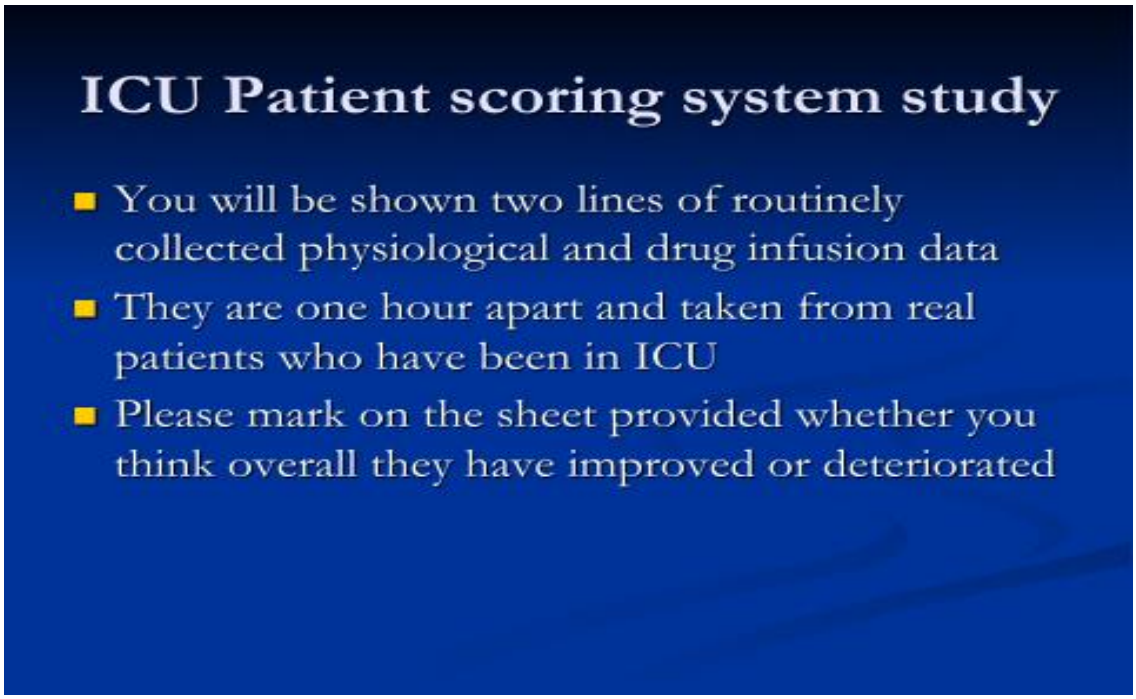


Figure 6-3: Example of a case shown to the consultants

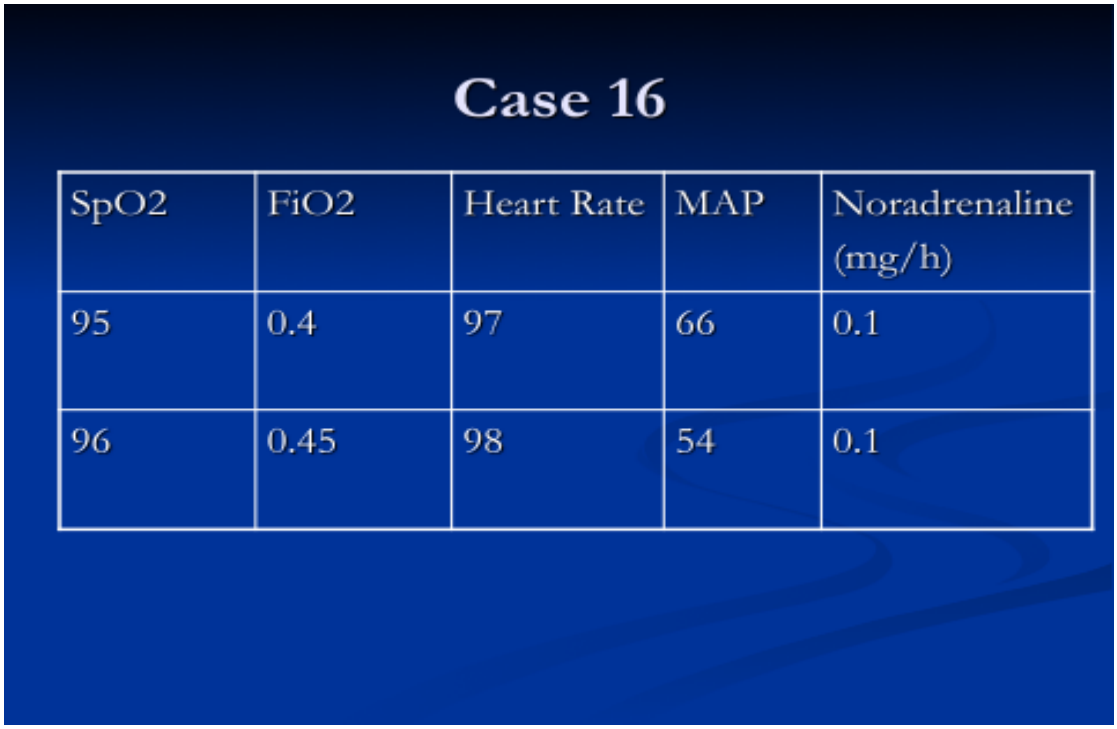


Figure 6-4: Extract of scoring system sheet given to the consultants

ICU Patient Scoring System

Please mark with an X in the appropriate box whether in your opinion the patient has improved or deteriorated

Case Number	Improved	Deteriorated
1		
2		
3		
4		
5		
6		

Figure 6-5: Extract of scoring system answer template used for marking

Case Number	Improved	Deteriorated	Actual Change
1	**X**		**E - C**
2	**X**		**D - B**
3		**X**	**C - E**
4		**X**	**A - C**
5	**X**		**E - C**
6	**X**		**C - A**
7		**X**	**A - C**
8		**X**	**C - E**

The consultants' results were then transcribed into an Excel spreadsheet for analysis.

6.3.2. One step change experiment

There were many more examples of a one step change in the 6827 time point data set. The methodology was different for the selection. A smaller spreadsheet containing every different type of one step change was prepared. This comprised 1151 time points. These were as follows, given as category and ordered numbers within the new smaller database: A-B (2-60), B-A (61-120), B-C (121-252), C-B (253-391), C-D (392-594), D-C (595-812), D-E (813-980), E-D (981-1151). An extract of the 1151 time point dataset is shown in table 6-4. Note that there are 8 one step change category possibilities. Again 4 examples of each type of change was required. Using a random number generator 4 examples for each class was selected. The integrity of the data from the original 6827 time point data base was checked. If each of the 4 examples had all the same core parameters described then these examples were kept. If not the random generator picked another 4 examples and so on until 4 examples contained no missing data. The same process was used for the selection of pairs for each category. A detailed description of the number of times 4 examples had to be picked at random for each class is shown in table 6-5. Once there were 4 examples for each class they were randomised using a random number generator [127] to be shown the consultants. An example of the numbers generated is shown in figure 6-6.

Table 6-4: Extract from the smaller data base showing all examples of a one step change

	Case	Time of Timepoint	Adren.	FiO2	HR	Mean	SpO2	Hypothesis
2	481	########		0.4	76	82		AB
3	535	########		0.4	76	72		AB
4	543	########		0.4	84	89		AB
5	587	########		0.35	89	99		AB
6	595	########		0.28	88	88		AB
7	601	########		0.28	89	96		AB
8	610	########		0.28	77	92		AB
9	629	########		0.35	76	91		AB
10	636	########		0.24	68	90		AB

Figure 6-6: Data produced by the random number generator

True Random Number Service Random Integer Set Generator

Here are your sets:

Set 1: 5, 12, 13, 53

Timestamp: 2012-05-13 11:31:00 UTC

Valid XHTML 1.0 Transitional | Valid CSS

Web Design by TSDA

Table 6-5: Clusters of 4 numbers generated and the reasons for rejection

Category	Randomisation cycle	Numbers generated	Accepted / Rejected with reason
AB	1	**5, 12, 13, 53**	Rejected - missing noradrenaline data in 13.
	2	**22, 36, 55, 57**	Accepted
BA	1	**72, 80, 84, 102**	Accepted
BC	1	**137, 148, 203, 227**	Rejected - missing noradrenaline data in 148
	2	**158, 183, 191, 225**	Rejected - missing mean arterial pressure data in 183
	3	**191, 200, 147, 241**	Accepted
CB	1	**301, 312, 326, 338**	Rejected - missing mean arterial pressure data in 312
	2	**276, 314, 324, 350**	Accepted
CD	1	**449, 539, 544, 589**	Rejected - missing FiO2 data in example 544
	2	**424, 458, 483, 564**	Accepted
DC	1	**647, 648, 691, 800**	Rejected - missing noradrenaline and mean arterial pressure data in 691
	2	**631, 636, 759, 797**	Accepted
DE	1	**816, 820, 823, 920**	Accepted
ED	1	**1086, 1096, 1097, 1098**	Rejected - missing FiO2 and oxygen saturation data in 1096
	2	**1073, 1089, 1127, 1134**	Rejected – adrenaline and noradrenaline data in 1127
	3	**991, 1053, 1102, 1104**	Rejected - missing mean arterial pressure data in 1104
	4	**981, 994, 1028, 1074**	Accepted

With the pairs selected for each category, the physiological and drug data were extracted from the smaller data base, and 32 slides prepared for review by the consultants. The order of the slides was randomised by the online random number generator. The slides were then shown to the same consultants who took part in the two step change experiment. The

process of annotation was identical. The complete slide show is shown in *Appendix XI*, the answer sheet in *Appendix XII* and the answer template in the *Appendix XIII.*

6.4. Results

6.4.1. Two step change experiment

Table 6-6 shows the result for each consultant, i.e. a score out of 24 for the number of changes they identified in the same direction as the computer prediction.

Table 6-6: Two step change experiment result

Consultant	1	2	3	4	5	6	7	8	9	10
Score	22	23	22	23	21	24	20	21	23	24

In 223/240 instances or 92.9% of cases, the consultants identified the change in the same direction as the computer prediction. A mean square contingency coefficient (phi coefficient), which is a measure of the association of two binary variables, was applied to the results. This gives a number between -1 and +1, where +1 is 100% agreement. The coefficient for the two step experiment was 0.85, p=0.000 which is highly statistically significant.

6.4.2. One step change experiment

Table 6-7 shows the result for each consultant, i.e. a score out of 32 for the number of changes they identified in the same direction as the computer prediction.

Table 6-7: Two step change experiment result

Consultant	1	2	3	4	5	6	7	8	9	10
Score	26	31	28	29	29	29	28	29	30	32

In 291/320 instances or 90.9%, of cases the consultants identified the change in the same direction as the computer prediction. The phi coefficient is 0.82, p=0.000 which is highly statistically significant.

6.4.3. Agreement for each type of category within the one and two step change experiments

A subsequent analysis was undertaken to ascertain if certain category changes were more prone to agreement or disagreement in both the one and two step change experiments. Table 6-8 illustrates the agreement for each type of category in the one step experiment. In

the one step experiment there were 4 examples of each type of change shown to 10 consultants giving a total of 40 instances of each type of change reviewed.

Table 6-8: Agreement across the different one step category changes

Type of change	**Number of instances where clinician disagreed with computer prediction (%)**	**Number of instances where clinician agreed with computer prediction (%)**
A-B	4/40 (10)	36/40 (90)
B-C	4/40 (10)	36/40 (90)
C-D	0/40 (0)	40/40 (100)
D-E	3/40 (7.5)	37/40 (92.5)
E-D	5/40 (12.5)	35/40 (87.5)
D-C	6/40 (15)	34/40 (85)
C-B	4/40 (10)	36/40 (90)
B-A	3/40 (7.5)	37/40 (92.5)

Taking all the deteriorations and improvements together, 11/160 (6.9%) of deteriorations and 18/160 (11.2%) of improvements were not identified by the clinicians. In a small sample there is less agreement with the score when it is improving than deteriorating. Further the highest disagreement appears to be instances where the patient as judged by the score is highly unstable but improving (E-D and D-C). Otherwise there is a scattering of disagreement throughout the different classes.

Table 6-9 shows a similar analysis is performed for the two step experiment. In the two step experiment there were 4 examples of each type of change shown to 10 consultants giving a total of 40 instances of each type of change reviewed.

Table 6-9: Agreement across the different one step category changes

Type of change	Number of instances where clinician disagreed with computer prediction (%)	Number of instances where clinician agreed with computer prediction (%)
A-C	2/40 (5)	38/40 (95)
B-D	6/40 (15)	34/40 (85)
C-E	0/40 (0)	40/40 (100)
E-C	6/40 (15)	34/40 (85)
D-B	0/40 (0)	40/40 (100)
C-A	3/40 (7.5)	37/40 (92.5)

Again taking all the deteriorations and improvements together, 8/120 (6.7%) of deteriorations and 9/120 (7.5%) of improvements were not identified correctly. Again

within the limitations of this experiment clinicians identified slightly fewer improvements than deteriorations, the category with the highest frequency being E-C i.e. very unstable to moderately unstable. The percentage of the improvements missed is lower in the two step than one step experiment.

Although overall there was a high level of agreement between the score's classification of improvement or deterioration, there is a suggestion that the model does not fit so well with the clinicians in cases where there is improvement from the very unstable state. This could potentially be improved by a future experiment giving a clinician the context in which the instability is occurring. This could be done by showing lines of data before and after the two lines in question as well as supplying a medical summary of the patient's condition."

6.5. Discussion

This work represents the first stage of the validation process. As this work concerns the development of a new physiological scoring system of instability there is no gold standard available with which to compare it with i.e. so called criterion validity. Therefore validation has to be a mixture of other techniques and future work will focus on theses areas.

Face validation: Does the test cover subjectively what it is supposed to be measuring? In order to test for face validity clinicians will be shown examples of changes within the score and no change within the score. They will be given a simultaneous clinical commentary about the patient's state and asked whether they think "on the face of it" that changes within the score reflect what is happening clinically i.e. does it appear to capture clinical improvement or deterioration.

Content Validity: Does the score represent all aspects of the instability it is trying to capture? The parameters which comprise the final ICU-PSS score will be shown to a group of experienced clinicians. They will be asked whether they feel that the parameters chosen to capture instability reflect what they themselves would have chosen if they had been designing a score. In a sense the score already has some indirect content validity given that when forming their rule base to score the data sets in its construction, two clinicians other than myself chose the same (although obvious) markers of instability i.e. heart rate, mean arterial pressure, inspired oxygen concentration, oxygen saturation and inotrope requirements.

Construct validity: Does the score measure what it proposes to identify and measure? As described construct validity comprises discriminant and convergent validity. A discriminant validation experiment was undertaken and successfully showed that when the score increases or decreases clinicians can (in the absence of clinical information) detect improvement or deterioration when the score changes by one or two steps. The disagreement is most marked when the score predicts an improvement from a very unstable state (E to D or E to C). As described some of the parameter bands are very narrow within these ranges and where the score may increase or decrease by a category with very little change in the parameters if they happen to sit very close to a boundary. The clinician may not detect or agree that there has been a change. This effect could possibly be reduced by the clinician having contemporaneous clinical information or data shown the patient's physiological state before or after the period in question. To complement this discriminant validity experiment a convergent validity experiment would be useful i.e. when there is no change in the score the clinicians no not detect a difference. Clinicians will be shown random examples of no change i.e. A to A, B to B etc. and asked whether the patient has deteriorated, improved or their physiological state is unchanged. Clinical history and trending information will be important to again help overcome the situation where parameters in the period of interest in the data shown are very close to a boundary.

The above are tests of validity. The score will also have to be shown to be reliable. If all the relevant data is present, collected properly from working equipment and processed appropriately by a computer algorithm then for given combinations of data there should be a consistent and reliable output. In terms of the clinicians a further experiment will be conducted to assess if they are consistent and reliable in their assessment. This could be done by showing a (large) series of lines of data, possibly with parameters around the middle of ranges. The clinician would then be asked to say whether the lines of data represented A (stable) through to (E) unstable. The same lines of data would be shown on more than one occasion to ascertain how reliable the clinician was with their own opinion and how reliably different clinicians when shown a line of data at a stability level mark it as such.

This experiment was designed to test whether clinicians not involved with its development could detect a one step and two step improvement or deterioration within the score. The clinicians could only review the same data with which the score was calculated. This was successful with a high level of agreement. However, the study has a number of weaknesses. In practice clinicians do not make decisions based on isolated data, they

examine trends. Further they have access to the patient's relevant medical history. In future experiments, if trends prior to and after the change of interest along with a clinical summary are included for the clinicians this could increase the agreement between the score's prediction and clinical impression further.

Part of the explanation for disagreement in the experiment undertaken may result from some of the bands for the parameters being narrow, particularly in the "A" and "D" ranges (stable). For example in the final rule base for the ICU-PSS an "A" oxygen saturation is 95-100% and "B" oxygen saturation 93-94%. Similarly an "A" heart rate is 56-89 and a "B" heart rate 51-55. Imagine the situation where the clinician is shown a line of data where the heart rate is 54 and the saturation 94%. If the next line of data has a heart rate of 56 and a saturation of 94% then the clinician may not record any change but the computer prediction had increased from a B to an A. Similarly for changes at the very unstable end of the score an "E" heart rate is 0-40, a "D" heart rate is 41-45 and a "D" oxygen saturation 89-91%. Additional clinical and trending information may therefore help to reduce disagreement between the score and clinical impression where the lines of data fall very close to parameter boundaries.

In validity terms this was a type of construct validity (which comprises convergent and discriminant validity). However in this initial phase we only tested if where there was discriminant validity i.e. did the clinicians detect that two lines of physiological data are different. Convergent validity was not tested i.e. if the clinicians are shown two lines of physiological data which are judged in scoring terms to be the same that they identify them as such. This will be part of future work. This could be done by showing the clinicians a number of examples within the score an A to A, B to B etc. and asking them if the patient has improved, stayed the same or deteriorated. Given that the parameter bands discussed earlier are narrow in certain parts of the score this would be better done by giving a clinical history and trending information to decrease the chance of disagreement when a parameter lies close to the boundary between two categories.

In these experiments of discriminant validity the examples of lines of data were screened prior to being used in the slide shows. This was to be certain that they contained the minimum number of parameters required by the ICU-PSS to reliably calculate a score between time points. These were oxygen saturation, inspired oxygen concentration, heart rate, mean arterial pressure, noradrenaline and adrenaline requirements. If any of these

were missing between the two lines of data then the example was rejected as the score could give an incorrect prediction due to lack of data rather than an actual change. In a similar manner the examples were screened for nonsense values e.g. an oxygen saturation in single figures. Three examples are shown below in figure 6-7.

Figure 6-7: Examples of rejected pairs of lines of data

Insight Case Number 1255

FiO2	HR	MAP	Noradrenaline	SpO2
0.4	123	66	0.8	4
0.4	108	65	0.7	97

The prediction is E-C, but apparent improvement is due to a typo in the entry for SpO2 rather than the patient being in stability level E to start with.

Insight Case Number 2059

FiO2	HR	MAP	Norad	SpO2
0.45	121	61		98
0.5				

Prediction is D-B, but apparent improvement is due to missing Heart rate, mean arterial pressure and oxygen saturation.

Insight Case Number 1428

FiO2	HR	MAP	Norad	SpO2
0.5	99	79	1.8	98
0.5	99	79		98

Prediction is D-B, but apparent improvement is due to missing noradrenaline.

The complete record of every case which was rejected in the selection process and the reasons is recorded in the *Appendix XIV* of the thesis. Despite algorithms developed to handle missing values the rules only allow for a single missing value in a sequence and so there will still be instances which the algorithms could not have dealt with.

To recap the algorithms developed to extrapolate and deal with missing single time points to enable the testing of the quantitative score were also applied to the datasets that were scored in the design of the qualitative score. To summarise, it was agreed that if there was a single missing value within in a sequence for a particular parameter then the previously recorded value would be used. This applied to heart rate, mean arterial pressure, oxygen saturation, inspired oxygen concentration, temperature and inotrope doses. For urine output if the preceding value was less than 100mls and there was a single missing time point then the previous value was used. If more than one missing time point is was assumed that nursing staff were recording a cumulative total and an average taken. Fluids were not extrapolated. Consistency was also applied to drug doses i.e. drugs of a particular

concentration running at a particular rate were all converted to mg/h dose of the drug. Inspired oxygen concentration was recorded in an inconsistent manner (as a fraction or percentage). It was agreed that all fractions would be converted to a percentage prior to analysis. This is summarised in the table 6-10.

Table 6-10: Summary of handling of a single missing data point

Parameter	**Action for a Single Missing Value**
Adrenaline	Use previously recorded value
Noradrenaline	Use previously recorded value
Propofol	Use previously recorded value
Alfentanil	Use previously recorded value
Heart Rate	Use previously recorded value
SpO2	Use previously recorded value
FiO2	Use previously recorded value
Urine	If previous value less than 100 replace missing value with previous value. Otherwise look forward in the data to the next value, take the average of it over the missing values and then replace missing values with that average.
Temperature	Use previously recorded value
MAP	Use previously recorded value
Fluids	Do nothing

With future work the algorithms will be made more sophisticated by having increased trending. Currently the nurses at the bed space enter the data manually into CareVue. If there is missing data at a particular time it may be possible to extrapolate from data which is auto charted by the System but not verified by nursing staff.

As can be seen there are considerable challenges in the validation of a new score of instability particularly in the absence of a previously validated gold standard. Several other different validation methods will have to be used in a series of further experiments to achieve this aim.

6.6. Conclusion

The Intensive Care Unit Patient Scoring System has undergone the first stage of validation and shown to be clinically credible by 10 Intensive Care consultants not involved in the score's development. This was due to a high level of agreement between a series of improvements and deteriorations in levels of the score and their clinical acumen during two discrimination validation experiments.

6.7. Acknowledgements

I co-conceived the idea for validation with Prof. J. Kinsella and Prof. D. Sleeman. The initial data bases were produced by computing science colleagues. I selected appropriate pairs of data for the two step and one step experiments, and prepared and randomised the slides. I conducted the two experiments with 5 colleagues in Crosshouse Hospital and 5 colleagues in the Western Infirmary, Glasgow.

Chapter 7: Applications of the Intensive Care Unit Patient Scoring System Identifying and predicting myocardial damage in the critically ill patient using physiological scoring

7.1. Abstract

7.1.1. Background

The validated quantitative score was now applied to clinical problems. Myocardial damage is common in Intensive Care patients and is associated with a poor outcome. I hypothesised that through analysis of physiological disturbance alone, it should be possible to detect when myocardial damage is occurring in the absence of cardiac biomarkers. Further, it should then be possible to predict when myocardial damage is occurring.

7.1.2. Methods

Two clinicians reviewed physiological data sets from Intensive Care patients. They initially identified periods of physiological disturbance they believed could be associated with myocardial damage occurring. On subsequent more detailed analysis, they characterised this physiological disturbance in a rule base using a combination of ranges of parameters from the ICU-PSS occurring for a set duration. This rule base was then used to scan further physiological data sets. The association between the rule base "firing" within 72 hours (natural decay of troponin) before or after a positive troponin, in a sequence of troponin rises, was established. In a second experiment to predict where myocardial damage is occurring, the rule base was applied to data sets, to ascertain if it "fired" in the 72 hours prior to the first troponin rise in a sequence of high troponins.

7.1.3. Results

In the detection of myocardial damage, the rule set correctly fired in 25/33 (75.8%), 95% CI: 57.7% to 88.9% cases of high troponin sequences (true positive) and did not fire in 8/33 (24.2%) cases of high troponin sequences (false negative). The rule set did not fire in 4/20 (20%), 95% CI: 5.9% to 43.7% of sequences of negative troponins (true negative) and did fire in 16/20 (80%) of sequences of negative troponins (false negative). Positive predictive value of the test 61% (95% CI: 44.5% to 75.8%). Negative predictive value of the test 33.3% (95% CI: 10.1% to 65%).

In the prediction of myocardial damage an extended rule set fired in 14/16 (87.5%), 95% CI: 61.6% to 98.1% cases before the first troponin in a sequence of positive troponins (true positive), and did not fire in 2/16 (12.5%) cases before the first troponin in a sequence of a

positive troponins (false negative). The extended rule set did not fire in 8/14 cases (57.1%), 95% CI: 28.9% to 82.2% before the first troponin in a sequence of negative troponins (true negative), and did fire in 6/14 (42.9%) of cases before the first troponin in a sequence of negative troponins. Positive predictive value of the test 70% (95% CI: 45.7% to 88%). Negative predictive value of the test 80% (95% CI: 44.4% to 96.9%).

7.1.4. Conclusion

This preliminary work leads to a hypothesis that it may possible to detect and predict myocardial damage using physiological scoring alone

7.2. Introduction

With the Intensive Care Unit Patient Scoring System developed and validated, I next wanted to use the score to examine certain unanswered critical care questions. Firstly, was it possible to detect myocardial damage occurring in ICU patients from physiological disturbance alone, rather than rely on cardiac biomarkers such as troponin (described in this chapter). Secondly, was it then possible to predict the occurrence of myocardial damage.

As described in the introduction, myocardial damage is an independent risk factor for both short and long term mortality in critically ill patients[80]. Systematic screening demonstrates that myocardial damage is common in the critically ill[82]. Despite a significant influence on outcome, I have not been able to identify any published studies on the optimum management for mortality reduction in critically ill patients, once myocardial damage has taken place. It would therefore be useful to identify impending damage.

Two recent consensus conferences have led to a new classification of myocardial infarction[75, 76] based on pathophysiology. Type I myocardial damage results from the rupture of an atherosclerotic plaque, with subsequent ischaemia and necrosis of myocardial cells in the territory of a coronary artery. This is a rare occurrence in the ICU population[79]. I postulated that it is more often a myocardial oxygen supply and demand mismatch, Type II damage in this classification, which is the cause of myocardial injury in the critically ill. A large proportion of critically ill patients have markedly deranged physiology. They may simultaneously have extremes of heart rate and blood pressure, either high or low, occurring in different combinations, coupled with hypoxia or poor tissue oxygen utilisation. This can result in global myocardial ischaemia due to a supply and demand mismatch rather than ischaemia due to a disruption in coronary artery flow [128].

I postulated that if the majority of myocardial damage occurring in the critically ill is due to a supply and demand imbalance, then physiological derangement occurring before and around the time of myocardial damage should be able to be characterised and detected by means of a physiologically based rule system such as the ICU-PSS. It should then be possible to apply the rule base to predict when myocardial damage is going to occur. In other words, could the rule set detect where myocardial damage was occurring (association) within a sequence of positive troponins and secondly could the rule set predict the occurrence of the first troponin rise in a sequence of positive troponins (causation). Due to the 72 hour decay of troponin within the blood, discrete sequences of

raised troponins (or negative troponins) were used for analysis (more detail is provided in the methods section). Specifically taking each hypothesis in turn:

Detection (Association) of physiological disturbance with myocardial damage

If the hypothesis is true that most troponin rises are caused by type two damage (myocardial oxygen supply and demand imbalance) then the physiological disturbance leading to this imbalance may be detected by rules capturing this disturbance. The rules would examine the time period 72 hours before a troponin rise (within a sequence) since with the natural decay of troponin it is still detectable 72 hours after an initial rise (possibly actually a little longer at very low levels). The rules capturing physiological disturbance would also examine the period after a troponin rise to capture the possibility of a significant cardiac event itself causing cardiovascular instability. For example this might include a patient who has developed cardiogenic shock and is hypotensive and hypoxic, or a patient who has damaged their conducting system and has a brady/tachy arrhythmia. This is a less likely scenario but it was felt important to capture it in the proposed model.

Prediction (Causation) of myocardial damage with physiological disturbance

In order to test a second hypothesis that it is possible to predict where myocardial damage is occurring based on physiological derangement, only the 72 hours before the first troponin rise within a sequence would be examined. Again, 72 hours would be chosen as the time period due to the natural decay of troponin within the blood.

7.3. Methods

7.3.1. Detection of myocardial damage

Data sets were collected from 51 critically ill patients with dialysis dependent renal failure who had routine serial troponin values recorded in the Intensive Care Unit at Glasgow Royal Infirmary. This was a far greater number of data sets than used previously. The data were initially stored in the CareVue system, anonymised, and extracted by the database managers as Excel spreadsheets, before being processed using ACHE (Architecture for Clinical Hypothesis Examination) described previously [88]. Handling the quantity of data used in this study manually would have been an extremely difficult task and error-prone.

In this study, the ACHE pre-processing tool was used to produce formatted Excel spreadsheets, with data comprising routinely collected physiological parameters and interventions. These were inspired oxygen concentration, oxygen saturation, heart rate, heart rate delta (the change in heart rate between two sequential hours), mean arterial pressure, mean arterial pressure delta, urine output, central venous pressure, doses of inotropes, vasoconstrictors, fluid administered and troponin values.

Given an increasing body of literature on the significance of a raised troponin in the critically ill, and to facilitate this and other studies, troponin-I measurement is performed three times per week in this ICU as well as when clinically indicated. Troponin I levels were considered to be negative if they were less than 0.05 micromoles/litre. Below this level there is high level of laboratory error associated with the assay. They were considered raised (positive) if greater than or equal to 0.05 micromoles/litre.

The data were presented for interpretation using the INSIGHT data display and manipulation system previously developed by the group [126]. Figure 7-1 shows an example of the data displayed in the INSIGHT system. All of the data sets can be viewed in the appendix.

Figure 7-1: Layout of data as presented in the INSIGHT system

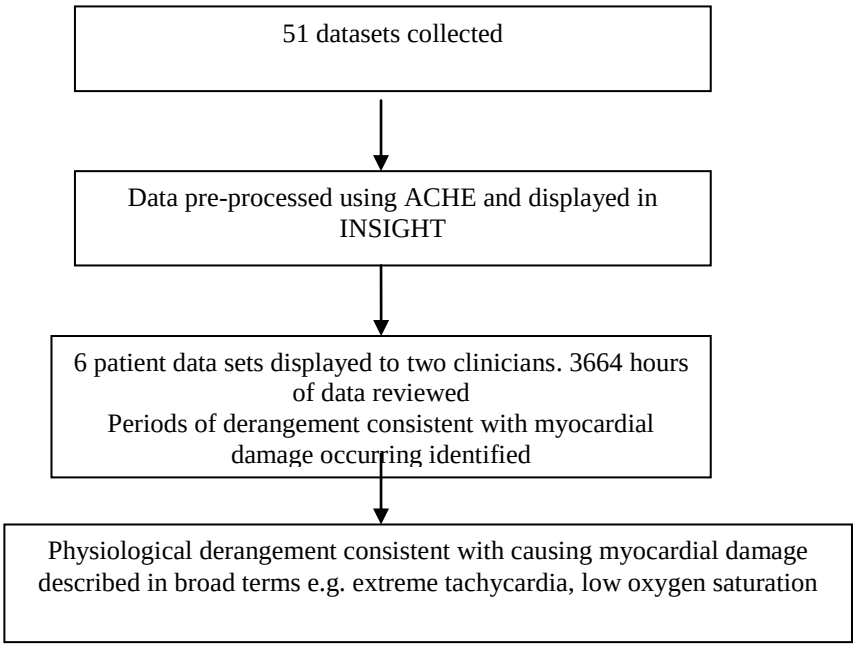
Case	Adren	Dobut.	FiO2	HR	HR Delta %	Mean	Mean Delta %	Norad	SpO2	SpO2 Delta %	Troponin	Urine	Hypothesis
1438			0.60	123.00	-8.89	61.00	-16.44		93.00	-2.11		80.00	no_event
1439			0.60	123.00	0.00	64.00	4.92		96.00	3.23		110.00	no_event
1440			0.60	129.00	4.88	78.00	21.88		92.00	-4.17		150.00	no_event
1441			0.65	130.00	0.78	59.00	-24.36		93.00	1.09		50.00	event
1442			0.65	127.00	-2.31	58.00	-1.69		96.00	3.23		65.00	event
1443			0.65	125.00	-1.57	54.00	-6.90		98.00	2.08		90.00	event
1444			0.65	125.00	0.00	57.00	5.56		97.00	-1.02		110.00	event
1445			0.00	126.00	0.80	58.00	1.75		97.00	0.00		80.00	event
1446			0.65	132.00	1.76	61.00	5.17		97.00	0.00		160.00	no_event
1447			0.65	140.00	6.06	67.00	9.84		97.00	0.00		110.00	no_event
1448			0.65	143.00	2.14	65.00	-2.99		92.00	-5.15		160.00	event
1449			0.90	147.00	2.80	62.00	-4.62		97.00	5.43		75.00	event
1450			0.90	143.00	-2.72	67.00	8.06		97.00	0.00		75.00	event
1451			0.80	141.00	-1.40	84.00	25.37		96.00	-1.03		150.00	event
1452			0.90	143.00	1.42	67.00	-20.24		97.00	1.04			event
1453			0.75	142.00	-0.70	74.00	10.45		95.00	-2.06		80.00	event
1454			0.65	141.00	-0.70	78.00	5.41		96.00	1.05		80.00	event
1455			0.50	138.00	-2.13	62.00	-20.51		98.00	2.08		105.00	no_event
1456			0.50	131.00	-5.07	63.00	1.61		100.00	2.04		125.00	no_event
1457			0.40	132.00	0.76	74.00	17.46		100.00	0.00		130.00	no_event
1458			0.40	133.00	0.76	74.00	0.00		93.00	-7.00		75.00	no_event
1459												70.00	no_event
1460			0.45	121.00		63.00			96.00				no_event
1461			0.40	122.00	0.83	66.00	4.76		97.00	1.04		90.00	no_event
1462			0.40	120.00	-1.64	69.00	4.55		99.00	2.06		125.00	no_event
1463			0.40	119.00	-0.83	64.00	-7.25		98.00	-1.01		120.00	no_event
1464			0.40	120.00	0.84	67.00	4.69		100.00	2.04		115.00	no_event
1465			0.40	118.00	-1.67	66.00	-1.49		100.00	0.00		100.00	no_event
1466			0.40	123.00	4.24	70.00	6.06		96.00	-4.00		130.00	no_event
1467			0.40	119.00	-3.25	58.00	-17.14		97.00	1.04		90.00	event
1468			0.40	118.00	-0.84	65.00	12.07		98.00	1.03		75.00	no_event
1469			0.40	121.00	2.54	71.00	9.23		99.00	1.02		110.00	no_event
1470			0.40	112.00	-7.44	64.00	-9.86		99.00	0.00		110.00	no_event
1471			0.40	111.00	-0.89	62.00	-3.13		99.00	0.00		80.00	no_event
1472			0.40	111.00	0.00	61.00	-1.61		99.00	0.00		130.00	no_event
1473			0.40	121.00	9.01	74.00	21.31		98.00	-1.01	0.14	70.00	no_event
1474			0.40	117.00	-3.31	69.00	-6.76		98.00	0.00		80.00	no_event
1475			0.40	116.00	-0.85	71.00	2.90		96.00	-2.04		120.00	no_event
1476			0.40	123.00	6.03	71.00	0.00		96.00	0.00		65.00	no_event
1477													no_event
1478			0.40	126.00		77.00			97.00			105.00	no_event
1479													no_event
1480			0.40	119.00		64.00			97.00			65.00	no_event
1481			0.40	125.00	5.04	87.00	35.94		97.00	0.00		100.00	no_event
1482			0.40	118.00	-5.60	66.00	-24.14		97.00	2.06		40.00	no_event
1483			0.40	118.00	0.00	66.00	0.00		97.00	-2.02			no_event

In a preliminary review using INSIGHT, 6 randomly selected patient data sets, consisting of the entire sequence of predominantly hourly physiological and intervention data for the patient's ICU stays, were analysed by Prof. J. Kinsella and myself. We independently analysed 3664 hours of data in total, having confirmed that all relevant parameters which we both would need to undertake the assessment were being displayed. We independently identified time periods of physiological derangement which we considered to be consistent with potential myocardial damage. We summarised these time periods of physiological derangement in broad terms as follows:

- Low values of oxygen saturation, extreme values of heart rate and mean arterial pressure.
- Additionally, high inspired oxygen concentration or large doses of inotropes.

In the initial review, only positive troponins (range 0.05 to 49.99 micromoles/litre) were displayed. Troponins which were measure but negative were not displayed in the initial review of the data. Figure 7-2 describes the initial review of the data.

Figure 7-2: Summary of the initial review of the data



In the second phase of the experiment, the 51 data sets were divided randomly into two smaller sets, one comprising 17 (a training data set) and the other 34 (a testing set), using the one third / two thirds split which is commonly used in statistical and computational model building [129].

All the data from the training data set were reviewed by the two clinicians (6827 hours of data in total). This now included positive and negative troponins. On this occasion we were specific about ranges of derangement for each physiological parameter, by using ranges from the ICU-PSS. Specifically, we considered any value for oxygen saturation, mean arterial pressure or heart rate falling within the most extreme category (E) sufficient to be consistent with potentially causing myocardial damage. In addition, the situation where 2 of these values fell within the next category (D) was also considered to be significant. This degree of derangement, e.g. a low mean arterial pressure with hypoxia and tachycardia, would be typical of conditions which could lead to a myocardial oxygen supply and

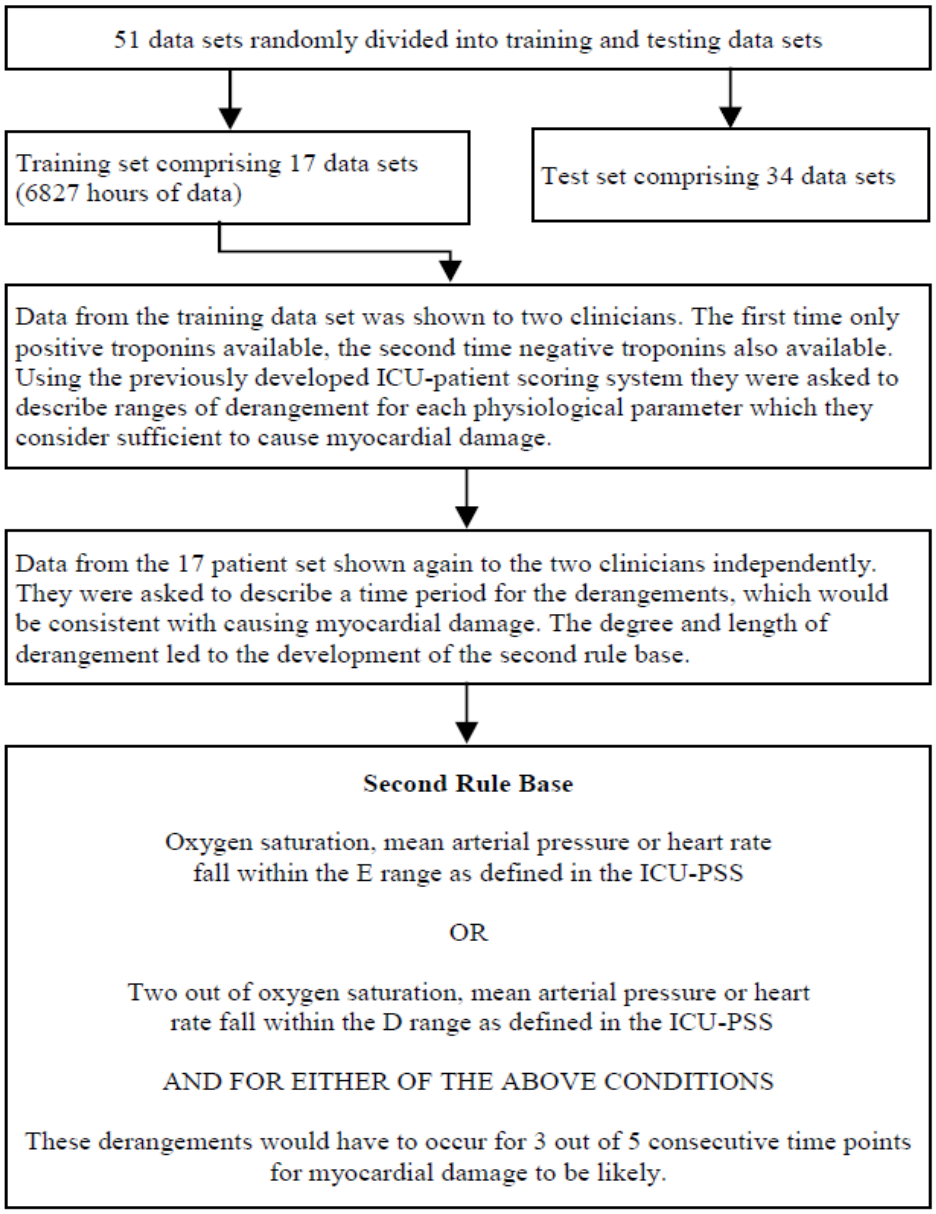
demand imbalance characteristic of type II myocardial damage. Level “D” and “E” rules are shown in figure 3 below.

Figure 7-3: Level “D” and “E” rules from the Intensive Care Unit Patient Scoring System

Oxygen Saturation 86 - 91%
Inspired Oxygen Concentration 0.70 - 0.89
Heart Rate 41 - 45 or 111 - 140 beats per minute
Mean Arterial Pressure 51 - 59 or 120 -129 mmHg

Having characterised the degree of derangement consistent with precipitating myocardial damage, we now considered the effect of duration of the physiological disturbance. Data from the17 patient training set were again reviewed independently by the 2 clinicians. Prof. Kinsella considered 4 hourly time points out of 6, and myself 3 out of 5 with the described derangements could be sufficient to cause myocardial damage. After discussion it was agreed to accept 3 out of 5 time points (Figure 7-4).

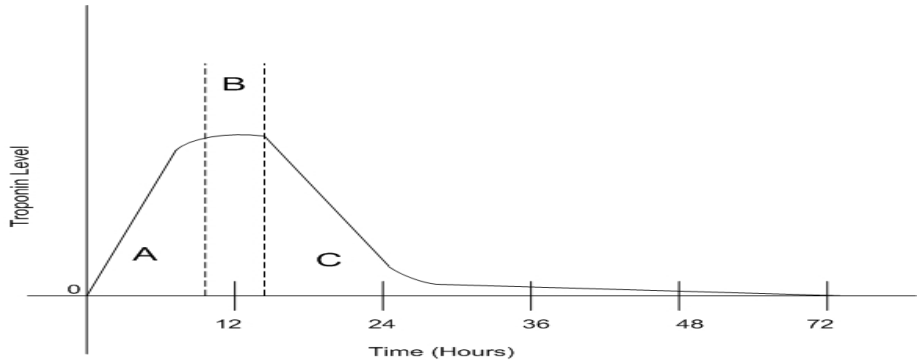
Figure 7-4: Summary of the second phase of the experiment



In the subsequent analysis, unless all the conditions specified in a rule were met, the rule set did not "fire." More detail of this, and the computing aspects of the various processes described above, can be found in a recent publication by our group[130].

When applying this rule set, the typical profile of troponin concentrations (rises and decays) in the blood were considered (Figure 7-5).

Figure 7-5: Typical rise and fall of troponin within the blood after myocardial damage (from Sleeman Moss, Sim, Kinsella: Predicting Adverse Events [130])



A = Initial rise of troponin, B= plateau phase, C= decay of troponin in blood

Troponins are sampled 3 times per week at Glasgow Royal Infirmary (Monday, Wednesday and Friday) as part of routine care and when clinically indicated. Although the standard rise and decay model of troponin within the blood was used in the hypothesis, given the constraints of the frequency with which the test is undertaken there was the possibility that there would only be one, more likely two and if fortunate three troponins around a single damage causing event. Further by chance there was the possibility of having two very similar troponins if we happened to sample at the start of the rise (A in the figure 5 above) and at the end of the decay (C in figure 5 above). Some examples from the actual data are shown below. The pattern of recorded troponins in patient 2660 (figure 7-6) and patient 2203 (figure 7-7) approximate to the rise and decay of troponin in the idealised graph above. In patient 2260 more samples have been taken by chance in the rise phase and in patient 2203 in the decay phase. More frequent troponin sampling may result in closer similarities to the shape of the idealised graph.

Figure 7-6: A real patient example of a troponin rise and decay with predominant sampling in the rise phase

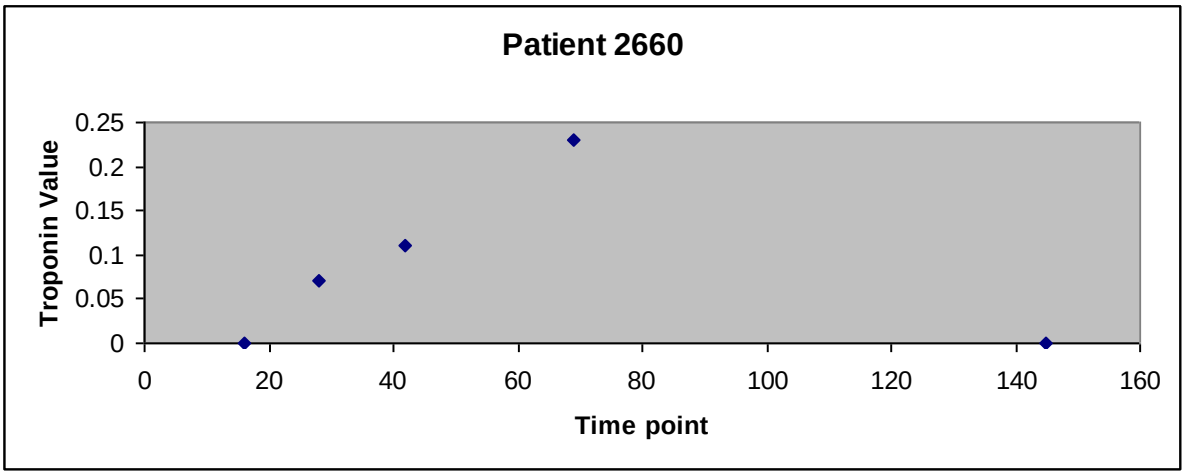
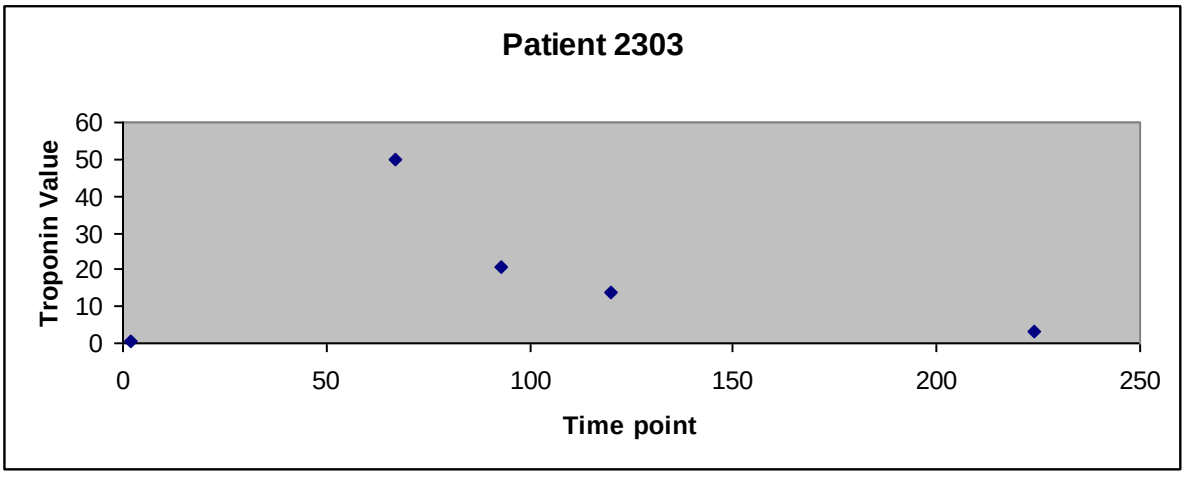
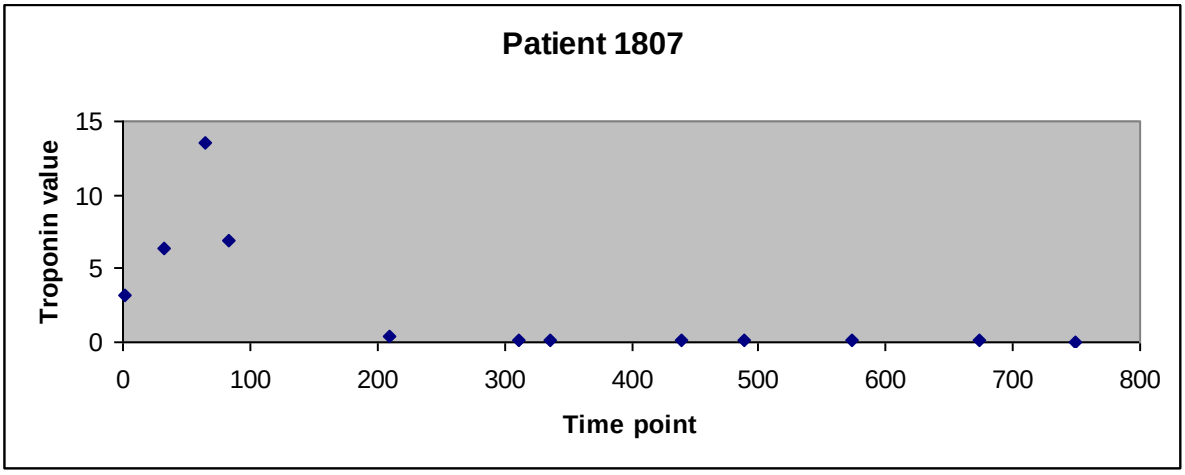


Figure 7-7: A real patient example of a troponin rise and decay with predominant sampling in the decay phase



Patient 1807 (figure 7-8) illustrates a potential problem in the model. In this case there has been a very large rise in troponin. Although the initial fall is quick there are still very low levels of troponin detected for a large number of time points afterwards. This could be due to the delayed clearance of troponin in patients with acute renal failure. Again, greater sampling could help distinguish between delayed excretion and a new ischaemic event i.e. if the time points were separated by a troponin which has been measured and undetected.

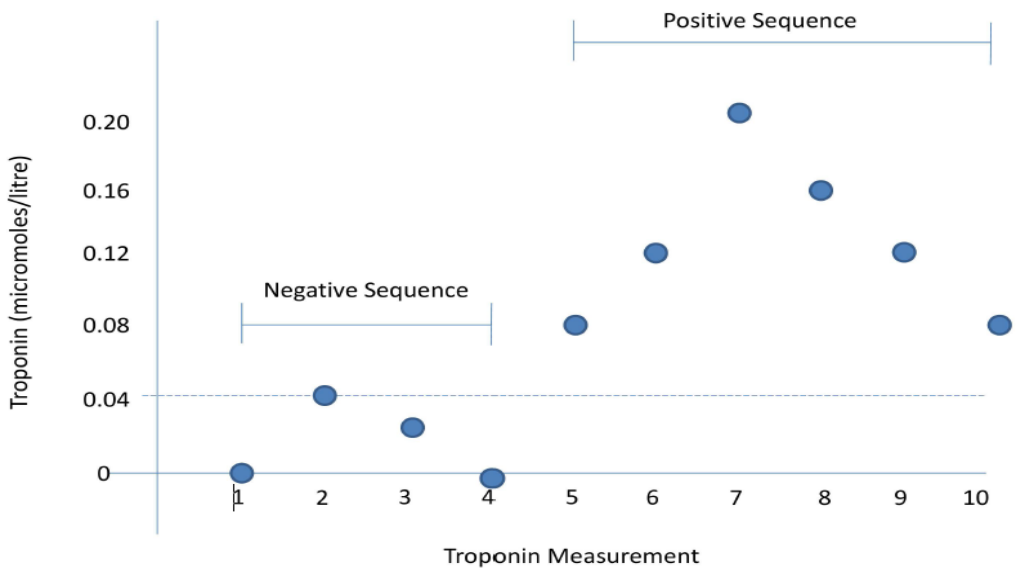
Figure 7-8: A real patient example showing a prolonged decay phase



7.3.1.1. Definition of sequences of troponin

As it can take up to 72 hours for the troponin from an initial rise to decay from the blood, in the subsequent analysis of the rule base it was applied not to individual positive or negative troponins but to sequences of positive or sequences of negative troponins as a whole within the data set. A sequence of positive troponins represents an initial rise greater or equal to 0.05 micromoles/litre, followed by subsequent decay of troponin. A sequence of negative troponins is where there have been regular troponin assays performed over a given time period, but no rise detected. These definitions are represented pictorially in figure 7-9.

Figure 7-9: Positive and Negative sequences of troponin (from Moss, Sleeman, Sim, Kinsella: Using Cardiovascular Derangements to Predict Raised Troponin Levels [131])



The next stage was to determine if *at least one* of the positive troponins in a sequence of positive troponin levels was either preceded or followed (within 72 hours) by a firing of a rule. The reciprocal hypothesis was also tested, i.e. none of the negative troponins within a sequence causes the firing of a rule. On this basis the following terms were defined:

A *true positive* is a sequence of positive troponin values, at least one of which is preceded or followed within 72 hours by the conditions specified in the rule base.

A *true negative* is a sequence of negative troponin values, none of which is preceded or followed within 72 hours by the conditions specified in rule base.
A *false positive* is a sequence of negative troponin values, at least one of which is preceded or followed within 72 hours by the conditions specified in the rule base.

A false *negative* is a sequence of positive troponin values, none of which is preceded or followed within 72 hours by the conditions specified in the rule base.

The patients in this study had dialysis dependent renal failure, which is well known to affect the troponin decay curve in the blood. However, as per our definition of a *true positive*, in this study we are describing a new positive troponin within a sequence of positive troponins occurring after a recorded negative troponin. That is to say a *true*

positive for the purposes of study represents a new myocardial event, rather than reflecting delayed clearance of troponin as a result of renal failure.

7.3.2. Methods - Prediction (causation) of myocardial damage due to physiological disturbance

In order to test the system for the prediction of myocardial damage, the testing data set was examined again using the extended rule base. As above, the testing set contained 33 sequences of high troponins and 22 sequences of negative troponins. Due to inconsistency of recorded data immediately after the patient was admitted, (a period when nursing and medical staff can be very busy and not all data are necessarily recorded), the first 12 time points (approximately the first 12 hours after admission) were removed from the analysis. This decreased the number of sequences to 27 high and 19 low.

For the purposes of the analysis:
A *true positive is* a firing of the extended rule base in the 72 hours prior the first raised troponin in a sequence of high troponins.
A *false positive is* a firing of the extended rule base in the 72 hours prior the first negative troponin in a sequence of negative troponins.
A *true negative* is where there is no firing of the extended rule base in the 72 hours prior the first negative troponin in a sequence of negative troponins.
A *false negative* is where there is no firing of the extended rule base in the 72 hours prior to the first raised troponin in a sequence of high troponins.

The first raised troponin in a sequence of high troponins would be represented by time point 5 and the first negative troponin in a sequence of negative troponins by time point 1 in figure 7-9 above.

7.4. Results

7.4.1. Detection (association of physiological disturbance) and Myocardial Damage

The 17 patient training data set contained 6827 time points of patient data, including 68 troponin recordings. Fourteen sequences of positive troponin values and 14 sequences of negative troponins (from regular testing) were identified in total. Two of the sequences of positive troponins were not included in the analysis as they occurred within the first six hours after admission to Intensive Care (and the physiological derangement leading to this may have occurred prior to ICU admission).

7.4.1.1. Analysis of where the troponin rises occurred in the training and testing data sets

The following analysis shown in figure 10 demonstrates where troponin rises occurred within the data sets. It is perhaps easier to see visually where the rises and falls (as akin to the idealised figure shown earlier) are happening in this manner. The very large number of time points makes this difficult in graphical form. The first column shows basic information about length of stay of each patient. The second column shows where the troponins were measure during the patient stay and whether they were positive or negative recorded as 0). The figure in brackets beside the actual time is the time point within the patient stay where a troponin rise or negative troponin is occurring. This is included to make it easier to see where the first troponin in a sequence has occurred. This is because in some of the analysis if the first troponin rise was within 72 hours, that sequence was disregarded as it is possible any cardiovascular instability causing it may have happened before the patient was admitted to Intensive Care.

Figure 7-10. Occurrence of troponin rises within the training and testing data sets

The 17 patient "training" data set

Demographics	Actual time into admission (time point)	Troponin Rises
Patient 1667 Admit Day 1, 01.18 Discharge Day 5, 15.00 120 time points recorded	Day 3, 05.40(56) Day 4, 06.05(84) Day 5, 0600 (110)	0.3 0.1 0.05
Patient 1713 Admit Day 1, 1500 Discharge Day 16, 1700 257 time points	Day 2, 05.30 (17) Day 6, 06.00 (119) Day 8, 06.00 (169) Day 10, 06.00 (219)	0.07 0.04 0.23 0.07
Patient 1883A Admit Day 1, 07.00 Discharge Day 34, 15.00 916 time points	Day 3, 06.00 (79) Day 6, 05.30 (133) Day 7, 13.00 (169) Day 9, 06.00 (217) Day 11, 06.00 (277) Day 12, 06.00 (301) Day 13, 06.00 (328) Day 16, 06.00 (406) Day 18, 06.00 (460) Day 20, 06.00 (511) Day 21, 06.00 (537) Day 21, 08.35 (540) Day 23, 06.00 (590) Day 25, 06.00 (640) Day 27, 06.00 (695) Day 30, 06.00 (775) Day 32, 06.04 (826) Day 34, 06.00 (878)	0.24 0.35 0.11 0.15 0.08 0 0 0 0 0.04 0.06 0.06 0.07 0 0 0 0 0
Patient 1906 Admit Day 1, 13.00 Discharge Day 74, 09.27 1792 time points	Day 4, 05.58(67) Day 6, 06.00 (124) Day 9, 07.15 (203) Day 11, 06.00 (256) Day 13, 06.00 (307) Day 16, 06.00 (386) Day 18, 06.04 (437) Day 19, 04.00 (464) Day 20, 06.00 (497) Day 23, 06.00 (575) Day 25, 06.00(626) Day 27, 06.00 (677) Day 30, 08.00 (759) Day 32, 06.00 (809) Day 34, 06.00(859) Day 37, 06.00 (936) Day 39, 06.00 (986) Day 41, 06.00(1045) Day 50, 06.00 (1119) Day 54, 06.00 (1221) Day 57, 04.00(1307) Day 59, 04.00 (1363) Day 60, 05.36 (1421) Day 63, 06.00 (1498) Day 65, 06.00 (1548) Day 67, 05.30 (1597) Day 70, 06.00 (1675) Day 74, 06.00 (1772)	1.54 2.71 0.84 0.35 0.17 0.09 0.08 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.09 0 0 0 0 0
Patient 1933 Admit Day 1, 18.00 Discharge Day 8, 19.00 184 time points	Day 1, 22.00 (7) Day 2, 06.00 (16) Day 4, 06.00 (67) Day 6, 06.00 (118) Day 7, 06.00 (145)	0.16 0.23 0.26 0.13 0.12
Patient 1948 Admit Day 1, 07.00 Discharge 2, 03.00 12 time points	Day 1, 17.00 (1)	1.1
Patient 1969 Admit Day 1, 20.00 Discharge Day 9, 18.00 202 time points	Day 2, 05.30 (12) Day 3, 06.00 (39) Day 5, 06.00 (89) Day 8, 06.00 (164)	0 0 0 0
Patient 2121 Admit Day 1, 13.00 Discharge Day 3, 01.00 39 time points	Day 1, 14.38 (3) Day 2, 06.00 (19)	0 0

Patient 2138A Admit Day 1, 00.00 Discharge Day 27, 13.35 692 time points	Day 1, 06.00 (8) Day 3, 05.37 (60) Day 5, 09.00 (110) Day 7, 06.00 (162) Day 10, 06.00 (244) Day 14, 06.00 (352) Day 16, 15.00 (415) Day 17, 06.21 (433) Day 19, 06.00 (490) Day 21, 06.00 (546) Day 24,06.00 (622)	0.3 0.18 0.14 0.15 0.05 0 0 0 0 0 0
Patient 2174 Admit Day 1, 12.00 Discharge Day 10, 14.00 232 time points	Day 1, 12.40 (2) Day 3, 06.00 (46) Day 8, 06.00 (173) Day 10, 06.00 (223)	0 0 0 0
Patient 2188 Admit Day 1, 05.00 Discharge Day 2, 08.00 33 time points	Day 1, 05.45 (2)	2.26
Patient 2189 Admit Day 1, 18.00 Discharge day 10, 08.00 233 time points	Day 2, 06.20 (16) Day 4, 06.00 (67) Day 4, 19.00 (81) Day 6, 06.00 (118) Day 9, 06.00 (191)	0.04 0.33 0.18 0.11 0.07
Patient 2284A Admit Day 1, 17.00 Discharge Day 14, 11.40 339 time points	Day 10, 06.00 (15) Day 4, 05.30 (89) Day 6, 06.00 (145) Day 6, 08.30 (148) Day 6, 19.00 (163) Day 8, 05.58 (205) Day 9, 02.30 (253) Day 10, 06.00 (282) Day 11, 06.00 (333)	0.04 0 0 0 0 0.14 0.14 0.22 2.57
Patient 2303A Day 1, 18.55 Discharge 27, 18.00 686 time points	Day 1, 19.00 (2) Day 4, 06.00 (67) Day 5, 06.00 (93) Day 6, 06.00 (120) Day 10, 06.00 (224) Day 11, 06.00 (249) Day 12, 06.00 (275) Day 13, 06.00 (304) Day 14, 05.30 (332) Day 15, 06.00 (363) Day 18, 06.00 (440) Day 20, 06.00 (490) Day 22, 06.00 (540) Day 25, 06.58 (623) Day 27, 06.05 (674)	0.66 50 20.8 14 3 2.71 1.66 2.12 1.52 0.83 0.61 0.48 0.26 0.09 0.06
Patient 2342A Admit Day 1, 21.00 Discharge Day 11, 14.00 237 time points	Day 2 06.00 (7) Day 2 19.00 (21) Day 3 04.54 (31) Day 6 05.30 (114) Day 8 06.00 (164) Day 10 05.00 (213)	0 0 0.04 0 0 0
Patient 2585 Admit Day 1, 18.00 Discharge Day 3, 20.00 58 time points	Day 3, 06.006 (43)	3.18
Patient 2644 Admit Day 1, 04.02 Discharge day 31, 14.00 794 time points	Day 1, 06.15 (3) Day 4, 06.00 (79) Day 6, 06.00 (129) Day 8, 06.00 (180) Day 11, 06.00 (270) Day 13, 06.00 (321) Day 14, 06.00 (350) Day 15, 06.00 (375) Day 18, 06.00 (451) Day 20, 04.00 (499) Day 22, 05.30 (551) Day 25, 06.00 (626) Day 27, 06.00 (676) Day 29, 06.00 (727)	0 0 0 0.14 0.07 0.31 0.1 0.04 0 0 0 0 0 0.07

The 34 patient "testing set"

Demographics	Actual time into admission (time point)	Troponin Rises
Patient 1536 Admit Day 1, 12.00 Discharge 66, 12.00 1684 time points	Day 2, 06.00 (20) Day 3, 06.34 (47) Day 5, 05.57 (98) Day 7, 06.00 (151) Day 11, 06.00 (264) Day 12, 06.00 (289) Day 17, 05.00 (415) Day 19, 06.00 (467) Day 21, 06.00 (521) Day 24, 06.00 (596) Day 28, 06.00 (697) Day 31, 06.00 (774) Day 33, 06.00 (825) Day 35, 06.00 (877) Day 38, 06.00 (953) Day 42, 06.00 (1058) Day 45, 06.00 (1138) Day 47, 06.00 (1189) Day 48, 06.00 (1214) Day 49, 06.00 (1241) Day 51, 06.00 (1291) Day 54, 06.00 (1373) Day 56, 06.00 (1423) Day 66, 05.00 (1677)	0.34 0.22 0.04 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Patient 1697 Admit Day 1, 08.00 Discharge Day 3, 11.59 59 time points	Day 1, 09.00 (2)	0
Patient 1748 Admit Day 1, 17.00 Discharge Day 20, 22.00 503 time points	Day 3, 06.00 (39) Day 3, 11.40 (46) Day 5, 06.00 (91) Day 10, 06.00 (224) Day 12, 06.00 (279) Day 15, 06.00 (358) Day 17, 05.44 (409) Day 19, 06.00 (461)	0.06 0.06 0.04 0 0 0 0 0
Patient 1757 Admit Day 1, 10.40 Discharge Day 2, 13.00 30 time points	Day 1, 10.40 (1) Day 2, 06.00 (22)	0 0
Patient 1774 Admit Day 1, 04.00 Discharge Day 26, 22.28 672 time points	Day 3, 06.00 (55) Day 5, 06.00 (105) Day 7, 06.00 (155) Day 14, 06.00 (335) Day 15, 05.50 (360) Day 16, 06.00 (389) Day 17, 05.37 (414) Day 19, 06.00 (465) Day 24, 05.30 (603) Day 26, 05.00 (653)	0.16 1.19 9.4 1.66 2.4 4.01 2.38 1.86 0.95 0.51
Patient 1822 Admit Day 1, 04.00 Discharge Day 37, 17.10 990 time points	Day 3, 05.00 (57) Day 4, 06.00 (92) Day 5, 06.00 (120) Day 8, 06.00 (204) Day 10, 06.00 (262) Day 12, 06.00 (315) Day 15, 06.00 (398) Day 20, 06.00 (532) Day 22, 06.00 (585) Day 24, 06.00 (636) Day 26, 06.00 (687) Day 26, 08.20 (690) Day 29, 06.00 (763)	0.06 0 0.09 0 0 0 0.05 0 0 0.05 0 0 0
Patient 1965 Admit Day 1, 13.00 Discharge Day 4, 20.00 87 time points	Day 2, 06.00 (20) Day 4, 06.00 (73)	0.18 0.95
Patient 2017 Admit Day 1, 19.00 Discharge Day 10, 14.00 261 time points	Day 1, 19.45 (3) Day 4, 06.00 (88) Day 6, 06.00 (139) Day 7, 06.00 (191) Day 8 14.30 (232)	0.12 14.5 5.49 4.26 2.61

Patient 2030 Admit Day 1, 15.00 Discharge Day 21, 10.16 528 time points	Day 2, 06.00 (17) Day 7, 06.00 (142) Day 11, 06.00 (244) Day 14, 06.00 (327) Day 16, 06.00 (376) Day 18, 06.00 (427) Day 21, 05.10 (503)	0.51 0.07 0 0 0 0 0
Patient 2158 Admit Day 1, 23.25 Discharge Day 9, 10.00 208 time points	Day 2 24/4/9 06.00 (9) Day 5 27/4/9 05.40 (97) Day 6 28/4/9 05.28 (129) Day 7 29/4/9 04.49 (156) Day 8 30/4/9 06.49 (185) Day 9 1/5/9 06.00 (204)	0 0 0 0 0 0
Patient 2265 Admit Day 1, 14.00 Discharge Day 45, 22.00 1134 time points	Day 4, 06.00 (72) Day 5, 06.00 (97) Day 5, 20.08 (113) Day 6, 06.00 (124) Day 7, 06.00 (150) Day 12, 06.00 (277) Day 13, 06.00 (302) Day 14, 06.00 (328) Day 15, 06.00 (356) Day 17, 06.00 (407) Day 19, 06.00 (457) Day 24, 05.00 (585) Day 26, 06.00 (636) Day 27, 06.00 (661) Day 31, 04.56 (761) Day 33, 06.00 (818) Day 35, 06.00 (869) Day 38, 07.00 (947) Day 39, 06.00 (971) Day 40, 06.00 (990) Day 42, 06.00 (1040) Day 45, 06.00 (1121)	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Patient 2313 Admit Day 1, 15.00 Discharge Day 30, 00.20 801 time points	Day 2, 06.00 (17) Day 4, 06.00 (67) Day 9, 05.29 (207) Day 11, 06.00 (259) Day 13, 06.00 (310) Day 16, 06.00 (391) Day 18, 06.00 (446) Day 20, 06.00 (499) Day 23, 09.57 (582) Day 24, 06.00 (603) Day 25, 06.00 (631) Day 27, 06.00 (690) Day 30, 06.00 (773)	0 0 0 0 0 0 0 0 0.07 0 0 0 0.09
Patient 2328 Admit Day 1, 04.00 Discharge Day 10, 08.00 247 time points	Day 1, 07.48 (5) Day 1, 14.00 (13) Day 2, 07.00 (31) Day 3, 06.00 (55) Day 5, 06.00 (110) Day 6, 10.55 (140) Day 7, 05.08 (161) Day 10, 05.25 (244)	0.61 1.23 4.37 2.31 1.77 1.48 1.24 0.66
Patient 2457 Admit Day 1, 23.00 Discharge Day 17, 16.00 332 time points	Day 3, 04.13 (60) Day 5, 06.16 (114) Day 7, 06.00 (164) Day 10, 06.00 (244) Day 12, 06.00 (297)	0.39 0.16 0.08 0 0
Patient 2607 Admit Day 1, 22.00 Discharge Day 9, 00.00 104 time points	Day 2, 01.00 (4) Day 4, 06.00 (60)	0 2.91
Patient 2660 Admit Day 1, 17.00 Discharge day 11, 12.00 253 time points	Day 2, 06.00 (16) Day 2, 16.30 (28) Day 3, 06.00 (42) Day 4, 06.00 (69) Day 7, 06.00 (145) Day 9, 06.00 (196) Day 11, 06.00 (246)	0 0.07 0.11 0.23 0 0 0
Patient 2698 Admit Day 1, 13.00 Discharge Day 3, 12.00 52 time points	Day 1, 21.50 (11) Day 3, 05.00 (44)	0.08 0.31

Patient 1684 Admit Day 1, 23.00 Discharge Day 3, 14.00 44 time points	Day 2, 00,15 (3) Day 2, 06.00 (9) Day 4, 06.00 (35)	0.48 1.85 4.26
Patient 1689 Admit Day 1, 23.00 Discharge Day 3, 09.00 37 time points	Day 3, 06.00 (34)	0.2
Patient 1720 Admit Day 1, 16.56 Discharge Day 19, 19.00 463 time points	Day 3, 09.00 (44) Day 4, 03.39 (64) Day 6, 06.00 (118) Day 16, 05.30 (375) Day 18, 06.00 (425)	2.35 1.25 0.57 0.05 0.05
Patient 1721 Admit Day 1, 02.00 Discharge Day 3, 20.00 73 time points	Day 2, 21.51 (48) Day 3, 06.18 (58)	0 0
Patient 1726 Admit Day 1, 14.00 Discharge 1 Day 2, 16.00 34 time points	Day 2, 07.06 (24)	0.07
Patient 1727 Admit Day 1, 13.00 Discharge Day 11, 04.00 243 time points	Day 2, 06.00 (22)	1.73
Patient 1750 Admit Day 1, 07.00 Discharge Day 5, 21.17 151 time points	Day 2, 06.00 (30) Day 4, 06.00 (101)	0 0.08
Patient 1807 Admit Day 1, 04.00 Discharge Day 30, 15.00 760 time points	Day 1, 04.56 (2) Day 2, 07.11 (32) Day 3, 11.10 (64) Day 4, 06.00 (83) Day 9, 05.30 (210) Day 13, 06.00 (311) Day 14, 05.42 (336) Day 18, 06.00 (439) Day 20, 06.00 (489) Day 23, 06.00 (573) Day 27, 06.00 (674) Day 30, 06.00 (749)	3.24 6.38 13.6 6.88 0.44 0.19 0.19 0.13 0.11 0.09 0.09 0.06
Patient 1818 Admit Day 1, 17.00 Discharge Day 5, 18.00 109 time points	Day 2, 17.14 (31) Day 3, 05.00 (43) Day 5, 06.00 (96)	0.44 0.27 0.07
Patient 1951 Admit Day 1 18.00 Discharge Day 21 20.00 220 time points	Day 14, 16.03 (45) Day 15, 06.00 (59) Day 17, 06.00 (107) Day 20, 06.00 (180)	0.06 0.12 1.78 0.45
Patient 2039 Admit Day 1, 21.00 Discharge Day 18, 06.00 434 time points	Day 1, 22.00 (3) Day 2, 06.00 (11) Day 3, 06.00 (36) Day 4, 10.30 (68) Day 5, 09.42 (92) Day 6, 06.00 (113) Day 7, 07.27 (140) Day 10, 06.00 (214) Day 13, 06.00 (303) Day 15, 06.00 (359) Day 17, 06.00 (409)	0.87 2.23 1.46 0.62 0.36 0.41 0.24 0.08 0.18 0.06 0
Patient 2231 Admit Day 1, 14.00 Discharge Day 11, 18.00 258 time points	Day 1, 14.00 (1) Day 4, 06.00 (69) Day 6, 06.00 (119) Day 8, 06.00 (169) Day 11, 06.00 (244)	0 0 0 0 0

Patient 2273 Admit Day 1, 14.00 Discharge Day 18, 14.47 440 time points	Day 1, 18.35 (7) Day 1, 20.20 (11) Day 2, 05.14 (25) Day 3, 06.00 (55) Day 3, 14.45 (66) Day 5, 06.00 (108) Day 8, 05.30 (183) Day 10, 00.00 (228) Day 10, 06.00 (234) Day 10, 10.00 (238) Day 12, 05.00 (283) Day 15, 06.00 (359)	0.12 0.16 0.2 0.14 0.13 0.08 0.14 0.09 0.1 1 0.04 0
Patient 2506 Admit Day 1, 11.00 Discharge Day 7, 19.00 181 time points	Day 2, 06.00 (31) Day 5, 06.00 (110)	2.59 0.48
Patient 2524 Admit Day 1, 06.00 Discharge Day 7, 14.00 168 time points	Day 2, 06.00 (27) Day 5, 05.08 (107) Day 7, 06.00 (159)	0.85 0.32 0.15
Patient **2547** Admit Day 1, 06.00 Discharge Day 4, 18.00 102 time points	Day 2, 09.15 (38) Day 4, 06.19 (88)	2 1.65
Patient 2554 Admit Day 1, 15.00 Discharge Day 5, 11.30 113 time points	Day 2, 06.00 (17) Day 5, 06.00 (108)	0.07 0.05

The rule base was initially run against the 17 patient training data set to see if there was an association between it firing and the presence of actual myocardial damage as evidenced by raised troponins. The results are shown in table 7-1.

Table 7-1: Results when the rule base is run on the 17 patient data training set

	Sequences of High Troponins	**Sequences of Negative Troponins**
Rule set "fires"	8 out of 12 (66.7%) i.e. True Positive	10 out of 14 (71.4%) i.e. False Positive
Rule set does not "fire"	4 out of 12 (33.3%) i.e. False Negative	4 out of 14 (28.6%) i.e. True Negative

After this initial analysis, and in an attempt to increase the true positive rate, a further rule base was produced to include derangements from category C on the ICU-PSS, thus making the original rule base more extensive. This extended rule base was run again on the 17 patient data set and the results shown in table 7-2.

Table 7-2: Results when the extended rule base is run against the 17 patient training data set

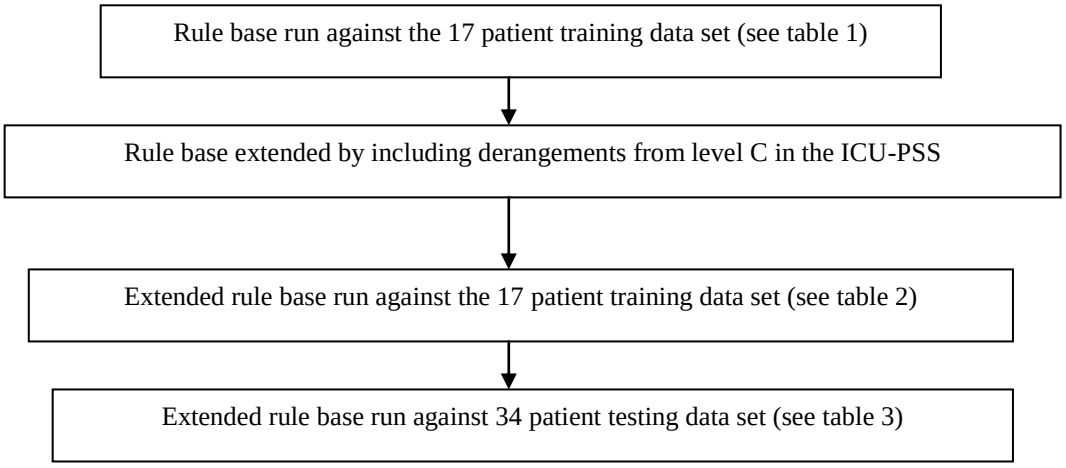
	Sequences of High Troponins	**Sequences of Negative Troponins**
Rule set "fires"	12 out of 12 (100%) i.e. True Positive	13 out of 14 (92.9%) i.e. False Positive
Rule set does not "fire"	0 out of 12 (0%) i.e. False Negative	1 out of 14 (7.1%) i.e. True Negative

The more extended rule base was then run against data sets from the remaining 34 patients (the testing data set described above). These 34 patient dataset contained 11,776 time points of patient data, including 198 troponin readings. From these data sets 33 sequences of positive troponin values and 22 sequences of negative troponin values were identified (see figure 10 above for the precise details of where these sequences occurred). Two out of the 22 sequences of negative troponins were removed as they occurred within the first 6 hours after admission. The results are shown in table 7-3 and the different rule bases are summarised in figure 7-11.

Table 7-3: Results when the more sophisticated rule base is run against the 34 patient testing set

	Sequences of High Troponins	**Sequences of Negative Troponins**
Rule set "fires"	25 out of 33 (75.8%) i.e. True Positive	16 out of 20 (80%) i.e. False Positive
Rule set does not "fire"	8 out of 33 (24.2%) i.e. False Negative	4 out of 20 (20%) i.e. True Negative

Figure 7-11: Sequence of testing and final modification to the rule base



7.4.2. Results - Prediction of myocardial damage

Table 7-4 shows the analysis of 72 hours prior to the first raised troponin in the 27 sequences of high troponins, and first negative troponin in the 19 sequences of negative troponins.

Table 7-4: Applying the extended rule base to the 72 hours prior to the 27 sequences of high and 19 sequences of negative troponins

	Sequences of High Troponins	**Sequences of Negative Troponins**
Extended rule base fires before the first troponin of the troponin sequence	14 out of 27 (51.9%) i.e. True Positive	6 out of 19 (31.6%) i.e. False Positive
Extended rule base does not fire before the first troponin of the troponin sequence	13 out of 27 (48.1%) i.e. False Negative	13 out of 19 (68.4%) i.e. True Negative

The first observation is that the false negative rate is high at 48.1%. That is to say, there is no firing of the extended rule base in the 72 hours prior to the first raised troponin in a sequence of high troponins. A further analysis showed that in 11 out of the 13 cases, the first troponin rise in a sequence of raised troponins occurred early in the patient's stay in ICU. Specifically, the mean time point for the first high troponin in the 13 sequences identified as true positives was 208, yet only the 33rd time point for the first raised troponin in sequences identified as false negatives. It is therefore possible that the physiological derangement leading to the myocardial damage was occurring prior to the patient's admission to ICU. It is also worth noting that in an analysis of the 13 sequences comprising the true negatives there were 5 instances where the first negative troponin in the sequence occurred within 72 hours of the patient's admission into ICU. Therefore there could have been a firing of the rule base before the patient's admission, which could mean that some of the true negatives are actually false positives. The results are presented again in table 7-5, with a reduced number of sequences reflecting those taken out of the analysis due to being early in the patient's admission to ICU. That is to say, the cardiovascular derangement could have happened before the patient's admission (if we take 72 hours as the time for decay of troponin).

Table 7-5: Results with sequences of troponin (high and negative) occurring early in the patient's admission to ICU removed

	Sequences of High Troponins	**Sequences of Negative Troponins**
Extended rule base fires before the first troponin of the troponin sequence	14 out of 16 (87.5%) i.e. True Positive	6 out of 14 (42.9%) i.e. False Positive
Extended rule base does not fire before the first troponin of the troponin sequence	2 out of 16 (12.5%) i.e. False Negative	8 out of 14 (57.1%) i.e. True Negative

7.5 Discussion.

This preliminary work leads to a hypothesis that it may possible to detect myocardial damage using physiological scoring alone based on commonly recorded ICU physiological parameters and drug infusion data, rather than traditional biomarkers. Further in a similar manner it may be possible to predict where myocardial damage is going to occur based on physiological disturbance alone.

However, in the evaluation of a new test, e.g. occurrence or non occurrence of an event with a binary predictor e.g. presence or absence (alternatively above or below a cut off point) it is common to summarise the data in a two by two contingency table [132] as shown in table 7-6.

Table 7-6: A two by two contingency table

	Outcome/Event present	Outcome/Event absent
Predictor / test positive	A (True positive)	B (False positive)
Predictor / test Negative	C (False negative)	D (True negative)

From this table a number of statistical descriptions can be defined (this shall be described in relation to this work in the context of prediction of troponin positive events).

A *true positive* (A) is where the rule set fires before the first positive troponin within a sequence of positive troponins.
A *false positive* (B) is where the rule set fires before the first negative troponin within a sequence of negative troponins.
A *false negative*(C) is where the rule set does not fire before the first positive troponin within a sequence of positive troponins.
A *true negative* (D) is where the rule set does not fire before the first negative troponin within a sequence of negative troponins.

From these basic definitions the *sensitivity* and *specificity* of the rules set to predict myocardial damage can be defined:

Sensitivity is the ability of the rule set to correctly fire before the first positive troponin within a sequence of positive troponins i.e. A/(A+C) in the above table. For example if the rule set were to have a 70% sensitivity then it would fire in 70% of instances before the

first positive troponin in a sequence of positive troponins but not fire in 30% of cases before the first positive troponin within a sequence of positive troponins.

Specificity is the ability of the rule set not to fire before the first negative troponin within a sequence of negative troponins i.e. D/(D +B). For example if the rule set had an 80% specificity it would not fire in 80% of instances before the first negative troponin within a sequence of negative troponins but would fire in 20% of cases before the first negative troponin within a sequence of negative troponins.

For this "test" although the ideal would be a high sensitivity and high specificity, in practice a high sensitivity is more important (initially) than a high specificity. This is because the practical implications of a false positive in a future system would be non-invasive bedside tests e.g. 12 lead ECG or an ECHO. What is perhaps more relevant to a clinician is the *positive predictive* and *negative predictive* value of a test.

Positive predictive value is where the rule set fires, what is the likelihood of this being before the first troponin in a sequence of positive troponins i.e. A / (A+B).

Negative predictive value is where the rule set does not fire, what is the likelihood of this being before the first negative troponin in a sequence of negative troponins i.e. D/ (C+D).

In this subsequent analysis the confidence intervals are now given along with the positive and negative predictive value of the rule base. (Two types of confidence intervals can be constructed around proportions (asymptotic and exact). Asymptotic assumes a normal approximation of the sampling distribution. When the sample size is small this normal assumption cannot be made and exact e.g. 95% confidence intervals are more appropriate.) Confidence intervals were calculated using Medcalc[133].

Tables 7-7 and 7-8 show the results when the rule base and extended rule base (to include derangements from level C in the scoring system) are run against the training data set (17 patients). This was to test if a positive troponin within a sequence was preceded or followed by a firing of these rule bases i.e. association of myocardial damagae with physiological disturbance. Given that the numbers in the 2x2 table are small and the confidence intervals large, nothing meaningful can be drawn from the data. The most

relevant set of results are in Table 7-9 as this shows the data when the extended rule set is run against the 34 patient testing data set.

7.5.1 Myocardial detection (association)

Table 7-7: Results when the rule base is run on the 17 patient training data set

	Sequence of High Troponins	**Sequence of Negative Troponins**
Rule set "fires"	8/12 (66.7%) i.e. True Positive	10/14 (71.4%) i.e. False Positive
Rule set does not "fire"	4/12 (33.3%) i.e. False Negative	4/14 (28.6%) i.e. False Negative
Sensitivity	66.7%	95% CI: 35% to 89.9%
Specificity	28.6%	95% CI: 8.6% to 58.1%
Positive predictive value	44.4%	95% CI: 21.6% to 69.2%
Negative predictive value	50%	95% CI: 16% to 84%

Table 7-8: Results when the extended rule base is run against the 17 patient training data set

	Sequence of High Troponins	**Sequence of Negative Troponins**
Rule set "fires"	12/12 (100%) i.e. True Positive	13/14 (92.9%) i.e. False Positive
Rule set does not "fire"	0/12 (0%) i.e. False Negative	1/14 (7.1%) i.e. False Negative
Sensitivity	100%	95% CI: 73.4% to 100%
Specificity	7.1%	95% CI: 1.19% to 33.9%
Positive predictive value	48%	95% CI: 27.8% to 68.7%
Negative predictive value	100%	95% CI: 16.6% to 100%

Table 7-9: Results when the more sophisticated rule base is run against the 34 patient testing set

	Sequence of High Troponins	**Sequence of Negative Troponins**
Rule set "fires"	25/33 (75.8%) i.e. True Positive	16/20 (80%) i.e. False Positive
Rule set does not "fire"	8/33 (24.2%) i.e. False Negative	4/20 (20%) i.e. False Negative
Sensitivity	75.8%	95% CI: 57.7% to 88.9%
Specificity	20%	95% CI: 5.9% to 43.7%
Positive predictive value	61%	95% CI: 44.5% to 75.8%
Negative predictive value	33.3%	95% CI: 10.1% to 65%

There is a suggestion from the data that the sensitivity of the rule set is moderately high at 75.7% but in the context of moderately wide confidence intervals. The specificity and negative predictive value of the rule set is low. The confidence intervals of the positive predictive value are so wide that no weight can be put on the figure of 61%.

A similar analysis is shown now for the rule base and myocardial prediction.
Table 7-10 shows the results to ascertain if the extended rule set fires (or not) in the 72 hours before a rise in the first troponin within a sequence of troponins (or the first negative troponin within a sequence of negative troponins). Again the confidence intervals are wide so nothing meaningful can be drawn from the results.

Table 7-11 shows the results when troponins (negative and positive) occurring early within the patient stay are removed (as any cardiovascular disturbance causing them will not necessarily be detected by the rule set). This had the effect of reducing the number of sequences of high troponins to 16 and negative troponins to 14.

7.5.2 Myocardial prediction (causation)

Table 7-10: Applying the extended rule base to the 72 hours prior to the 27 sequences of high and 19 sequences of negative troponins

	Sequence of High Troponins	**Sequence of Negative Troponins**
Rule set "fires"	14/27 (51.8%) i.e. True Positive	6/19 (31.6%) i.e. False Positive
Rule set does not "fire"	13/27 (48.1%) i.e. False Negative	13/19 (68.4%) i.e. False Negative
Sensitivity	51.8%	95% CI: 32% to 71.3%
Specificity	68.4%	95% CI: 43.5% to 87.4%
Positive predictive value	70%	95% CI: 45.7% to 88%
Negative predictive value	50%	95% CI: 29.9% to 70.1%

Table 7-11: Results with sequences of troponins (high and negative) occurring early in the patient's admission to ICU removed

	Sequence of High Troponins	**Sequence of Negative Troponins**
Rule set "fires"	14/16 (87.5%) i.e. True Positive	6/14 (42.9%) i.e. False Positive
Rule set does not "fire"	2/16 (12.5%) i.e. False Negative	8/14 (57.1%) i.e. False Negative
Sensitivity	87.5%	95% CI: 61.6% to 98.1%
Specificity	57.1%	95% CI: 28.9% to 82.2%
Positive predictive value	70%	95% CI: 45.7% to 88%
Negative predictive value	80%	95% CI: 44.4% to 96.9%

As might be expected from the small numbers the confidence intervals are sufficiently wide so that no meaningful conclusions can be drawn. There is a perhaps however a signal towards sensitivity (87.5% with confidence intervals 61.6% to 98.1%).

The rule base from the ICU-PSS underpinning the work, is distilled from the collective experience of three senior intensive care clinicians, and validated by 10 ICU consultants from two other hospitals. Electronic patient records in intensive care are a relatively new phenomenon. Analysing data on this scale would have been impossible using traditional paper records. For this exploratory study, over three thousand hours of data were examined. This required the use of various sophisticated computing tools, because manual manipulation of these data was impracticable. For that reason, the storage, processing and

manipulation of data in this way to examine a clinical problem has not previously been undertaken. Other groups have tried to detect myocardial damage automatically with ECG analysis using neural networks with some success[134]. Our system has the advantage of using physiological disturbance alone, and of potentially being able to detect impending damage before ECG abnormalities occur.

For myocardial association (detection) the false negative rate (i.e. the final rule set does not fire in certain cases when myocardial damage has subsequently occurred.) is moderately high. There are a number of possibilities. In this study I have used hourly data points, as currently physiological values have to be entered manually into the electronic patient record, then verified by the nurse at the bed space. Human nature being what it is, if there have been fluctuations in the parameter around the time of recording, there is a danger that the less extreme variation will be recorded. Also, there is a risk of missing some brief but significant physiological disturbance. It possible, but unlikely, that a patient could have 50 minutes of profound hypotension occurring between two hourly time points. This extreme disturbance would not be "seen" by the system. Increasing the frequency of observations, increasing the number of clinicians, and using more patient data sets will allow further refinement of the model in order to reduce the false negative rate.

An alternative explanation for the false negatives is that, in a subset of ICU patients with sepsis, the supply / demand imbalance characteristic of type II damage may be only part of the reason for myocardial injury. There is evidence, using coronary sinus thermodilution catheters, that coronary artery blood flow is not significantly different between septic and non septic ICU patients when their heart rate is less than 100 beats per minute, and may be greater in the septic group when the heart rate exceeds 100 beats per minute. Microvascular thrombi formation in the absence of overt physiological disturbance may play a role[135.] All of these factors may be exacerbated by the myocardial depressant cytokines (interleukins-2, 4, 6, 8, 10 and tumour necrosis factor alpha) characteristic of severe sepsis[136.] If such microvascular thrombi leading to myocardial damage occur in the absence of physiological disturbance, the system will not detect these events.

There is a high false positive rate. Not all physiological derangement which causes the rule base to trigger will result in myocardial damage in every individual. It may be that these are patients with no underlying flow limiting coronary artery disease who are able to cope with more extreme physiological stressors. Ultimately, however, a high false positive rate is less concerning, as the rule base firing will be a prompt to simple and non-invasive

investigation. The frequency of false positives may fall with further experience and refinement of the system.

In terms of myocardial prediction, the patient in Intensive Care is often sedated, and may not display typical symptoms of myocardial ischaemia. The current literature does not suggest any effective treatments for type II myocardial damage once it has occurred in the intensive care population[137]. Moreover, it is now well documented that the occurrence of myocardial damage does considerably increase predicted mortality and the length of ICU stay [138,139].

The results from the prediction study are encouraging. However the false positive rate only increases from 51.9% to 87.5% once the sequences of troponin (high and negative) occurring early in the patient's admission to ICU are removed. However it is my aim that any future system utilising the rule base could be used in areas other than ICU. For example, there is no reason why a future system could not be used in a high dependency area or acute medical receiving ward, and perhaps start to pick up the characteristic physiological disturbances leading to type II myocardial damage.

This study had several weaknesses that could be addressed in future work. Although 52 patient data sets were used, the number of positive and negative sequences was small (27 sequences of positive and 19 sequences of negative troponins). In statistical terms this makes it more likely that the confidence intervals will be wider. Further, more than one sequence could have come from the same patient and it is possible that it is the same continued myocardial supply/demand imbalance occurring at different times causing ongoing damage rather than discrete events. With routine 3 times per week troponin sampling it is possible that a troponin that was only very mildly elevated had decayed and was undetected when sampled even though the rule set had correctly "fired." Nursing staff at Glasgow Royal either record manually or verify physiological data automatically downloaded into CareVue at hourly time points. It is possible that there could be profound periods of physiological disturbance occurring between the points that the nursing staff record that the rule set does not "see." The combination of rules chosen and their duration is only one of a large number of other combinations and variations. It may be that other combinations with different durations model better.

The work was also undertaken retrospectively. Future work would be done prospectively using a much larger cohort of patients. Greater numbers would have the advantage of lack

of reliance on more than one event within the same patient. Greater troponin sampling frequency would identify more clearly if the rise and decay of troponin in Intensive Care patients is as per the ideal figure seen in textbooks. Renal impairment is common and the decay phase in particular may be prolonged. Further, increased sampling would decrease the chance of missing a small troponin rise that has already decayed by the time it is sampled under the current regime. The rule set could run on auto charted data from CareVue and be calculated automatically many times an hour. This would decrease the chance of profound physiological disturbance being missed between two hourly time points currently recorded. It would be interesting to model different combinations of rules and of different duration to ascertain if they give a better fit. Finally it would be useful to measure the new generation of highly sensitive troponins in addition to the current troponin test to see if the rules are detecting or predicting a troponin rise, but it is not being seen due to the sensitivity of the troponin test itself.

7.5.3. Future work

My future aim is to develop a bedside system which, using data from routine monitoring attached to the patient, will detect characteristic patterns of physiological derangement which lead to myocardial damage. Having demonstrated that it is possible to detect myocardial damage using physiological disturbance alone, I will work to refine this theoretical model, and incorporate it into a real time bedside system which alerts the health care professional to impending myocardial damage. This will initially prompt simple bedside tests, e.g. a 12 lead ECG, bedside echocardiography or a troponin measurement.

This could, at first with high fidelity simulation and then with further study, lead to interventions at an earlier stage, e.g. increase in FiO2, fluid boluses, beta blockade and so on. The type of intervention would depend on the type of derangement and would be occurring before damage has occurred. The system could also be used in different clinical environments, e.g. high dependency or other areas with a suitable level of monitoring.

It is now clear that a raised troponin is an independent risk factor for poorer ICU, hospital and long term outcome. It is also clear that, once it has occurred, there is little evidence on how to treat a patient or indeed how to reduce the mortality risk. Therefore a system which can predict impending damage and trigger further intervention may be clinically useful. Indeed the detection and timeliness of intervention in these areas may ultimately affect outcome.

7.6. Conclusion

These preliminary studies lead to the hypothesis that it may be possible to detect myocardial damage and predict its occurrence, based on physiological disturbance alone. This will require further study based on analysis of prospectively collected and much larger data sets

7.7. Acknowledgements

I co-conceived the hypothesis tested in this chapter. The data sets for analysis were prepared by our computing science colleagues. In the initial phase both Prof. Kinsella and myself analysed 3664 predominantly hourly time points of physiological and intervention data to characterise the types of physiological disturbance associated with myocardial damage. In the second, more detailed phase, I analysed a further 6827 predominantly hourly time points of data, and in discussion with Prof Kinsella we agreed what rules occurring in what duration, from the ICU-PSS, would constitute the rule set which would be used to detect and predict myocardial damage. I agreed with the computing scientists what characterised a sequence of troponins based on the natural decay in blood. The rule set was applied on my behalf to the data sets by the computing scientists.

Chapter 8: Patient Outcomes: Associations between mean ICU-PSS score, troponin rise during admission and outcome

8.1. Abstract

8.1.1. Introduction

The APACHE II score combines physiological derangement with diagnostic information to reach an APACHE II predicted mortality. I hypothesised that there is a correlation between the patient's overall cardiovascular state (as measured by the ICU-PSS) and patient outcome. Further, as myocardial damage in the critically ill is caused by an imbalance between myocardial oxygen supply and demand,
there should also be a correlation between the size of a troponin rise (representing a greater and more sustained physiological derangement) and outcome.

8.1.2. Methods

In a first study, a data base was created from ward watcher of 54 patients. APACHE II score, predicted mortality, whether medical or surgical diagnosis, and outcome were recorded. Physiological data from this cohort was extracted from CareVue. The hourly ICU-PSS score was converted to a numeric scale (A=1 to E=5) and a mean score of the ICU-PSS was calculated for periods during the stay. In a second study, data from 100 consecutive ICU admissions from July to October 2009 was extracted from the Ward Watcher System and the CareVue. These included APACHE II score, APACHE II predicted mortality, size of first troponin rise, day of first troponin rise, highest troponin rise, date of highest troponin rise, and patient outcome (alive / dead). The troponins were grouped into 4 ranges (<0.04, 0.04-0.19, 0.20-1.99, ≥2.0) and mean APACHE II score, mean APACHE II predicted mortality, and ICU mortality calculated.

8.1.3. Results

In the first study, 26/54 patients had a medical and 28/54 a surgical diagnosis. 17/26 medical and 8/28 surgical patients died. Mean values for survivors and non survivors in the different groups at different time periods were calculated. In this preliminary work the computer programme used to calculate these values could not, at this stage, apply appropriate statistical tests.
For all patients: 29 alive, 25 dead. Mean score Day 1, 3.79 (alive)/4.28(dead). Mean Score day 1-2, 3.79 (alive)/4.29 (dead). Mean score total stay, 3.12 (alive)/4.23 dead.

For patients with a medical diagnosis: 9 alive, 17 dead. Mean score Day 1, 3.39 (alive)/4.32(dead). Mean Score day 1-2, 3.44 (alive)/4.29 (dead). Mean score total stay, 2.76 (alive)/4.28 dead.
For patients with a surgical diagnosis: 20 alive, 8 dead. Mean score Day 1, 3.98 (alive)/4.18(dead). Mean Score day 1-2, 3.95 (alive)/4.30 (dead). Mean score total stay, 3.29 (alive)/4.13 dead.

In the second study, 23/100 patients were excluded from analysis (no troponin data or excluded from APACHE II scoring). The mean APACHE II predicted mortality, and actual ICU mortality for the different troponin ranges were respectively <0.04 (24%, 13.3%, n=30), 0.04-0.19 (42%, 13.3%, n=15), 0.2-1.99 (38.5%, 22.7%, n=22) and ≥2.0 (50.7%, 40%, n=10).

8.1.4. Conclusion

This preliminary work leads to the hypothesis that there is a correlation between patient's mean cardiovascular scores (as captured by the ICU-PSS) and, in certain groups, outcome. There is an association between level of troponin rise and Intensive Care Mortality.

8.2. Introduction

A patient's APACHE II score as calculated in the first 24 hours is combined with an APACHE II diagnosis to give an APACHE II predicted mortality. This is the gold standard prediction of patient outcome currently in use in Scotland. The higher the APACHE II score, the higher the predicted mortality. Unit predicted mortality can be calculated and compared with actual mortality to give a standardised mortality ratio (SMR) [11]

I postulated therefore that there is a correlation between the patient's overall cardiovascular state (as measured by the ICU-PSS) and patient outcome. Similarly as most troponin rises in Intensive Care are caused by a myocardial supply / demand imbalance, I postulated that there should also be a correlation between the size of a troponin rise (representing a greater and more sustained physiological derangement) and outcome.

8.3. Methods

8.3.1. Possible correlation between mean ICU-PSS score and outcome

I prepared a data set from 54 anonymised patients admitted to Glasgow Royal Infirmary Intensive Care Unit. The data collected included their APACHE II Score, APACHE II predicted mortality, reason for admission to the unit (APACHE III definition), corresponding APACHE II diagnosis and their outcome (dead or alive). These data were extracted from the Ward Watcher System. The full data set can be seen in *Appendix XV*. An extract is shown in table 8-1.

Table 8-1: An extract from the 54 patient data set

Patient-ID	Start/Fin	Outcome	APACHE II	Predicted Mortality	Med Diag.
1536	13	Alive	29	77.2	Medical
1667	1697	Dead	33	84.5	Medical
1689	1861	Alive	22	58.9	Medical
1695	18614	Dead	31	75.4	Medical
1697	1898	Alive	24	35.5	Medical
1711	18654	Dead	20	35.5	Medical
1720	2214	Alive	28	63.7	Surgical
1721	2677	Alive	18	44.4	Surgical
1726	2750	Alive	17	26.2	Surgical
1727	2784	Dead	29	69.6	Medical
1742	18666	Alive	16	23.3	Surgical
1748	3027	Alive	23	62.3	Surgical
1750	3530	Dead	18	29.1	Medical
1757	3681	Alive	22	14	Medical

Physiological data from the Electronic Patient Record of each of these patients was then analysed. The predominantly hourly time point ICU-PSS score (A-E) was converted to a numerical score with A=1 point, B=2 points and so on. The mean score of these time points during various parts of the patient stay was then calculated. These were a mean score for the first day in ICU, day 1-2 and the total stay. These conversions and calculations would be extremely difficult manually and were done automatically by a computer workbench, the "I-predictor" created by our computing science colleagues. This programme also allowed examination of correlations between ICU-PSS scores and the demographic data base I had created [140].

The patients were separated into those with either a medical or surgical diagnosis and as to whether their outcome was survival or death. Mean scores were then calculated for these groups as a whole for day 1, day 1-2 and total stay in Intensive Care.

8.3.2. Possible correlation between size of troponin rise and outcome

I audited 100 consecutive ICU admissions from July to October 2009. Data were extracted from the Ward Watcher System and the Electronic Patient Record. This work was made possible, as from 2009, troponins have been recorded regularly as part of routine clinical care, as well as when there is a specific clinical indication. The parameters were, date of admission to ICU, date of discharge from ICU, length of stay (days), age, sex, APACHE II score, APACHE II predicted mortality, size of first troponin rise, day of first troponin rise, highest troponin rise, date of highest troponin rise, and patient outcome (alive / dead). The full data set can be viewed in *Appendix XVI*. An extract is shown in table 8-2 below.

Table 8-2: Extract from data set for analysis between troponin rises and outcome

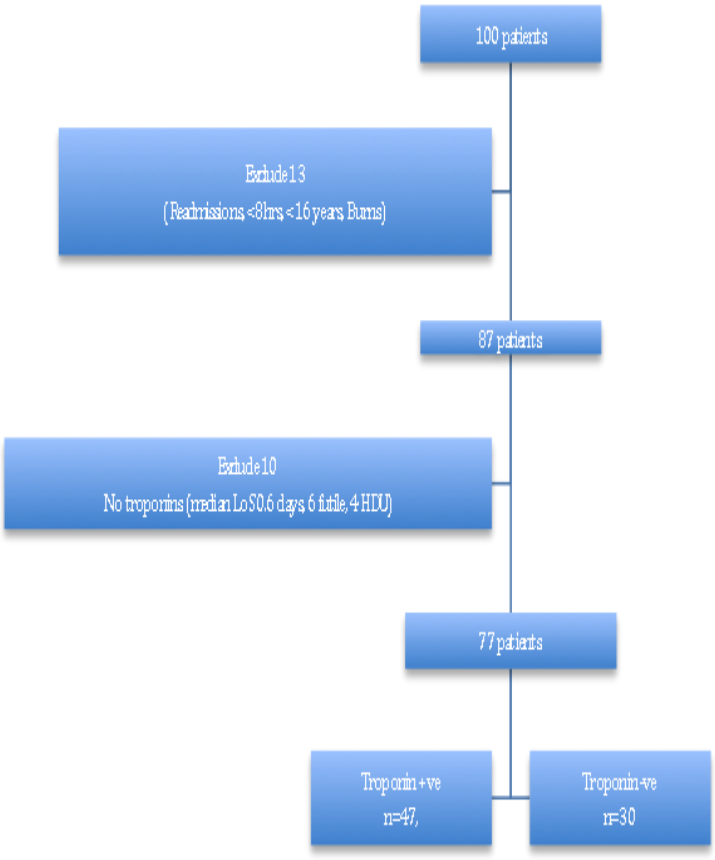
Date of admission	Date of discharge	Length of stay (days)	Age	M/F	Apache II	Predicted mortality	1st Troponin	Day of 1st troponin rise	Highest troponin	Day of highest troponin rise	Alive (1)/ Dead (0)
12/10/2009	15/10/2009	2.9	70	F	28	59.4	5.01	1	6.54	2	0
10/10/2009	14/10/2009	4	58	F	Readmission	Readmission	3.21	1	3.21	1	1
09/10/2009	14/10/2009	4.5	76	M	Readmission	Readmission	0.13	1	0.13	1	1
08/10/2009	09/10/2009	0.9	33	M	11	6.3	<0.04	Y	<0.04	Y	1
08/10/2009	12/10/2009	3.6	35	F	19	27	0.27	2	0.27	2	1
08/10/2009	14/10/2009	6.1	43	F	33	75.6	<0.04	Y	<0.04	Y	0
06/10/2009	08/10/2009	1.5	77	F	16	23.3	23.8	1	23.8	1	1

In the table a troponin value of <0.04 indicates a troponin sampled on routine clinical testing but in which there is no rise. In such circumstances a "Y" in the day of first

troponin, and day highest troponin rise columns, merely indicates that a troponin was recorded and is negative i.e. there is no day of first rise.

Of the 100 patients on whom data were collected, 13 were excluded as there was no APACHE II data (readmissions, length of stay < 8 hours, age < 16 years or burns). A further 10 patients were excluded as there was no troponin data available. The mean length of stay in this group was 0.6 days and they were not admitted long enough to have a sample collected for routine testing. 77 patients were analysed in total. This is represented in figure 8-1.

Figure 8-1: Flow chart summarising the 23 patients excluded from the analysis



8.4. Results

8.4.1. Correlation between mean ICU-PSS and outcome

54 patients were analysed. 28/54 patients had a surgical and 26/54 had a medical diagnosis. Table 8-3 shows the raw data, with mean scores for medical, surgical and all patients in both outcome alive and outcome dead categories.

Table 8-3: Overall outcome per diagnostic category

Time Period	**All patients**		**Medical diagnosis**		**Surgical diagnosis**	
	Alive	Dead	Alive	Dead	Alive	Dead
Number of patients	29	25	9	17	20	8
Day 1 Mean ICU-PSS score	3.79	4.28	3.39	4.32	3.98	4.18
Day 1-2 Mean ICU-PSS score	3.79	4.29	3.44	4.29	3.95	4.30
Total stay in ICU Mean ICU-PSS score	3.12	4.23	2.76	4.28	3.29	4.13

8.4.2. Correlation between troponin rise and outcome

A positive troponin I was greater than 0.04 micrograms / litre. For the purposes of the analysis I arbitrarily assigned those patients who had no troponin elevation or those who had a troponin rise into 4 categories. These were a troponin of less than 0.04 micrograms / litre (no rise), 0.04-0.19 micrograms / litre (low), 0.2-1.99 micrograms / litre (medium) and greater or equal to 2.00 micrograms / litre (high).Table 8-4 shows the results of assigning the 77 patients included in the analysis to these categories.

Table 8-4: Analysis of the 77 patients by troponin range

	Troponin <0.04	**Troponin 0.04-0.19**	**Troponin 0.20-1.99**	**Troponin ≥2.0**
Number of patients	30	15	22	10
Mean APACHE II	16	20.9	22	24.8
Mean APACHE II predicted hospital mortality (%)	24	42	38.5	50.7
Actual ICU mortality (%)	**13.3**	**13.3**	**22.7**	**40**

The range of troponin rises was 0.04-23.8 micrograms/litre. The low troponin group had an ICU mortality of 13.3%, the medium group 22.7% and the high group 40%. The 3 categories of troponin positive patients were more ill than the troponin negative patients, having both higher APACHE II score and predicted mortalities. From these results, there is a correlation between an increasing troponin rise and ICU mortality.

8.5. Discussion

8.5.1. ICU-PSS and patient outcome

In this experiment the predominantly hourly time points of the patient's stay was scored using the ICU-PSS on the A-E scale. This was converted to a numerical value, A=1, B=2 etc. This approach has simplicity but has a number of drawbacks. A non-linear categorical scale is being converted into a numerical score from which a mean is calculated. The data was subject to the extrapolation and nonsense values algorithms described earlier in the thesis but these have their limitations. Clearly if a "mean" is being calculated the A to E score has to be accurate before conversion to a numerical value. Two other checks before using this programme have to be undertaken. The users have to be satisfied themselves that the data is satisfactory by manually checking for significant missing values. Secondly the programme can itself flag to the user where there is missing data.

Unfortunately at the stage of development of the system, it was only able to convert the categorical scale into a numerical value and calculate a mean. It was not able to calculate standard deviations or interquartile ranges. This will clearly be important for future development. The table below shows the raw data, with mean scores for medical, surgical and all patients in both outcome alive and outcome dead categories.

As can be seen there is a suggestion that the mean scores are worse in the non-survivors than survivors. However, no appropriate statistical tests have been applied. This system does not take into account the different lengths of stay of the various patients. Future work could focus on this problem in a different manner by examining the different proportions of the patient's stay spent in different categories of the score. This would avoid the difficulties above of converting ordinal to numerical data.

Therefore, in this preliminary work there may be a correlation between mean cardiovascular scores (as captured by the ICU-PSS) and, in certain groups their outcome. Further work with appropriate statistical tests will be required to confirm or refute this hypothesis. This would be of interest as the ICU-PSS is a purely physiologically based score and is diagnosis independent. Compare this to the APACHE system. This requires diagnostic criteria to be combined with an APACHE score to obtain a predicted mortality, and therefore has its own inherent problems. There can be inconsistency and inaccuracy in applying diagnostic criteria, as the Ward Watcher data are often entered by inexperienced junior staff. There can also be inaccuracy and inconsistency in entering the physiological data. It a study by Goldhill *et al* [141], checking points assigned for 8 physiological

variables, applying strict APACHE II criteria 20.6%, of these points were higher and 6.7% of the points lower than originally entered into the study ICU's data base. This had the effect of raising that ICU's predicted mortality from 24.8 to 27.8%. Further, in the APACHE II score, the difficulty in determining inspired oxygen concentration in spontaneously breathing patients has an impact on the acute physiology score component [142].

There would be several advantages of the ICU-PSS in the future as a predictor of mortality. There are none of the inconsistencies in data entry or ambiguity in the components of the score, it is diagnosis independent, it could be calculated automatically and can be calculated at various points throughout the patient stay (potentially giving a more accurate prediction of outcome with time).

8.5.2. Troponin rise in ICU and outcome

The results demonstrate an increasing ICU mortality as the ranges of troponin also increase. Previous studies have examined this phenomenon and found similar trends, but some of these were in purely medical ICUs [80]. These results confirm that the trend in mortality holds true in a mixed Scottish ICU with a higher than average APACHE II score.

The recent universal definition of myocardial infarction has helped clarify thinking on this issue [76]. Critically ill patients have disordered cardiovascular function characterised by different combinations of hypo- or hypertension, tachy- or bradycardia and a high incidence of coexisting cardiovascular disease. A rise in cardiovascular biomarkers is taken to indicate myocardial damage, but the elevated levels found in multiple organ failure, sepsis and burns do not necessarily indicate the development of a myocardial infarction, due to lack of ECG changes or a specific wall motion abnormality on echocardiography.

Although I found no change in ICU mortality in patients in the range 0.04-0.19 micrograms / litre versus patients with no troponin rise, a recent much larger study of 663 patients by Reynolds [76] showed a trend towards lower odds of hospital survival in patients with minor elevation in the range 0.05-0.12 micrograms / litre (again using Troponin-I). The study was undertaken in a mixed medical and surgical unit where patients have troponin sampled daily. Interestingly 52% of their patients had a troponin rise there whilst in Intensive Care, which is similar to the rate I found in the ICU of Glasgow Royal Infirmary. Further, in a recent study conducted in the Beth Israel Deaconess Medical Centre, Velasquez *et al* examined a cohort of 3250 patients who had one or more troponins

measured during their admission [139].1219 of those had at least one positive result. They demonstrated an increase in all cause mortality at 1000 days in patients with minor troponin rises (0.01 to 0.1ng/ml in their assay).

In future work, I will extend the data base and focus on minor troponin elevation ranges only to ascertain if the effect demonstrated in other studies holds true for a mixed Scottish ICU population. Small troponin rises are likely to be overlooked by most clinicians as there is no evidence how to treat them in the absence of overt coronary artery disease or on the mechanism of their effect on mortality.

8.6. Conclusions

This preliminary work leads to the hypothesis that there is a correlation between patient's mean cardiovascular scores (as captured by the ICU-PSS) and, in certain groups outcome. This will have to be investigated further using appropriate statistical tests applied to larger datasets. There is an association between level of troponin rise and Intensive Care Mortality.

8.7. Acknowledgements

I prepared the data base for the ICU-PSS outcome study and the troponin outcome study. I co-conceived the idea for the ICU-PSS outcome study with Profs. John Kinsella and Derek Sleeman. I co-conceived the idea for the troponin outcome study with Prof. John Kinsella. The database I created for the ICU-PSS study was interrogated by our computing science study to ascertain the relationship between mean ICU-PSS and outcome in different groups of patients.

Chapter 9: Final Discussion and future direction

The initial aim of this thesis was to ascertain if haemodialysis in Intensive Care patients is a haemodynamically unstable therapy. From an initial and simple experiment, it became clear that a more sophisticated scoring system would have to be designed to answer this question. This resulted in the development of a novel quantitative score for cardiovascular instability. Although this approach was shown to have some merit, it became apparent that, to overcome the deficiencies of this score and of other currently available scores in Intensive Care, a different strategy was required. This led to the development and first stages of development of a novel quantitative score underpinned by a sophisticated physiological rule base, to summarise the overall state of a patient. This was a major part of the work of this thesis.

The qualitative score has the advantage over the quantitative score that it captures the expertise of several clinicians. It was shown to be clinically credible in a series of studies with clinicians from different hospitals not involved in the development of the score. This score comes closer to an ideal score. It is calculated from routinely collected data and takes into account the amount of physiological and pharmacological support a patient is receiving (treatment effect). It could theoretically be calculated an infinite number of times, be automated, and displayed in real time. Early work suggests that it may discriminate outcomes in a diagnosis independent fashion. This could be of importance, as currently available scores such as APACHE II exclude certain diagnostic groups, e.g. burns. Further, it can be calculated from time zero and patients are not excluded from using the ICU-PSS because they have been admitted for less than 8 hours. Future work with the score will look to establishing if combining the score with, e.g., the APACHE II diagnostic codes will improve prediction of patient outcomes. The score's positive predictive power may also increase over time after admission, with repeated calculation.

Using physiological data on the scale in this body of work presented enormous challenges which for all practical purposes would have been impossible with traditional paper based records. Our collaborators were computing science colleagues who had very little knowledge of critical care and its terminology. A lot of very basic terms which are obvious to practising clinicians were alien to them and a lot of time was spent educating them in the "domain." For any final bedside score the issue of missing data or consistently missing parameters will have to be dealt with. Rules were introduced so that parameters were converted and presented in a consistent manner. Rules were also introduced to deal with single missing values in data sets and to exclude "nonsense" values. This will have to be

taken further in the future. It would also be useful to have an increased sampling frequency. The limit in this research was the hourly frequency of recordings currently undertaken by nursing staff. In the future it may be possible to make use of the auto chart function of CareVue. This will be particularly important in the prediction of type II myocardial damage.

Quantitative score - The weakness of this score is that the parameters and weightings were derived by a single researcher. Using this work as a starting point the score will be refined by conducting a brainstorming session with a group of clinicians. During this session they will be asked to define regularly recorded parameters and appropriate ranges. These will then be incorporated into a new set of rules and tested on real patient data sets to determine if they show trends in improvements and deteriorations. The major difficulty in developing the scores has been a lack of a standard reference with which to compare. In the absence of this the new score could be tested in a series of clinical scenarios with a separate group of clinicians not involved in its development. This would assess whether the score is clinically credible to them. At this stage any final refinements to the weightings of parameters could be made.

Qualitative score - It would be useful to expand the number of clinicians scoring data sets and undergoing sessions with INSIGHT to expose inconsistencies. Since the initial work was done with the INSIGHT system, a comprehensive user manual has been written and it can now be used by the individual themselves independently from the computing scientists. This will help to remove any bias. It would be interesting to use other methodologies to define the rules in an A to E score. This could include facilitating a round table discussion of "experts" in the field. This method would have better face and content validity but would be more difficult to test there is inconsistency between an abstract discussion and clinical reality. A Delphi exercise could also be done but this may require several rounds to refine rules. The anonymity of responses would help reduce bias as the responder would not be directly influenced by a peer. Experts could also be shown mock scenarios of an accelerated patient stay in Intensive Care and asked to summarise why they are improving or deteriorating and describe why. This would help eliminate the problem of describing instability in the abstract.

Validation of the qualitative score – Only a series of discrimination experiments as the first stage of full validation have been undertaken thus far. To recap a suggested scheme for completing this process: For face validation clinicians will be shown examples of changes

within the score and no change within the score. They will be given a simultaneous clinical commentary about the patient's state and asked whether they think "on the face of it" that changes within the score reflect what is happening clinically i.e. does it appear to capture clinical improvement or deterioration. For content validation parameters which comprise the final ICU-PSS score will be shown to a group of experienced clinicians. They will be asked whether they feel that the parameters chosen to capture instability reflect what they themselves would have chosen if they had been designing a score. In a sense the score already has some indirect content validity given that when forming their rule base to score the data sets in its construction, two clinicians other than myself chose the same (although obvious) markers of instability i.e. heart rate, mean arterial pressure, inspired oxygen concentration, oxygen saturation and inotrope requirements.

As above discriminant validity experiments have already been undertaken. These successfully showed that when the score increases or decreases clinicians can (in the absence of clinical information) detect improvement or deterioration when the score changes by one or two steps. To complement this discriminant validity experiment a convergent validity experiment would be useful i.e. when there is no change in the score the clinicians no not detect a difference. Clinicians will be shown random examples of no change i.e. A to A, B to B etc. and asked whether the patient has deteriorated, improved or their physiological state is unchanged. Clinical history and trending information will be important to again help overcome the situation where parameters in the period of interest in the data shown are very close to a boundary. The score will also have to be shown to be reliable. If all the relevant data is present, collected properly from working equipment and processed appropriately by a computer algorithm then for given combinations of data there should be a consistent and reliable output. In terms of the clinicians a further experiment will be conducted to assess if they are consistent and reliable in their assessment. This could be done by showing a (large) series of lines of data, possibly with parameters around the middle of ranges. The clinician would then be asked to say whether the lines of data represented A (stable) through to (E) unstable. The same lines of data would be shown on more than one occasion to ascertain how reliable the clinician was with their own opinion and how reliably different clinicians when shown a line of data at a stability level mark it as such.

Myocardial association and prediction - Due to the large confidence intervals around both the positive predictive and negative predictive value of the rule base, this initial work was hypothesis generating. To overcome the weaknesses within the original studies which

might have led to this, future work would be done prospectively using a much larger cohort of patients. This would avoid reliance on detecting more than one event within the same patient. Greater troponin sampling frequency would identify more clearly if the rise and decay of troponin in Intensive Care patients is as per the ideal figure seen in textbooks and would decrease the chance of missing a small troponin rise that has already decayed by the time it is sampled under the current regime. The rule bases would be tested in a non-renal failure population to avoid the confounding factor of prolonged troponin delay. The rule set could run on auto charted data from CareVue and be calculated automatically many times an hour which would decrease the chance of profound physiological disturbance being missed between two hourly time points currently recorded. It would be interesting to model different combinations of rules and of different durations to ascertain if they give a better fit. It would also be useful to measure the new generation of highly sensitive troponins in addition to the current troponin test to see if the rules are detecting or predicting a troponin rise, but it is not being seen due to the sensitivity of the troponin test itself. This would initially require an analysis of the typical profile for rise and fall of highly sensitive troponins within the blood after a myocardial event.

Outcomes - This will be approached in a different manner. Rather than mean Intensive Care Unit - Patient Scoring System score over time, I will examine the percentage time in a 24 hour period spent at a particular level in the score and its association with outcome.

The future - I plan to develop a bedside monitor using the score in real time to give an overall summary of the physiological state of the patient. Initially, I will test a prototype in a high fidelity simulation to ascertain if clinical behaviour is altered with and without the assistance of the monitor. I hope that it can be introduced into clinical practice as an aid to less experienced staff who may not recognise deterioration in apparently “normal” physiological parameters, while the amount of support the patient requires is silently increasing.

In summary, in this thesis I have described the development and first stages in the validation of a novel scoring system for patients in Intensive Care which goes some way to addressing the problems of currently available scores. This could lead to a commercially available bedside monitor capable of increasing patient safety and of improving clinical outcomes.

Appendices

Appendix I

Patient 708 - Testing of the quantitative score on unextrapolated data

Time	Adrenaline	Noradrenaline	Fluids	Propofol	Alfentanil	HR	SpO2	FiO2	Urine	Temp	MAP	Dialysis	Score
Day 1													
19/12/2006 19:45:37						114.0	90.0						**2**
19/12/2006 20:00:00			500.0			111.0	92.0	100.0		37.7	62.0		**11**
19/12/2006 21:00:00			500.0			116.0	79.0	100.0			68.0		**12**
19/12/2006 22:00:00	1.0	2.0	500.0	60.0	1.5	99.0	69.0	100.0	80.0	38.1	62.0		**19**
19/12/2006 23:00:00	2.0	2.0	500.0	60.0	1.5	108.0	83.0	100.0	10.0		62.0		**20**
Day 2													
20/12/2006 00:00:00	2.8	4.0	250.0	60.0	1.5	110.0	100.0	100.0	15.0		59.0		**19**
20/12/2006 01:00:00	2.8	4.0	350.0	60.0	1.5	112.0	83.0	100.0	10.0	38.3	59.0		**23**
20/12/2006 02:00:00	2.8	4.0	100.0	60.0	1.5	109.0	83.0	100.0	25.0		63.0		**18**
20/12/2006 03:00:00	2.8	4.0	100.0	60.0	1.0	107.0	75.0	100.0	15.0	38.8	67.0		**20**
20/12/2006 04:00:00	2.8	4.0	100.0	60.0	1.0	112.0	76.0	100.0	0.0		80.0		**15**
20/12/2006 05:00:00	2.8	4.0	50.0	60.0	1.0	118.0	80.0	100.0		38.9	82.0		**13**
20/12/2006 06:00:00	2.8	4.0	100.0	60.0	1.0	121.0	82.0	100.0	35.0		83.0		**13**
20/12/2006 07:00:00	2.8	4.0	100.0	60.0	1.0	124.0	82.0	100.0	15.0	38.9	64.0		**20**
20/12/2006 08:00:00	2.8	4.0	50.0	60.0	1.0	126.0	85.0	100.0	10.0	39.8	66.0		**20**
20/12/2006 09:00:00	2.8	4.0	50.0	60.0	1.0	127.0	86.0	100.0			69.0		**17**
20/12/2006 10:00:00	2.8	4.0	50.0	60.0	1.0	123.0	85.0	100.0	10.0	39.2	63.0		**20**
20/12/2006 11:00:00	2.8	4.0	50.0	60.0	1.0	118.0	91.0	100.0			58.0		**16**
20/12/2006 12:00:00	2.2	4.0	50.0	60.0	1.0	116.0	95.0	100.0	5.0	38.7	59.0		**18**
20/12/2006 13:00:00	2.2	4.0	50.0	60.0	1.0	117.0	81.0	100.0	0.0		58.0		**20**
20/12/2006 13:37:00						117.0	81.0	100.0			58.0		**9**
20/12/2006 14:00:00	2.2	4.5		60.0	1.0	115.0	82.0	100.0		38.6	58.0		**18**
20/12/2006 15:00:00	2.2	5.0		60.0	1.0	114.0	89.0	100.0	5.0		54.0		**19**
20/12/2006 16:00:00	2.2	5.0		60.0	1.0	113.0	91.0	100.0	0.0	38.9	57.0		**20**
20/12/2006 17:00:00	2.2	5.0		60.0	1.0	112.0	77.0	100.0			63.0		**17**
20/12/2006 18:00:00	2.2	5.0		60.0	1.0	112.0	90.0	100.0	30.0	38.8	53.0	Dialysis	**18**
20/12/2006 19:00:00	2.2	5.0		60.0	1.0	111.0	91.0	100.0			53.0	Dialysis	**16**
20/12/2006 20:00:00	2.2	5.0		60.0	1.0		92.0	100.0				Dialysis	**4**
20/12/2006 21:00:00	2.2	5.0		60.0	1.0	111.0	93.0	100.0	0.0	38.5	52.0		**20**
20/12/2006 22:00:00	2.2	5.0		60.0	1.0	112.0	91.0	100.0	0.0		51.0		**19**
20/12/2006 23:00:00	2.2	5.0		60.0	1.0	112.0	87.0	100.0			50.0		**17**
Day 3													
21/12/2006 00:00:00	2.2	5.0		60.0	1.0	113.0	88.0	100.0	0.0	38.5	53.0		**21**
21/12/2006 01:00:00	2.2	5.0		60.0	1.0	113.0	88.0	100.0	0.0		51.0		**20**
21/12/2006 02:00:00	2.2	5.0		60.0	1.0	113.0	91.0	100.0			47.0		**19**
21/12/2006 03:00:00	2.2	5.0		60.0	1.0	114.0	93.0	100.0			48.0		**19**
21/12/2006 04:00:00	2.2	5.0		60.0	1.0	113.0	93.0	100.0		39.4	46.0		**20**
21/12/2006 05:00:00	2.2	5.0		60.0	1.0	114.0	89.0	100.0	0.0		45.0		**23**
21/12/2006 06:00:00	2.2					114.0	88.0	100.0		40.3	46.0		**18**
21/12/2006 07:00:00	2.2	5.0		60.0	1.0	113.0	89.0	100.0	0.0		47.0		**23**
21/12/2006 08:00:00	2.2	5.0		60.0	1.0	112.0	89.0	100.0		40.8	47.0		**22**
21/12/2006 09:00:00						113.0	89.0	100.0			47.0		**11**

Appendix II

Patient 708 - Quantitative score applied to final extrapolated data

Time	Adrenaline	Noradrenaline	Fluids	Propofol	Alfentanil	HR	SpO2	FiO2	Urine	Temp	MAP	Dialysis	Score
Day 1													
19/12/2006 19:45:37						114.0	90.0						**2**
19/12/2006 20:00:00			500.0			111.0	92.0	100.0		37.7	62.0		**11**
19/12/2006 21:00:00			500.0			116.0	79.0	100.0		37.7	68.0		**12**
19/12/2006 22:00:00	1.0	2.0	500.0	60.0	1.5	99.0	69.0	100.0	80.0	38.1	62.0		**19**
19/12/2006 23:00:00	2.0	2.0	500.0	60.0	1.5	108.0	83.0	100.0	10.0		62.0		**20**
Day 2													
20/12/2006 00:00:00	2.8	4.0	250.0	60.0	1.5	110.0	100.0	100.0	15.0		59.0		**19**
20/12/2006 01:00:00	2.8	4.0	350.0	60.0	1.5	112.0	83.0	100.0	10.0	38.3	59.0		**23**
20/12/2006 02:00:00	2.8	4.0	100.0	60.0	1.5	109.0	83.0	100.0	25.0	38.3	63.0		**19**
20/12/2006 03:00:00	2.8	4.0	100.0	60.0	1.0	107.0	75.0	100.0	15.0	38.8	67.0		**20**
20/12/2006 04:00:00	2.8	4.0	100.0	60.0	1.0	112.0	76.0	100.0	0.0	38.8	80.0		**16**
20/12/2006 05:00:00	2.8	4.0	50.0	60.0	1.0	118.0	80.0	100.0	0.0	38.9	82.0		**16**
20/12/2006 06:00:00	2.8	4.0	100.0	60.0	1.0	121.0	82.0	100.0	35.0	38.9	83.0		**14**
20/12/2006 07:00:00	2.8	4.0	100.0	60.0	1.0	124.0	82.0	100.0	15.0	38.9	64.0		**20**
20/12/2006 08:00:00	2.8	4.0	50.0	60.0	1.0	126.0	85.0	100.0	10.0	39.8	66.0		**20**
20/12/2006 09:00:00	2.8	4.0	50.0	60.0	1.0	127.0	86.0	100.0	10.0	39.8	69.0		**20**
20/12/2006 10:00:00	2.8	4.0	50.0	60.0	1.0	123.0	85.0	100.0	10.0	39.2	63.0		**20**
20/12/2006 11:00:00	2.8	4.0	50.0	60.0	1.0	118.0	91.0	100.0	10.0	39.2	58.0		**19**
20/12/2006 12:00:00	2.2	4.0	50.0	60.0	1.0	116.0	95.0	100.0	5.0	38.7	59.0		**18**
20/12/2006 13:00:00	2.2	4.0	50.0	60.0	1.0	117.0	81.0	100.0	0.0		58.0		**20**
20/12/2006 13:37:00	2.2	4.0		60.0	1.0	117.0	81.0	100.0	0.0		58.0		**20**
20/12/2006 14:00:00	2.2	4.5		60.0	1.0	115.0	82.0	100.0	0.0	38.6	58.0		**21**
20/12/2006 15:00:00	2.2	5.0		60.0	1.0	114.0	89.0	100.0	5.0	38.6	54.0		**20**
20/12/2006 16:00:00	2.2	5.0		60.0	1.0	113.0	91.0	100.0	0.0	38.9	57.0		**20**
20/12/2006 17:00:00	2.2	5.0		60.0	1.0	112.0	77.0	100.0	0.0	38.9	63.0		**21**
20/12/2006 18:00:00	2.2	5.0		60.0	1.0	112.0	90.0	100.0	30.0	38.8	53.0	Dialysis	**18**
20/12/2006 19:00:00	2.2	5.0		60.0	1.0	111.0	91.0	100.0	30.0		53.0	Dialysis	**17**
20/12/2006 20:00:00	2.2	5.0		60.0	1.0	111.0	92.0	100.0	30.0		53.0	Dialysis	**17**
20/12/2006 21:00:00	2.2	5.0		60.0	1.0	111.0	93.0	100.0	0.0	38.5	52.0		**20**
20/12/2006 22:00:00	2.2	5.0		60.0	1.0	112.0	91.0	100.0	0.0		51.0		**19**
20/12/2006 23:00:00	2.2	5.0		60.0	1.0	112.0	87.0	100.0	0.0		50.0		**20**
Day 3													
21/12/2006 00:00:00	2.2	5.0		60.0	1.0	113.0	88.0	100.0	0.0	38.5	53.0		**21**
21/12/2006 01:00:00	2.2	5.0		60.0	1.0	113.0	88.0	100.0	0.0		51.0		**20**
21/12/2006 02:00:00	2.2	5.0		60.0	1.0	113.0	91.0	100.0	0.0		47.0		**22**
21/12/2006 03:00:00	2.2	5.0		60.0	1.0	114.0	93.0	100.0	0.0		48.0		**22**
21/12/2006 04:00:00	2.2	5.0		60.0	1.0	113.0	93.0	100.0	0.0	39.4	46.0		**23**
21/12/2006 05:00:00	2.2	5.0		60.0	1.0	114.0	89.0	100.0	0.0	39.4	45.0		**24**
21/12/2006 06:00:00	2.2	5.0		60.0	1.0	114.0	88.0	100.0	0.0	40.3	46.0		**25**
21/12/2006 07:00:00	2.2	5.0		60.0	1.0	113.0	89.0	100.0	0.0	40.3	47.0		**25**
21/12/2006 08:00:00	2.2	5.0		60.0	1.0	112.0	89.0	100.0	0.0	40.8	47.0		**25**
21/12/2006 09:00:00						113.0	89.0	100.0	0.0		47.0		**14**

Appendix III

Document for computing scientists explaining the quantitative score

Explanation of Scoring System

Heart Rate (HR)
Needs no explanation. A very fast of slow rate can impair the filling of the heart and may represent an abnormal heart rhythm.

Mean Arterial Blood Pressure (MAP)
This is probably more relevant in the Intensive Care setting than systolic and diastolic blood pressure. For your interest it is derived from the equation:

Mean Arterial Pressure = Diastolic Pressure + 0.333 (Systolic Pressure – Diastolic Pressure)

Too low a pressure and you do not perfuse vital organs e.g. kidneys, too high a pressure and you increase your chance of heart attack, stroke etc.

Central Venous Pressure (CVP)
This gives an indication of filling of the right side of the heart i.e. blood coming back to the heart from the body. Usually measured via a line inserted into the internal jugular or subclavian veins. If when a patient's blood pressure is low you give them a fluid bolus and their CVP remains unchanged then this is an indication that they can cope with more fluid before commencing a drug to increase the blood pressure (inotropes). If the CVP increases dramatically then this is an indication that they are well filled with fluid.

Cardiac Output (CO)
This is the product of your heart rate and stroke volume (the amount of blood ejected from your heart with each contraction). If you have a heart attack and a failing heart is struggling to eject blood then your cardiac output will be low.

Cardiac Index (CI)
A way of comparing people of different body sizes. Derived by taking cardiac output and dividing by body surface area (for an average 70kg man this is $1.7m^2$).

Stroke Volume
The amount of blood ejected from the heart with each contraction. Within certain limits, the more blood returning to the heart the greater the stroke volume (Frank-Starling relationship).

Stroke Volume Variance
Reflects the variation in stroke volume caused by changes in intrathorcic pressure when e.g. a patient is being ventilated. The greater the variance the more likely it is that the patient still requires extra fluid.

Systemic Vascular Resistance (SVR)
In simplistic terms the resistance against which the heart must contract to eject blood into the body. If a patient is septic they are often vary vasodilated and may have a low SVR. To raise their blood pressure appropriate therapy would comprise fluid and a drug which "tightens up" their circulation e.g. noradrenaline (see later).

Systemic Vascular Resistance Index (SVRI)
As for cardiac index this corrects SVR for body surface area.

Oxygen Delivery (D02)
This is the amount of oxygen delivered to the peripheral tissues and is obtained by multiplying the arterial oxygen content of blood (20mls/100mls blood) and the cardiac output (5 litres) giving a figure of 1000mls/min.

Oxygen Saturation (Sp02)
Oxygen is transported in the blood by being bound to haemoglobin (as well as a small dissolved fraction). The oxygen saturation is the %haemoglobin saturation with oxygen. If you draw a graph of haemoglobin saturation (%) against oxygen tension (kPa) (which drives the oxygen to bind with the haemoglobin) then you get a sigmoid shaped graph. This explains why a saturation of 75% although not a low number is actually very serious. The saturation is measured by a finger or ear probe, which uses infrared light.

Temperature
Self-explanatory except to say that extremes of high and low are bad for many body systems e.g. at high temperatures various enzymatic processes start to become disturbed.

Urine Output
Self-explanatory. Roughly speaking you need to perfuse your kidneys with a mean pressure of 60mmHg to produce urine. This may be more in a patient with high blood pressure.

Propofol 1%
A phenolic derivative used by injection to induce general anaesthesia. It may be used in lower concentration to sedate patients in Intensive Care. It causes a dose-dependent reduction in vascular tone that reduces systemic vascular resistance (SVR), central venous pressure (CVP) and cardiac output (CO). Heart rate remains relatively unchanged.

Alfentanil
A synthetic opiate. Used in higher doses during general anaesthesia but in lower doses to sedate patients in Intensive Care. May cause vasodilation (hence lower SVR), slowing of the heart rate and low blood pressure (hypotension).

Adrenaline
A naturally occurring catecholamine. It is a positive inotrope (a drug which increases the force of contraction of the heart) and hence raises blood pressure. At higher doses it also increases systemic vascular resistance (SVR). It is used in Intensive Care to raise blood pressure in patients with low cardiac output when they are adequately filled with fluid.

Noradrenaline
Another catecholamine. Like adrenaline it is a positive inotrope. However its main effect is to “tighten up” the peripheral circulation and is thus used to raise blood pressure in patients where it is low because of a low systemic vascular resistance e.g. in sepsis. In practice patients often require a mixture of adrenaline and noradrenaline for blood pressure support.

Appendix IV

Complete transcript of discussion with the computer scientists which lead to classification of broad levels of stability

Patient 733

Overall: Medium Worst

General Description:

Patient is unstable until around the morning of the 13th when patient starts to stabilise. Apart from the odd 'blip' patient is on a general upwards trend.

Detailed Description:

Time & Date	Comments	Condition of Patient
8/01/07 10:00:00	Probably intubated, hence Propofol which affects the blood pressure. Blood pressure low so gave some fluids	Getting Worse
8/01/07 11:00:00	FiO2 probably turned down to check oxygen saturation	
8/01/07 13:00:00 - 18:00:00	Slightly worse as blood pressure not increased whilst on a higher amount of Noradrenaline.	Worse
8/01/07 19:00:00 - 9/01/07 00:00:00	Worse than when patient first admitted. Heart rate is higher and oxygen level is very bad	Worse
9/1/07 02:00:00 – 9/01/07 06:00:00	Slightly better oxygen saturation and blood pressure better	Slight improvement
09/01/07 07:00:00 - 09/01/07 14:00:00	Much worse. Very bad at 13:30. Oxygen very bad, Noradrenaline has increased whilst blood pressure has decreased	Much worse
09/01/07 16:01:00 - 09/01/07 19:00:00	DIALYSIS Not much better. Dialysis may have been predicted because urine output and blood pressure not good	Worse
09/01/07 20:00:00 - 10/01/07 00:00:00	Not much change after dialysis. Urine low, it appears kidneys have taken a hit. Possibly Septic	No change
10/01/07 11:00:00 - 10/01/07 11:00:00	DIALYSIS Little better on dialysis. Noradrenaline down, oxygen down, but Noradrenaline increased. Blood Pressure okay but on increased Noradrenaline Wobble at 11:09:00 – prob down to cardiovascular problems	No change
10/01/07 17:00:00	Bit better after dialysis	No change/ Slight improvement
11/01/07 02:00:00 - 11/01/07 05:00:00	Better	Improvement
11/01/07 06:00:00 - 11/01/07 10:00:00	DIALYSIS Slight increase in oxygen.	No change

	Blood Pressure down a little Noradrenaline increased	
11/01/07 11:00:00 - 11/01/07 13:00:00	Blood pressure increased after dialysis. Much better now as Noradrenaline has gone	Improvement
11/01/07 14:00:00 - 11/01/07 17:00:00	Stable	Stable/ No change
11/01/07 18:00:00 - 11/01/07 23:00:00	Wobble. Blood pressure down, FiO2 up. Worse again. Perhaps down to secretions in the lung.	Worse
Morning of 12/01/07	Much better	Big improvement
12/01/07 18:00:00	DIALYSIS Blood pressure okay but Propofol has been increased. Bit more unstable on dialysis. Oxygen increased, in response to major wobble on dialysis. Towards end of session starts to adapt to dialysis.	Worse on dialysis
13/01/07 02:00:00 - 13/01/07 10:00:00	Blood pressure low despite less Propofol. Not quite as good but less oxygen has been given.	Slight improvement
13/01/07 10:00:00 – 14/01/07 04:00:00	Better again, stable. Oxygen probably not going to get any lower than 35%.	Improvement/Stable
14/01/07 05:00:00 - 14/01/07 13:00:00	Good. Making urine again. Best patient has been.	Improvement
14/01/07 14:00:00 - 14/01/07 13:00:00	DIALYSIS Oxygen has been increased on dialysis. Alfentanil and Propofol lower on dialysis. Patient little bit worse on dialysis.	Worse
15/01/07 00:00:00 - 15/01/07 05:00:00	Propofol up a little, Blood pressure down a little	No change
15/01/07 06:00:00 - 15/01/07 15:00:00	DIALYSIS Initially fine on dialysis. Slight wobble at start as FiO2 increased back up. Recovered from wobble much quicker.	No change
15/01/07 16:00:00 – 16/01/07 06:00:00	No change. Everything okay. Controlled, stable.	No change
16/01/07 & 17/01/07	Oxygen fine, HR decreasing. Slowly things are getting better. Coping well with dialysis	Improvement
18/01/07	Fine, HR fine	No change
19/01/07 & 20/01/07	Blood pressure not changing much on dialysis	No change
21/01/07 & 22/01/07	Stability, low oxygen and heart rate fine.	Improvement
23/01/07	FiO2 of 28 often when taking the tube out of the patient. Starting to make urine.	Improvement

Patient 708

Overall: Worst

General Description:

Patient enters ICU almost as sick as a patient can be. They then deteriorate throughout the session resulting in what is suspected as the patient dying.

Time & Date	Comments	Condition of Patient
19/12/06 19:45:37 – 19/12/06 21:00:00	Heart rate high, oxygen saturation low. Aggressive fluids are given in response to high heart rate. Test patient's response to fluids. Despite being given the fluids, the blood pressure is low. Intubated at 8pm. Very sick patient. Despite incubation, oxygen saturation only went up to 92.	Very bad, getting worse
19/12/06 22:00:00	Central line put in. Blood pressure not responding to Adrenaline and Noradrenaline. FiO2 and SpO2 are grim. Most likely septic as a lot of adrenaline and nor adrenaline given.	Worse - Decreasing
20/12/06 02:00:00 - 20/12/06 04:00:00	100% oxygen given but saturation decreasing. Increase in adrenaline and nor adrenaline but blood pressure still low.	Worse - Decreasing
20/12/06 05:00:00 - 20/12/06 10:00:00	Getting worse. Blood pressure not moving. Heart rate increasing. Urine output tailing off. Oxygen saturation dire.	Worse - Decreasing
20/12/06 13:00:00 - 20/12/06 15:00:00	High amounts of adrenaline and Noradrenaline. Blood pressure grim. No change, very unwell.	Stable – No worse
20/12/06 16:00:00 - 20/12/06 17:00:00	Oxygen worse. Noradrenaline increased. Not enough blood pressure for urine.	Worse
20/12/06 18:00:00 - 20/12/06 20:00:00	DIALYSIS Get worse on dialysis. Blood pressure even lower. Oxygen saturation is slightly higher. Could be because fluid had built up in the lungs and has now been removed.	On balance, stable
Rest of session	Gradual deterioration. Oxygen saturation and blood pressure continue to decrease. Becoming more and more septic. Patient in the end dies, probably from a cardiac arrest/deciding not to increase drugs further.	Much worse

Patient 728

Overall: Least Worst

General Description:

Initially get worse upon admittance to ICU and then stabilise morning of the 9/01/07

Time & Date	Comments	Condition of Patient
05/01/07 17:00:00 - 05/01/07 18:00:00	Must be worried about blood pressure as a lot of fluid has been given to the patient. Quite bad situation.	Bad
05/01/07 18:00:00	Intubated. Blood pressure poor considering lots of fluid has been given.	
05/01/07 19:00:00 - 05/01/07 21:00:00	Blood pressure still low but given Noradrenaline. Heart rate is high. Overall deterioration.	Worse - Decreasing
05/01/07 22:00:00 - 05/01/07 23:00:00	No better. Oxygen saturation unchanged but FiO2 has decreased. Noradrenaline has been increased to maintain blood pressure.	No change
06/01/07 00:00:00 - 06/01/07 06:00:00	Blood pressure bit better and urine a bit better, heart rate is okay. Sedation has been lowered. Fluid down.	Bit more stable
06/01/07 15:00:00 - 06/01/07 18:00:00	Blood pressure worse. Heart rate okay. Sedation the same.	Stable
06/01/07 20:00:00 - 06/01/07 21:00:00	DIALYSIS Blood pressure copes well on dialysis. Oxygen saturation okay. Heart rate unchanged. Sedation cut back.	Stable
06/01/07 22:00:00 - 07/01/07 06:00:00	Blood pressure doesn't really alter. Noradrenaline turned down.	Stable
07/01/07 07:00:00 - 07/01/07 12:00:00	Oxygen much the same and saturation lower. Blood pressure quite low and Noradrenaline has been reduced. Overall improvement.	Slight improvement
07/01/07 13:00:00 - 07/01/07 18:00:00	DIALYSIS Blood pressure not altered and coped well on dialysis. At times slightly worse	Slightly worse
07/01/07 19:00:00 – 08/01/07 09:00:00	Wobble after dialysis Noradrenaline back on to maintain blood pressure. Deterioration.	Slightly worse
08/01/07 10:00:00 - 16/01/07 16:00:00	DIALYSIS Blood pressure lower. Still on Noradrenaline. Slight blood pressure hit.	Slightly worse
08/01/07 17:00:00 - 08/01/07 20:00:00	Blood pressure unaltered. Oxygen lower. Stable	Improvement
08/01/07 21:00:00 - 09/01/07 08:00:00	Bit of improvement. Oxygen decreased.	Improvement
09/01/07 09:00:00 - 09/01/07 14:00:00	DIALYSIS Blood pressure fine. Oxygen the same. Propofol	Stable

	decreased. Coped well on dialysis.	
Afternoon of 09/01/07 and Morning of 10/01/07	Noradrenaline stopped and blood pressure okay. Slight improvement.	Improvement
10/01/07 12:00:00 - 10/01/07 16:00:00	DIALYSIS Blood pressure doesn't change much.	Stable
Evening of 10/01/07 and morning of 11/01/07	Very stable. Producing reasonable volumes of urine.	Improvement
11/01/07 12:00:00 - 11/01/07 20:00:00	DIALYSIS Slight drop in blood pressure	Slightly worse
12/01/07	Fine low oxygen. Blood pressure fine. Heart Rate fine.	Improvement
12/01/07 16:00:00	DIALYSIS Slight drop of blood pressure	Slightly worse
Rest of session	Happier situation. Temperature coming down. Heart rate getting slower. Lot better.	Improvement

Appendix V

Assigning stability levels and starting to formulate ranges for parameters

Patient 708

Day	Comments – John	Level Given - John	Comments - Malcolm	Level Given - Malcolm
Day 1 19/12/06	First thing, given patient fluid in a large volume and ADR & NORAD and it has taken several hours to get on top of the situation. Low BP, High HR and a lot of treatment used to get those values. MAP, adequate value is between 60 -100. HR >100 abnormal. Fluid > 1litre – high amount Fluid > 700ml – Starting to worry Supporting evidence, Urine and FiO2. Main points: 1) Total fluids initially high	E	100% oxygen but saturation only up to 90%. Heart Rate is very high	E
Day 2 20/12/09	Still unstable despite need for fluids decreasing. NORAD increases and ADR decreased. Still high HR & BP not impressive. IF BP in the 50s – losing the battle. Looking at trends in particular, the running average for MAP & HR.	E	Low oxygen saturation despite still at 100% Oxygen. High Heart Rate and hypotensive despite both adrenaline and noradrenaline. 5) Oxygen 6) High HR despite ADR & NORAD 7) Blood Pressure 8) Urine Output Oxygen Saturation: 96-94: Not bad Below 94: Bad 90-84: Very Bad	E
Day 3 21/12/09	Remains unstable/dying. Increase in NORAD & ADR, BP decreasing and there is a fast HR. Average ADR & NORAD. Looking at trend of MAP.	6	Patient stays bad 4) Oxygen saturation bad and 100% oxygen 5) Blood Pressure & HR 6) Urine Heart Rate HR > 140: Bad as heart doesn't refill properly.	E

Patient 728

Day	Comments - John	Level Given - John	Comments - Malcolm	Level Given - Malcolm
Day 1 05/01/07	Getting fluid, starting NORAD, HR quite high (looks at peak). MAP quite low despite large amount of FLUID and NORAD	E	100% oxygen required. Despite a lot of fluid and a lot of norad. , blood pressure is low. 1) Oxygen 2) Blood Pressure low despite NORAD & FLUID 3) HR 4) Urine	E
Day 2 06/01/07	Beginning to stabilise. NORAD dose stable. BP miles better (in normal range). HR good. Looking at a balance of fluids, drugs and BP.	D	100% oxygen down to 75% and saturation okay. Blood Pressure bit better and requiring less fluid. 1) Oxygen and Saturation 2) Blood Pressure 3) Fluid MAP Look at a particular value 60 for normal people 70/80 normal for ICU	D
Day 3 07/01/07	Stable, off NORAD – good. Little dip in BP but still in okay range. HR okay. Balance of decreasing vasopressor.	D -> C	Oxygen down to 65% and saturation maintained. Blood pressure okay and heart rate okay. NORAD coming down. Urine variable 1) Oxygen and Saturation 2) NORAD 3) Urine	C
Day 4 08/01/07	Not getting much fluid, low level of NORAD, cardiovascular stable, MAP and HR in range.	C	1) Oxygen and Saturation 2) Blood Pressure, HR & NORAD 3) Urine	B
Day 5 09/01/07	Better – Oxygen not high, BP fine, NORAD disappears. Healthy BP – no change on dialysis.	B	Down to 50% oxygen and saturation still okay. Blood pressure good on smaller amount of inotrope. Heart rate coming down and urine still not quite right. 1) Oxygen and Saturation 2) HR 3) Urine	B
Day 6 10/01/07	Better – urine vol not happened as in kidney failure. All ranges normal. No ADR/NORAD, lower oxygen. Handled dialysis well. Fluid doesn't appear.	B	40% oxygen – as low as it gets. Good saturation. Blood pressure good, no inotrope used. Urine being made. 1) Oxygen and Saturation 2) Blood Pressure & inotrope 3) Urine	A

Day 7 11/01/07	Really stable. BP fine, HR fine. Handled dialysis no problems.	A	Still 40% oxygen and good saturation, MAP good and no inotrope given, urine being made. 1) Oxygen and Saturation 2) MAP & inotrope 3) HR 4) Urine	A
Day 8 12/01/07	Same as Day 7	A	Same as Day 7	A
Day 9 13/01/07	Same as Day 7	A	Same as Day 7	A
Day 10 14/01/07	Receiving a sedative just to keep him in bed.	A	Same as Day 7	A
Day 11 15/01/07	Same as Day 7	A	Same as Day 7	A
Day 12 16/01/07	80 an hour fluid – normal drip, probably pulled the tube out. Suddenly started producing urine. Wouldn't give that a score.	A	Same as Day 7	A

Appendix VI

Annotations of patient 720 by Dr. M. Sim

Time	HR	MAP	CVP	FiO2	SpO2	Urine (mls/h)	Prop. (mg/h)	Hart. (mls/h)	Gelo.	Alfent. (mg/h)	Dialysis	Adren. (mg/h)	Sodium Chloride (mls/h)	Norad (mg/h)	Score A-E
Day 1															
29/12/2006 17:00	102	78	7	1	96	60	100	0							**E**
29/12/2006 17:10	107	71			100										**E**
29/12/2006 18:00	104	74	7	1	95	30	100	125	500						**E**
29/12/2006 19:00	106	70	9	0.8	99	5	100	125	500						**D**
29/12/2006 20:00	108	69	9	0.7	98	0	100	125							**D**
29/12/2006 21:00	109	64	11	0.7	98	0	100	125							**D**
29/12/2006 22:00	109	64	12	0.6	99	240	20	125							**D**
29/12/2006 23:00	105	66	9	0.6	94	30	0	125		0.5					**D**
Day 2															
30/12/2006 00:00	106	72	10	0.6	93			125		0.5					**D**
30/12/2006 01:00	112	83	11	0.6	91	110		125		0.5					**D**
30/12/2006 02:00	110	71	11	0.6	91			125		0.5					**D**
30/12/2006 03:00	109	69	13	0.6	92	100		125		0.5					**D**
30/12/2006 04:00	105	59	7	0.7	93	50		125	250	0.5					**D**
30/12/2006 05:00	106	56	8	0.7	91	50		125	250	0.5					**D**
30/12/2006 06:00	109	55	8	0.7	93	40		125		0.5					**D**
30/12/2006 07:00	103	54	12	0.7	67	35		125		0.5					**D**
30/12/2006 08:00	110	64	13	0.7	97	45		125	500	0.5					**D**
30/12/2006 09:00	106	66	20	0.7	99	30		125		0.5					**D**
30/12/2006 10:00	103	65	23	0.7	100	45		125		0.5					**D**
30/12/2006 11:00	103	56	14	0.7	100	25		125	500	0.5					**D**
30/12/2006 12:00	100	56	13	1	99			125	500	0.5					**E**
30/12/2006 12:35		56	14	0.7											**E**
30/12/2006 13:00	99	54	12	1	99	30		125		0.5	Dialysis				**E**
30/12/2006 14:00	110	72	9	0.95	96	25		125	500	0.5	Dialysis	0.6			**E**
30/12/2006 15:00	127	75	4	0.95	98	30		125		0.5		0.6			**E**
30/12/2006 16:00	128	85		0.9	99	15		125		0.5		1			**E**
30/12/2006 17:00	120	99	10	0.9	95	20		125		0.5		1			**E**
30/12/2006 18:00	117	86	8	0.9	94	20		125		0.5		1			**E**
30/12/2006 19:00	110	77	7	0.9	98	15		60		1					**E**
30/12/2006 20:00	112	69	12	0.9	96	20				1		1	60		**E**
30/12/2006 21:00	111	69	15	0.9	92	15			500	1		1			**E**
30/12/2006 22:00	108	59	42	0.9	91	15				1		1	60		**E**
30/12/2006 23:00	115	78	16	1	91	15				1		1	60	0.4	**E**
Day 3															
31/12/2006 00:00	115	66	13	1	95	10				1		0.8	60	0.6	**E**
31/12/2006 01:00	113	63	14	1	95	0				1		0.8	60	0.6	**E**
31/12/2006 02:00	111	65	13	1	94	5				1		0.8	60	0.8	**E**
31/12/2006 03:00	111	67	19	1	96	20				1		0.6	60	0.9	**E**
31/12/2006 04:00	113	60	19	1	91	0				1		0.2	60	1.1	**E**
31/12/2006 04:41														1.4	**E**

31/12/2006 05:00	115	69	18	1	92	5			500	1		0.2	60	1.4	**E**
31/12/2006 05:18														1.3	**E**
31/12/2006 05:40										1				1.1	**E**
31/12/2006 06:00	114	76	24	1	98	0				1			60	1.1	**E**
31/12/2006 07:00	119	71	27	1	100	5				1			60	1	**E**
31/12/2006 08:00	111	63	26	1	100	40				1			60	1	**E**
31/12/2006 09:00	109	64	14	1	95	10				1			60	1	**E**
31/12/2006 10:00	108	76	21	1	97	10				1	Dialysis		60	1	**E**
31/12/2006 10:25	108	76	21	1	97										**E**
31/12/2006 11:00	111	66	17			0					Dialysis			1	**E**
31/12/2006 12:00	119	73		1	94	5			250		Dialysis	0.2		1.6	**E**
31/12/2006 13:00	115	71	17	1	95				250	1	Dialysis	0		1.6	**E**
31/12/2006 14:00	110	76	13	1	95	5				1	Dialysis			1.6	**E**
31/12/2006 14:25				1	94										**E**
31/12/2006 15:00	154	67	15	1	96	10				1	Dialysis			1.4	**E**
31/12/2006 16:00	139	70		1	95	10				1	Dialysis			1.4	**E**
31/12/2006 17:00	78	82	18	1	95	0				0.5				1.4	**E**
31/12/2006 18:00	77	89		1	95	0				0.5				1.4	**E**
31/12/2006 19:00	80	82	18	1	95	10				0.5				0.8	**E**
31/12/2006 20:00	84	81	16	1	96	10				1				0.8	**E**
31/12/2006 21:00	85	81		1	95	10				1				0.8	**E**
31/12/2006 22:00	90	72	12	1	95	10				1				0.8	**E**
31/12/2006 23:00	86	71	12	1	98	10				1				0.8	**E**
Day 4															
01/01/2007 00:00	89	74	11	1	97	5				1				0.8	**E**
01/01/2007 01:00	89	76		1	97	5				1				0.8	**E**
01/01/2007 02:00	90	69		1	96	5				1				0.8	**E**
01/01/2007 03:00	95	68	12	1	98	5				1				0.8	**E**
01/01/2007 04:00	95	71	14	1	97	0				1				0.8	**E**
01/01/2007 05:00	97	70	11	1	98	5				1				0.8	**E**
01/01/2007 06:00	96	65	11	1	99	0				1				0.8	**E**
01/01/2007 07:00	92	82	21	1	96	0				1				0.8	**E**
01/01/2007 08:00	94	87	22	1	97	25				1				0.8	**E**
01/01/2007 09:00	94	71	18	1	97	0								0.8	**E**
01/01/2007 09:27	94	71	18	1	97										**E**
01/01/2007 10:00	94	83	16	1	99	0					Dialysis			0.8	**E**
01/01/2007 11:00	83	75	16	1	97	25					Dialysis			0.8	**E**
01/01/2007 12:00	84	78	16	1	96	5					Dialysis			0.8	**E**
01/01/2007 13:00	77	77	11	0.95	93	10					Dialysis			0.8	**E**
01/01/2007 14:00	79	82	6	0.9	96	0					Dialysis			0.8	**D**
01/01/2007 15:00	72	89	9	0.85	95	0					Dialysis			0.8	**D**
01/01/2007 16:00	75	83	5	0.8	96	0					Dialysis			0.7	**D**
01/01/2007 17:00	80	91	6	0.75	94									0.7	**D**
01/01/2007 18:00	85	91	7	0.75	93									0.4	**D**
01/01/2007 19:00	85	86	12	0.8	93	15								0.3	**C**
01/01/2007 20:00	84	85	12	0.8	94	0								0.3	**C**
01/01/2007 21:00	86	84	12	0.8	95	0								0.3	**C**
01/01/2007 22:00	87	78	12	0.8	94	20								0.3	**C**
01/01/2007 23:00	88	66	11	0.8	93	0								0.3	**D**
Day 5															

02/01/2007 00:00	90	63	10	0.8	94	0								0.3	**D**
02/01/2007 01:00	92	64	10	0.85	96	0								0.3	**D**
02/01/2007 02:00	95	72	11	0.85	97	5								0.3	**D**
02/01/2007 03:00	97	71	12	0.85	97	0								0.3	**D**
02/01/2007 04:00	91	73	12	0.85	97	0								0.3	**D**
02/01/2007 05:00	94	73	12	0.85	98	0								0.3	**D**
02/01/2007 06:00	91	75	11	0.85	98	0								0.3	**D**
02/01/2007 07:00	91	80	15	0.85	98	20								0.3	**D**
02/01/2007 08:00	91	73	14	0.85	99	0								0.1	**D**
02/01/2007 09:00	90	76	15	0.85	99									0.1	**D**
02/01/2007 09:58	90	76	15	0.85	99										**D**
02/01/2007 10:00	91	71	14	0.7	95						Dialysis			0.1	**C**
02/01/2007 11:00	90	67	14	0.7	96	25					Dialysis				**C**
02/01/2007 12:00	91	68	12	0.7	100	0					Dialysis				**C**
02/01/2007 13:00	88	74	17	0.6	100	15					Dialysis				**B**
02/01/2007 14:00	90	74	16	0.5	100	0					Dialysis				**B**
02/01/2007 15:00	88	72	17	0.45	98	0					Dialysis				**B**
02/01/2007 16:00	87	72	16	0.55	100	15					Dialysis				**B**
02/01/2007 17:00	86	80	16	0.55	97	0									**B**
02/01/2007 18:00	83	76	16	0.55	96	0									**C**
02/01/2007 19:00	94	64	16	0.65	96	5									**C**
02/01/2007 20:00	90	69	14	0.65	96	0									**C**
02/01/2007 21:00	91	64	15	0.7	98	0									**C**
02/01/2007 22:00	94	67	19	0.7	97	0									**C**
02/01/2007 23:00	94	61	16	0.7	97	10									**C**
															C
Day 6															**C**
															C
03/01/2007 00:00	98	60	16	0.7	100	0									**C**
03/01/2007 01:00	103	60	17	0.7	100	0									**C**
03/01/2007 02:00	109	65	17	0.7	100	0			250						**C**
03/01/2007 03:00	106	65	15	0.7	97	20									**C**
03/01/2007 04:00	106	70	18	0.7	98	0									**C**
03/01/2007 05:00	107	73	16	0.7	98										**C**
03/01/2007 06:00	108	75	20	0.7	99										**C**
03/01/2007 07:00	105	70	16	0.7	100										**C**
03/01/2007 08:00	103	71	16	0.7	100	20									**C**
03/01/2007 09:00	106	69	14	0.7	100	10									**C**
03/01/2007 10:00	100	66	15	0.65	98	5									**C**
03/01/2007 10:03	100	66	15	0.65	98										**C**
03/01/2007 11:00	99	62	16	0.7	95	0									**C**
03/01/2007 12:00	101	60	14	0.75	95	5					Dialysis				**D**
03/01/2007 13:00	101	63	12	0.75	96						Dialysis				**D**
03/01/2007 14:00	103	64	15	0.75	96	7					Dialysis				**D**
03/01/2007 15:00	102	59	11	0.75	97	0					Dialysis				**D**
03/01/2007 16:00	101	61	10	0.75	98	5					Dialysis				**D**
03/01/2007 17:00	102	59	15	0.75	100	5					Dialysis				**D**
03/01/2007 18:00	100	55	10	0.75	100	0									**D**
03/01/2007 19:00	104	54	10	0.75	100	0									**D**
03/01/2007 20:00	102	49	11	0.8	92	0									**D**
03/01/2007 21:00	108	69	12	0.8	97	0			500						**D**
03/01/2007 21:20				0.8	93										**D**
03/01/2007 22:00	105	70	27	0.8	95										**D**
03/01/2007 23:00	107	78	19	0.8	98	20									**D**

Day 7															
04/01/2007 00:00	104	69	33	0.8	98	0									**D**
04/01/2007 01:00	106	69	18	0.8	98	0									**D**
04/01/2007 02:00	103	62		0.8	97	5									**D**
04/01/2007 03:00	106	68	17	0.8	98	0									**D**
04/01/2007 04:00	106	67	19	0.8	97	0									**D**
04/01/2007 05:00	112	63	17	0.85	95	0									**E**
04/01/2007 06:00	103	59	21	0.85	98	0					Dialysis				**E**
04/01/2007 07:00	108	57	19	0.85	100	20					Dialysis				**E**
04/01/2007 08:00	110	59	19	0.85	100	0					Dialysis				**E**
04/01/2007 09:00	108	56	18	0.85	100	0					Dialysis				**E**
04/01/2007 10:00	111	54	20	0.85	100	0					Dialysis				**E**
04/01/2007 11:00	113	54	23	0.85	100	0					Dialysis				**E**
04/01/2007 12:00	110	52	19	0.8	100	0									**E**
04/01/2007 12:08	113	54	23	0.85	100										**E**
04/01/2007 13:00	111	67	18	0.85	100	0									**E**
04/01/2007 14:00	109	57	19	0.85	96	0									**E**
04/01/2007 15:00	111	59	20	0.85	95	0									**E**
04/01/2007 18:00	115	61	15	0.95	98	0									**E**
04/01/2007 19:00	115	59	14	0.95	99	0									**E**
04/01/2007 20:00	116	58	15	0.95	99	15									**E**
04/01/2007 21:00	114	58	15	0.95	100	0									**E**
04/01/2007 22:00	114	52	14	0.95	100	0									**E**
04/01/2007 23:00	109	53	20	0.95	100	0									**E**
Day 8															
05/01/2007 00:00	112	61	20	0.95	100	0									**E**
05/01/2007 01:00	108	57	17	0.95	100	10									**E**
05/01/2007 02:00	107	51	16	0.95	98	0			250						**E**
05/01/2007 03:00	104	53	22	0.95	96									0.8	**E**
05/01/2007 04:00	109	62	19	0.95	96	0								0.4	**E**
05/01/2007 05:00	113	59	19	0.95	97	0								0.6	**E**
05/01/2007 06:00	117	64	17	1	97	0								0.6	**E**
05/01/2007 07:00	116	65	17	1	97	0								0.7	**E**
05/01/2007 08:00	115	59	17	1	96	0								0.7	**E**
05/01/2007 09:00	120	63	16	1	95	0								0.7	**E**
05/01/2007 10:00	116	58	13	1	93	0								0.7	**E**
05/01/2007 11:00	117	55	18	1	97	0					Dialysis			1.2	**E**
05/01/2007 12:00	123	81	18	1	98	15					Dialysis			0.8	**E**
05/01/2007 12:12	117	55	18	1	97										**E**
05/01/2007 13:00	117	65	17	0.95	95	0					Dialysis			1	**E**
05/01/2007 14:00	116	68	17	0.95	98	0					Dialysis			1	**E**
05/01/2007 15:00	115	66	17	0.9	96	15					Dialysis			1.1	**E**
05/01/2007 16:00	116	67	17	0.9	96						Dialysis			1.1	**E**
05/01/2007 17:00	113	65	17	0.9	97	0					Dialysis			1.1	**E**
05/01/2007 18:00	117	62	18	0.95	97									1.1	**E**
05/01/2007 19:00	116	60	17	0.95	96									1.1	**E**
05/01/2007 20:00	117	64	18	0.95	96	0								1.1	**E**
05/01/2007 21:00	118	61		0.95	96	20								1.3	**E**
05/01/2007 22:00	119	53	0	0.95	94	0								2	**E**
05/01/2007 23:00	122	62		0.95	95	0								2	**E**

Day 9															
06/01/2007 00:00	124	75	-2	0.95	93	0								2	**E**
06/01/2007 01:00	116														

Appendix VII

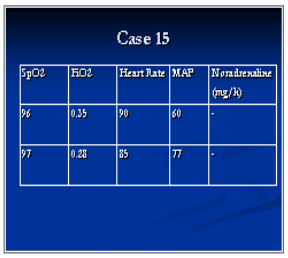
Database for selection of pairs for two step change experiment

Pair type	First of group	Last of group	# of pairs	Base-Line-in-Spreadsheet	Rand Num	(Ordered) Spreadsheet row number	**INSIGHT Case Num**	Batch	Status	Example
CA	846	853	8	846	1	846	**594**	1		**1**
CA				846	2	847	**628**	1		**2**
CA				846	3	848	**4049**	1		**3**
CA				846	4	849	**4251**	1		**4**
CA				846	5	850	**5040**	1		**5**
CA				846	6	851	**6462**	1		**6**
CA				846	7	852	**6507**	1		**7**
CA				846	8	853	**6688**	1		Not used
DB	2043	2074	32	2043	17	2059	**2976**	1		**8**
DB				2043	9	2051	**867**	1		**9**
DB				2043	23	2065	**3984**	1		**10**
DB				2043	10	2052	**1380**	1		Missing
DB				2043	16	2058	**2059**	1		Missing
DB				2043	30	2072	**6257**	1		**11**
DB				2043	28	2070	**5637**	1		**12**
DB				2043	18	2060	**3464**	1		**13**
DB				2043	12	2054	**1428**	1		Missing
DB				2043	20	2062	**3898**	1		Missing
DB				2043	13	2055	**1476**	1		Missing
DB				2043	22	2064	**3966**	1		Missing
DB				2043	3	2045	**582**	1		**14**
EC	5081	5095	15	5081	12	5092	**5871**	1		**15**
EC				5081	14	5094	**5891**	1		Missing
EC				5081	10	5090	**5494**	1		Missing
EC				5081	2	5082	**1255**	1		Missing
EC				5081	3	5083	**2088**	1		Missing
EC				5081	13	5093	**5874**	1		**16**
EC				5081	4	5084	**2403**	1		Missing
EC				5081	1	5081	**742**	1		**17**
EC				5081	7	5087	**4992**	1		Missing
EC				5081	15	5095	**6216**	1		**18**
EC				5081	5	5085	**2417**	1		Missing
EC				5081	6	5086	**4906**	1		**19**
EC				5081	11	5091	**5720**	1		Missing
AC	171	183	13	171	1	171	**589**	1		Missing but could extrapolate)
AC				171	2	172	**593**	1		ok
AC				171	3	173	**638**	1		ok
AC				171	4	174	**3478**	1		ok
AC				171	5	175	**4050**	1		ok
AC				171	6	176	**4232**	1		ok
AC				171	7	177	**4271**	1		ok
AC				171	8	178	**4276**	1		ok
AC				171	9	179	**5039**	1		ok
AC				171	10	180	**5064**	1		ok

AC				171	11	181	**5083**	1		ok
AC				171	12	182	**5091**	1		ok
AC				171	13	183	**6526**	1		ok
BD	805	839	35	805	29	833	**5634**	1		ok
BD				805	22	826	**3967**	1		Missing
BD				805	17	821	**3318**	1		Missing
BD				805	18	822	**3341**	1		Missing
BD				805	6	810	**630**	1		ok
BD				805	11	815	**1421**	1		Missing
BD				805	1	805	**11**	1		ok
BD				805	2	806	**445**	1		ok
BD				805	30	834	**5636**	1		Missing
BD				805	28	832	**5632**	1		ok
BD				805	10	814	**876**	1		ok
BD				805	23	827	**4042**	1		ok
BD				805	14	818	**1854**	1		Missing
CE	2014	2035	22	2014	3	2016	**619**	1		ok
CE				2014	13	2026	**4977**	1		Missing
CE				2014	6	2019	**3081**	1		Missing
CE				2014	10	2023	**4905**	1		ok
CE				2014	22	2035	**6776**	1		Missing
CE				2014	8	2021	**4880**	1		ok
CE				2014	18	2031	**5745**	1		Missing
CE				2014	11	2024	**4956**	1		ok
CE				2014	1	2014	**1**	1		Missing
CE				2014	21	2034	**6309**	1		Missing
CE				2014	9	2022	**4885**	1		Missing
CE				2014	17	2030	**5721**	1		Missing
CE				2014	15	2028	**5224**	1		Missing

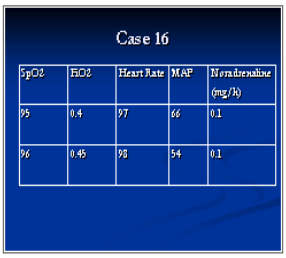
Appendix VIII

Slide show for ICU-PSS 2 step change experiment


Case 15

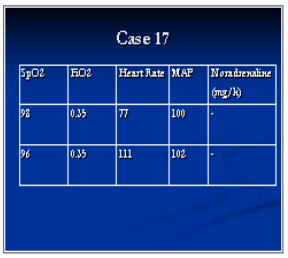
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
96	0.35	90	60	-
97	0.28	85	77	-

17


Case 16

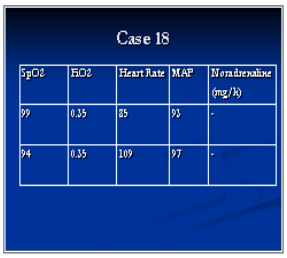
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
95	0.4	97	66	0.1
96	0.45	98	54	0.1

18


Case 17

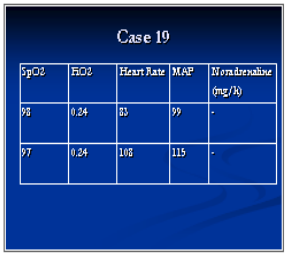
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
98	0.35	77	100	-
96	0.35	111	102	-

19


Case 18

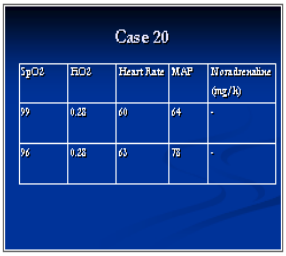
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
99	0.35	85	93	-
94	0.35	109	97	-

20


Case 19

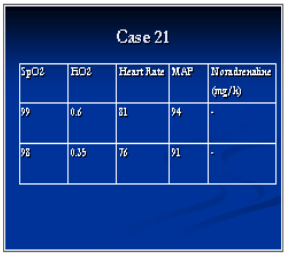
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
98	0.24	83	99	-
97	0.24	108	115	-

21


Case 20

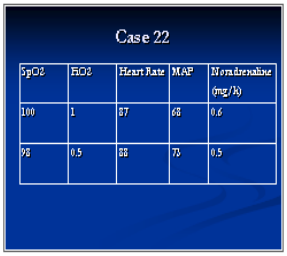
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
99	0.28	60	64	-
96	0.28	63	78	-

22


Case 21

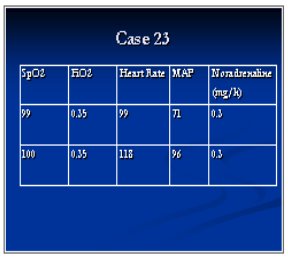
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
99	0.6	81	94	-
98	0.35	76	91	-

23


Case 22

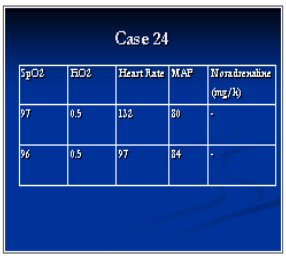
SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
100	1	87	68	0.6
98	0.5	88	73	0.5

24


Case 23

SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
99	0.35	99	71	0.3
100	0.35	118	96	0.3

25


Case 24

SpO2	FiO2	Heart Rate	MAP	Noradrenaline (mg/h)
97	0.5	132	80	-
96	0.5	97	84	-

26

Appendix IX

Scoring sheet for ICU-PSS 2 step experiment

ICU Patient Scoring System

Please mark with an X in the appropriate box whether in your opinion the patient has improved or deteriorated

Case Number	Improved	Deteriorated
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		

Appendix X

Answer template for ICU-PSS 2 step experiment

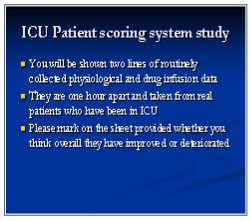
ICU Patient Scoring System

Answer template

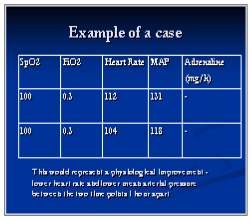
Case Number	Improved	Deteriorated	Actual Change
1	X		E - C
2	X		D - B
3		X	C - E
4		X	A - C
5	X		E - C
6	X		C - A
7		X	A - C
8		X	C - E
9	X		E - C
10	X		D - B
11		X	B - D
12	X		D - B
13		X	C - E
14		X	C - E
15	X		C - A
16		X	B - D
17		X	B - D
18		X	A - C
19		X	A - C
20	X		C - A
21	X		C - A
22	X		E - C
23		X	B - D
24	X		D - B

Appndix XI

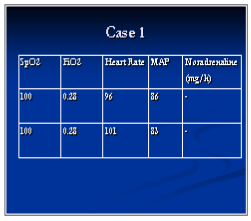
Slide show for ICU-PSS 1 step change experiment



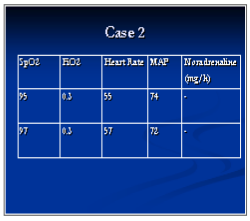
1



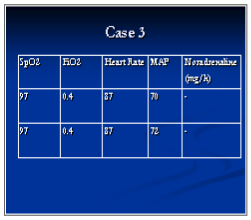
2



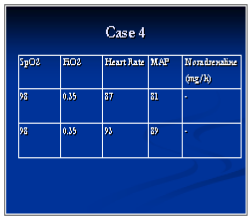
3



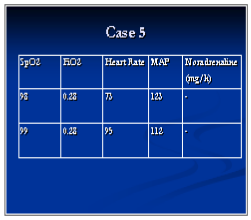
4



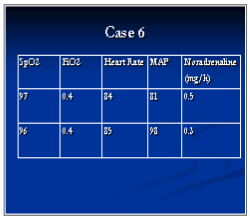
5



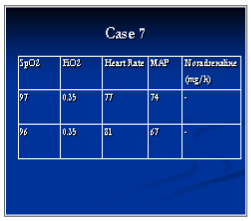
6



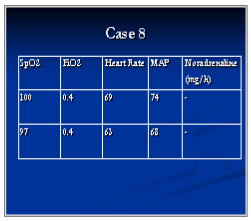
7



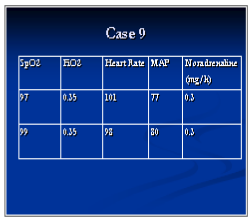
8



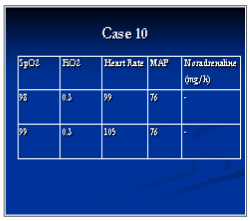
9



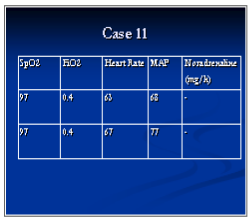
10



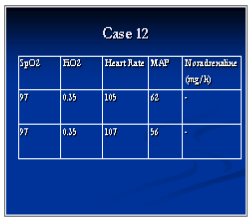
11



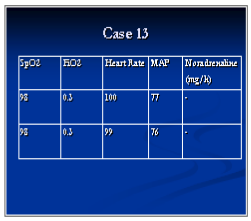
12



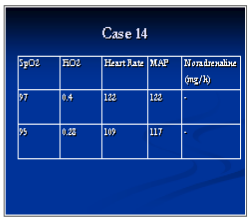
13



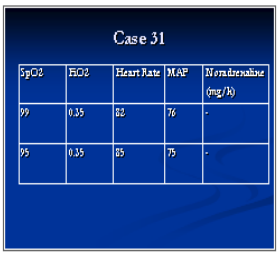
14



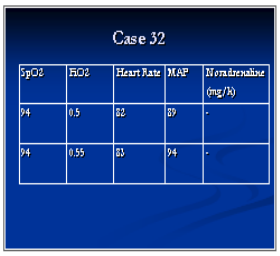
15



16



33



34

Appendix XII

Scoring sheet for ICU-PSS 1 step experiment

ICU Patient Scoring System: Please mark with an X in the appropriate box whether in your opinion the patient has improved or deteriorated

Case Number	Improved	Deteriorated
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		

Appendix XIII

Answer template for ICU-PSS 1 step experiment

Answer template for ICU-PSS one step change experiment

Case Number	Improved	Deteriorated	Actual Change
1		x	B-C
2	x		B-A
3	x		B-A
4		x	A-B
5	x		D-C
6	x		C-B
7		x	A-B
8		x	A-B
9	x		C-B
10		x	B-C
11	x		B-A
12		x	C-D
13	x		C-B
14	x		D-C
15	x		E-D
16	x		B-A
17	x		D-C
18		x	C-D
19	x		E-D
20		x	D-E
21		x	D-E
22		x	C-D
23		x	D-E
24	x		E-D
25		x	B-C
26	x		E-D
27	x		C-B
28		x	C-D
29	x		D-C
30		x	D-E
31		x	A-B
32		x	B-C

Appendix XIV

Reasons for rejection of pairs of data used in the 2 step changes experiment (with both Insight Case Numbers)

High Level Summary

Most of the rejected pairs are because apparent improvement or deterioration was caused by missing parameter(s) rather than actual improvement or deterioration.

I also for the purposes of this experiment required a minimum "core" set of parameters. These were FiO2, heart rate, MAP and SpO2. After the 1 stage change experiment is complete we can examine the clinician's ability to discriminate using less than this core set.

There were also two examples in the selected pairs in the B-D two step change which in fact from the spread sheet were D-D, so I ignored these.

For ease I have made a table of each rejected pair rather than you having to scroll through the larger data set each time.

Insight Case Number 1380-1381

FiO2	HR	MAP	Norad	SpO2
	134	61	1.2	95
0.5				95

Prediction is D-B, but apparent improvement is due to missing HR, MAP, and Noradrenaline/

Insight Case Number 2059-2060

FiO2	HR	MAP	Norad	SpO2
0.45	121	61		98
0.5				

Prediction is D-B, but apparent improvement is due to missing HR, MAP and SpO2/

Insight Case Number 1428-1429

FiO2	HR	MAP	Norad	SpO2
0.5	99	79	1.8	98
0.5	99	79		98

Prediction is D-B, but apparent improvement is due to missing noradrenaline.

Insight Case Number 3898-3899

FiO2	HR	MAP	Norad	SpO2
0.5	77	81	1.2	98
0.5	77	81		98

Prediction is D-B, but apparent improvement is due to missing noradrenaline.

Insight Case Number 1476-1477

FiO2	HR	MAP	Norad	SpO2
0.5	95	68	1.8	100
0.5	95	68		100

Prediction is D-B, but apparent improvement is due to missing noradrenaline.

Insight Case Number 3966-3967

FiO2	HR	MAP	Norad	SpO2
0.8	93	66	0.2	92
	90			

Prediction is D-B, but apparent improvement is due to missing FiO2, MAP, noradrenaline and SpO2.

Insight Case Number 5891-5892

Adrenaline	FiO2	HR	MAP	SpO2
1.4	0.4	109	63	99
	0.4	109	63	99

Prediction is E-C, but apparent improvement is due to missing Adrenaline

Insight Case Number 5494-5495

FiO2	HR	MAP	Noradrenaline	SpO2
0.7	109	70	1.8	98
	105	65		97

Prediction is E-C, but apparent improvement is due to missing FiO2 and Noradrenaline

Insight Case Number 1255-1256

FiO2	HR	MAP	Noradrenaline	SpO2
0.4	123	66	0.8	4
0.4	108	65	0.7	97

Prediction is E-C, but apparent improvement is due to a typo in the entry for SpO2

Insight Case Number 2088-2089

FiO2	HR	MAP	Noradrenaline	SpO2
1	104	87		100
	104	87		100

Prediction is E-C, but apparent improvement is due to missing FiO2

Insight Case Number 2403-2404

FiO2	HR	MAP	Noradrenaline	SpO2
0.45	156	61		98
0.45			0.8	98

Prediction is E-C, but apparent improvement is due to missing heart rate and MAP

Insight Case Number 4992-4993

FiO2	HR	MAP	Noradrenaline	SpO2
	95	130		100
	88	118		97

Prediction is E-C, which is probably correct but for the purposes of the experiment I wanted to be consistent and required a minimum data set of FiO2, HR, MAP and SpO2. After completion of the one step change experiment then we can focus on two and one step changes with fewer core parameters.

Insight Case Number 2417-2418

FiO2	HR	MAP	Noradrenaline	SpO2
0.45	145	78	0.8	95
			0.9	

Prediction is E-C, but apparent improvement is due to missing FiO2, HR, MAP and SpO2.

Insight Case Number 5720-5721

FiO2	HR	MAP	Noradrenaline	SpO2
0.65	109	69	2.8	94
0.65	109	69		94

Prediction is E-C, but apparent improvement is due to missing noradrenaline.

Insight Case Number 589-590

FiO2	HR	MAP	Noradrenaline	SpO2
0.35	87	98		98
	107	86		99

Prediction is A-C, but apparent FiO2 was missing and I wanted a minimum core data set (in this case it could have been extrapolated as it was 0.35 for several hours before and after).

Insight Case Number 3967-3968

FiO2	HR	MAP	Noradrenaline	SpO2
	90			
0.8	95	72	0.2	95

Prediction is B-D, but apparent deterioration is due to missing FiO2, MAP, noradrenaline and SpO2.

Insight Case Number 3318-3319

FiO2	HR	MAP	Noradrenaline	SpO2
0.35	126			97
0.35	127			100

Prediction is D-D, for some reason this appeared in the selected pairs document as a B-D so I moved on to the next one.

Insight Case Number 3341-3342

FiO2	HR	MAP	Noradrenaline	SpO2
0.3	131			100
0.3	134			100

Prediction is D-D, for some reason this appeared in the selected pairs document as a B-D so I moved on to the next one.

Insight Case Number 1421-1422

FiO2	HR	MAP	Noradrenaline	SpO2
0.4	90	74		97
0.4	98	82	1.8	97

Prediction is B-D, but apparent deterioration is due to missing noradrenaline.

Insight Case Number 5636-5637

FiO2	HR	MAP	Noradrenaline	SpO2
0.35	99	71		99
0.35	113	88	0.3	99

Prediction is B-D, but apparent deterioration is due to missing noradrenaline, but could have probably been extrapolated as was 0.3 for several hours before.

Insight Case Number 1854-1855

FiO2	HR	MAP	Noradrenaline	SpO2
0.4			0.1	99
0.4	122	69	0.4	98

Prediction is B-D, but apparent deterioration is due to missing heart rate and MAP.

Insight Case Number 4977-4978

FiO2	HR	MAP	Noradrenaline	SpO2
	101	118		100
	103	144		100

Prediction is C-E which is probably correct but the lack of FiO2 meant there was not the core data set.

Insight Case Number 3081-3082

FiO2	HR	MAP	Noradrenaline	SpO2
0.3	108			97
0.3	103	50	0.2	98

Prediction is C-E, but apparent deterioration is due to lack of MAP and noradrenaline.

Insight Case Number 6776-6777

Adrenaline	FiO2	HR	MAP	SpO2
0.4	0.1	95	88	97
	0.9	102	75	98

Prediction is C-E, but apparent deterioration is due to lack of adrenaline. It had only been started one hour before this so would be difficult to extrapolate.

Insight Case Number 5745-5746

FiO2	HR	MAP	Noradrenaline	SpO2
0.55	101			
1	98	73	1.7	97

Prediction is C-E, but apparent deterioration is due to lack of MAP, noradrenaline and SpO2.

Insight Case Number 1-2

FiO2	HR	MAP	Noradrenaline	SpO2
	100	105		100
1	85	145		100

Prediction is C-E, but apparent deterioration is due to lack of FiO2.

Insight Case Number 6309-6310

FiO2	HR	MAP	Noradrenaline	SpO2
0.55	84	71		100
	170			

Prediction is C-E, although real due to increase in heart rate there are missing core parameters of FiO2, MAP and SpO2.

Insight Case Number 4885-4886

FiO2	HR	MAP	Noradrenaline	SpO2
0.3	101			100
0.3	118	142		100

Prediction is C-E, although real due to increase in heart and high MAP there are missing core parameters i.e. MAP

Insight Case Number 5721-5722

FiO2	HR	MAP	Noradrenaline	SpO2
0.65	109	69		94
0.65	110	69	2.8	97

Prediction is C-E, but apparent deterioration is due to lack of noradrenaline which had been quite fluctuant before this so would have been hard to extrapolate.

Insight Case Number 5224-5225

FiO2	HR	MAP	Noradrenaline	SpO2
0.5	105			94
0.5	103	0		95

Prediction is C-E, but there is no MAP so missing core parameters.

MS 20/5/12

Appendix XV

Data base created to examine ICU-PSS and outcome

Table for outcomes

Patient-ID	Start/Fin	Outcome	APACHE II	Predicted Mortality	Med Diag.
1536	13	Alive	29	77.2	Medical
1667	1697	Dead	33	84.5	Medical
1689	1861	Alive	22	58.9	Medical
1695	18614	Dead	31	75.4	Medical
1697	1898	Alive	24	35.5	Medical
1711	18654	Dead	20	35.5	Medical
1720	2214	Alive	28	63.7	Surgical
1721	2677	Alive	18	44.4	Surgical
1726	2750	Alive	17	26.2	Surgical
1727	2784	Dead	29	69.6	Medical
1742	18666	Alive	16	23.3	Surgical
1748	3027	Alive	23	62.3	Surgical
1750	3530	Dead	18	29.1	Medical
1757	3681	Alive	22	14	Medical
1774	3711	Dead	22	42.4	Medical
1781	18719	Alive	17	26.2	Surgical
1807	4383	Alive	32	78	Surgical
1818	5142	Alive	19	48	Surgical
1822	5251	Dead	22	42.4	Medical
1933	8950	Alive	20	21.3	Surgical
1948	9134	Dead	17	36.2	Medical
1951	9146	Dead	22	58.9	Surgical
1965	9366	Dead	21	41.6	Surgical
1969	9453	Alive	30	49.3	Medical
1970	19258	Alive	14	18.6	Surgical
2017	9655	Alive	23	58.5	Surgical
2030	9916	Alive	38	92.6	Surgical
2039	10444	Dead	33	84.7	Medical
2119	19278	Dead	21	55.3	Surgical
2121	10878	Dead	23	58.5	Surgical
2138	10917	Alive	25	47.1	Surgical
2158	11609	Dead	20	47.6	Medical
2174	11817	Dead	9	9.9	Medical
2188	12049	Dead	45	97.2	Medical
2189	12082	Dead	31	68.1	Medical
2220	19615	Alive	12	14.6	Surgical
2231	12315	Alive	29	79.9	Surgical
2265	12573	Dead	35	79.3	Surgical
2273	13707	Alive	34	66.6	Medical

2284	14147	Dead	18	19.7	Surgical
2303	14486	Alive	32	82.5	Medical
2313	15172	Dead	11	8.1	Surgical
2327	20729	Dead	41	93	Surgical
2328	15973	Dead	43	86.7	Medical
2342	16220	Alive	30	72.6	Surgical
2457	16457	Alive	25	70	Surgical
2506	16789	Alive	36	55.8	Medical
2524	16970	Alive	28	72.7	Medical
2527	17138	Alive	24	65.7	Surgical
2585	17353	Dead	34	82.6	Medical
2607	17411	Dead	21	38.9	Medical
2644	17515	Alive	32	56.6	Surgical
2660	18309	Alive	25	53.1	Surgical
2698	18562	Dead	33	85.8	Medical

Appendix XVI

Data base created to examine troponin level and outcome

Date Admit	Date Out	L.O.S.	Age	M/F	Apache II	Pred. Mort.	1st Troponin	Day 1st troponin	Highest Troponin	Day Highest Troponin	Alive (1)/ Dead (0)
10/10/2009	19/11/2009	40.6	72	M	Readmission	Readmission	0.33	2	6.47	24	0
09/10/2009	14/10/2009	4.5	76	M	Readmission	Readmission	0.13	1	0.13	1	1
08/10/2009	09/10/2009	0.9	33	M	11	6.3	<0.04	Y	<0.04	Y	1
08/10/2009	12/10/2009	3.6	35	F	19	27	0.27	2	0.27	2	1
08/10/2009	14/10/2009	6.1	43	F	33	75.6	<0.04	Y	<0.04	Y	0
06/10/2009	08/10/2009	1.5	77	F	16	23.3	23.8	1	23.8	1	1
06/10/2009	07/10/2009	1.1	35	M	7	7.6	0.16	2	0.16	2	1
05/10/2009	06/10/2009	0.9	64	F	16	23.5	<0.04	Y	<0.04	Y	1
04/10/2009	08/10/2009	4	57	F	16	1.1	<0.04	Y	<0.04	Y	1
03/10/2008	02/11/2009	30	59	M	19	12.2	0..04	1	0.1	2	1
03/10/2009	04/10/2009	1.1	41	M	27	75.9	0.36	2	0.36	2	0
02/10/2009	09/10/2009	6.6	58	F	17	19.7	1.57	2	1.57	2	1
02/10/2009	08/10/2009	5.7	72	M	11	17.4	0.25	4	0.25	4	1
02/10/2009	04/10/2009	2	59	M	33	85.8	0.08	1	0.31	3	0
02/10/2009	02/10/2009	0.5	36	F	7	0.3	<0.04	Y	<0.04	Y	1
29/09/2009	09/10/2009	9.6	33	F	Readmission	Readmission	<0.04	Y	<0.04	Y	1
29/09/2009	01/10/2009	2.4	40	M	4	2.9	<0.04	Y	<0.04	Y	1
28/09/2009	29/09/2009	1.2	27	M	23	63.7	0.07	1	0.07	1	1
27/09/2009	29/09/2009	1.8	75	M	17	25.9	<0.04	Y	<0.04	Y	1
25/09/2009	25/09/2009	0.2	25	F	<8 hours	< 8 hours	N	N	N	N	1
24/09/2009	26/09/2009	1.9	82	M	16	23.5	0.08	2	0.08	2	1
24/09/2009	25/09/2009	1.2	47	M	27	75.9	N	N	N	N	0
22/09/2009	02/10/2009	9.8	73	M	25	53.1	0.07	2	0.23	4	1
22/09/2009	26/09/2009	3.9	69	M	20	35.5	0.88	2	0.88	2	1
26/09/2009	27/09/2009	0.7	59	M	7	4.7	N	N	N	N	1
21/09/2009	22/09/2009	1.5	72	M	15	28.2	0.7	1	4.48	1	1
20/09/2009	22/09/2009	1.9	16	M	23	45.7	N	N	N	N	1
20/09/2009	22/09/2009	1.9	22	M	20	1.9	1.28	1	1.28	1	1
20/09/2009	21/09/2009	0.9	60	F	18	1.4	<0.04	Y	<0.04	Y	1
19/09/2009	23/09/2009	3.9	33	F	11	12.9	<0.04	Y	<0.04	Y	1
19/09/2009	19/09/2009	0.6	84	F	25	35.9	N	N	N	N	0
14/09/2009	16/09/2009	1.7	71	M	23	46	0.16	3	0.16	3	1
14/09/2009	02/10/2009	18.6	76	M	20	35.5	0.47	5	0.47	5	1
13/09/2009	15/09/2009	1.7	39	F	23	45.7	1.03	1	1.03	1	1
13/09/2009	13/09/2009	0.2	44	M	< 8 hours	< 8 hours	N	N	N	N	0
11/09/2009	14/09/2009	2.9	53	M	23	41.1	0.29	4	0.29	4	1
10/09/2009	18/09/2009	8.2	75	F	25	68.9	0.07	2	0.08	3	1
09/09/2009	10/09/2009	1.2	78	M	39	92.9	0.43	1	0.43	1	1
08/09/2009	10/09/2009	1.9	50	F	17	16.1	<0.04	Y	<0.04	Y	1
07/09/2009	10/09/2009	3	63	M	13	14	<0.04	Y	<0.04	Y	1
07/09/2009	08/09/2009	1.6	75	F	36	83.2	6.3	1	6.8	1	0
06/09/2009	06/09/2009	0.1	24	M	<8 hours	< 8 hours	N	N	N	N	1
06/09/2009	06/09/2009	0.3	79	M	16	23.4	0.31	1	0.31	1	1
06/09/2009	07/09/2009	1.1	26	M	11	23.4	<0.04	Y	<0.04	Y	1
04/09/2009	09/09/2009	4	63	M	21	38.9	2.91	4	2.91	4	0
04/09/2009	05/09/2009	0.6	35	M	10	5.6	N	N	N	N	1
04/09/2009	09/09/2009	5.1	60	M	21	35.6	2	4	2	4	1
03/09/2009	04/09/2009	1.4	61	F	19	32.2	<0.04	Y	<0.04	Y	1

03/09/2009	06/09/2009	3.8	78	F	24	52.5	0.27	1	0.27	1	0
03/09/2009	04/09/2009	1.8	65	F	23	41.3	0.04	2	0.04	2	1
02/09/2009	03/09/2009	0.4	70	F	37	86.7	N	N	N	N	0
01/09/2009	01/09/2009	0.4	47	F	34	82.6	N	N	N	N	0
31/08/2009	01/09/2009	0.7	47	F	19	1.6	<0.04	Y	<0.04	Y	1
31/08/2009	07/09/2009	7	50	F	Readmission	Readmission	<0.04	Y	<0.04	Y	1
30/08/2009	03/09/2009	4.4	55	F	25	68.9	<0.04	Y	<0.04	Y	1
29/08/2009	31/08/2009	2.1	51	M	34	82.6	3.18	3	3.18	3	0
27/08/2009	29/08/2009	2.3	48	F	17	21.7	0.48	2	0.48	2	1
25/08/2009	27/08/2009	2.3	78	M	24	49.3	0.05	1	0.12	2	0
24/08/2009	16/09/2009	22.7	49	M	Burn	Burn	0.1	15	0.1	15	1
24/08/2009	28/08/2009	4	67	F	15	34	<0.04	Y	<0.04	Y	1
24/08/2009	28/08/2009	4	50	F	Burn	Burn	<0.04	Y	<0.04	Y	1
23/08/2009	26/08/2009	3	75	M	15	34	<0.04	Y	<0.04	Y	1
23/08/2009	04/09/2009	12.2	76	M	21	14	0.19	1	1.54	2	1
22/08/2009	24/08/2009	1.5	42	M	17	34.5	2.13	2	2.13	2	1
22/08/2009	28/08/2009	5.7	16	F	9	3.6	<0.04	Y	<0.04	Y	1
22/08/2009	23/08/2009	1.2	15	M	15 years old	15 years old	N	N	N	N	1
21/08/2009	22/08/2009	0.7	67	F	14	9.4	0.19	2	0.19	2	1
21/08/2009	30/08/2009	8.7	44	M	15	30.4	<0.04	Y	<0.04	Y	1
21/08/2009	21/08/2009	0.2	50	M	<8 hours	<8 hours	N	N	N	N	0
21/08/2009	22/08/2009	1.4	51	F	10	19.9	<0.04	Y	<0.04	Y	1
20/08/2009	24/08/2009	3.8	59	M	Readmission	Readmission	0.07	2	0.07	2	1
19/08/2009	21/08/2009	1.7	63	M	13	13.4	<0.04	Y	<0.04	Y	1
19/08/2009	25/08/2009	5.9	70	F	17	37	<0.04	Y	<0.04	Y	0
18/08/2009	21/08/2009	3.2	40	M	24	65.7	2	2	2	2	1
18/08/2009	19/08/2009	1	52	M	27	60.1	N	N	N	N	0
16/08/2009	20/08/2009	4.2	48	F	Burn	Burn	<0.04	Y	<0.04	Y	0
14/08/2009	15/08/2009	0.5	36	M	19	16.4	0.8	2	0.8	2	1
14/08/2009	14/08/2009	0.7	23	M	9	3.6	0.61	1	0.61	1	1
13/08/2009	06/09/2009	23.9	31	F	26	64.9	0.06	1	0.06	1	1
13/08/2009	23/08/2009	9.7	68	F	27	46	0.16	2	0.16	2	1
13/08/2009	17/08/2009	3.9	33	M	12	9.3	<0.04	Y	<0.04	Y	1
13/08/2009	19/08/2009	6.2	59	M	28	72.7	0.07	2	0.07	2	1
10/08/2009	22/08/2009	12.6	54	F	16	23.2	<0.04	Y	<0.04	Y	0
09/08/2009	12/08/2009	2.4	72	F	26	66.2	0.07	1	0.11	2	0
08/08/2009	11/08/2009	2.8	29	M	21	2.2	0.74	3	0.74	3	1
08/08/2009	08/08/2009	0.5	61	M	12	8.7	0.13	1	0.22	1	1
07/08/2009	08/08/2009	0.4	40	F	3	2.4	N	N	N	N	1
06/08/2009	15/08/2009	8.6	40	M	21	51.2	<0.04	Y	<0.04	Y	1
06/08/2009	12/08/2009	6.3	52	F	36	55.8	2.59	2	2.59	2	1
05/08/2009	06/08/2009	0.6	58	F	20	38.1	N	N	N	N	1
05/08/2009	10/08/2009	5.3	39	M	35	89	<0.04	Y	<0.04	Y	1
03/08/2009	08/08/2009	5.7	75	M	34	86.3	0.25	2	0.63	3	0
02/08/2009	06/08/2009	3.5	36	M	13	24.6	0.07	2	0.07	2	1
02/08/2009	05/08/2009	2.6	37	M	19	28.7	0.11	2	0.11	2	1
31/07/2009	01/08/2009	0.8	63	M	9	9.9	<0.04	Y	<0.04	Y	1
31/07/2009	06/08/2009	5.5	57	M	15	21	<0.04	Y	<0.04	Y	1
31/07/2009	01/08/2009	1	79	F	33	87.7	1.32	2	1.32	2	0
30/07/2009	02/08/2009	2.9	64	F	19	17.7	0.16	2	0.16	2	1
30/07/2009	20/08/2009	21.2	71	F	20	35.5	<0.04	Y	<0.04	Y	0
30/07/2009	31/07/2009	1.4	68	M	21	2.2	<0.04	Y	<0.04	Y	1

References

1. Vincent J.L., Artigas A. and Bihari D. Guidelines for the utilisation of intensive care units. *Intensive Care Medicine* 1994; 20: 163-164.
2. Day V. on behalf of the Department of Health. Comprehensive Critical Care. A review of adult critical care services. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4082872.pdf
3. Intensive Care Society. Levels of Critical Care for Adult Patients (Standards and Guidelines), 2009. http://www.ics.ac.uk/professional/standards_and_guidelines/levels_of_critical_care_for_adult_patients
4. The Critical Care Minimum Data Set (CCMDS). Department of Health 2006. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_116368
5. Lassen H.C.A. A preliminary report on the 1952 epidemic of poliomyelitis in Copenhagen with specific reference to the treatment of respiratory insufficiency. *Lancet* 1953; 1: 37-41.
6. Ibsen B. The anaesthetist's viewpoint on the treatment of respiratory complications in poliomyelitis during the epidemics in Copenhagen, 1952. *Proc R Soc Med* 1954; 47:72–74.
7. Reisner-Sénélar L. The birth of intensive care medicine: Björn Ibsen's records. *Intensive Care Med* 2011; 37(7):1084-86.
8. Eggert S.M. and Jerwood C. Percutaneous tracheostomy. *British Journal of Anaesthesia CEPD Reviews* 2003; 8(5): 139-142.
9. Reynolds L.A. and Tansey E.M. History of British Intensive Care, c.1950 – c.2000. *Wellcome Witness to Twentieth Century Medicine*, 2011; 42.
10. Progressive Patient Care: Interim report of a departmental working group. Monthly Bulletin MOH and PHLS 1962; 21: 218-26.
11. Scottish Intensive Care Society Audit Group. *Audit of Critical Care in Scotland 2013 Reporting on 2012*. http://www.sicsag.scot.nhs.uk/Publications/SICSAG-report-2013-web.pdf?1
12. Organisation for Economic Co-operation and Development Health data 2012. http://www.oecd.org/unitedkingdom/BriefingNoteUNITEDKINGDOM2012.pdf
13. Adhikari N., Fowler R., Bhagwanjee S. and Rubenfeld G. Critical care and the global burden of critical illness in adults. *Lancet* 2010; 376:1339-1346.

14. Personal communication with Mrs Marion MacDonald, Clinical Services Manager, Western Infirmary Glasgow.
15. Scottish Intensive Care Society Audit Group. http://www.sicsag.scot.nhs.uk/
16. Information Services Division Scotland. http://www.isdscotland.org/
17. Apgar V. A proposal for a New Method of Evaluation of the Newborn Infant. *Current Researches in Anaesthesia and Analgesia* 1953; 32(4): 260-7.
18. Ranson J.H., Rifkind K.M., Roses D.F., Fink S.D., Eng K. and Spencer F.C. Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology and Obstetrics* 1974; 139(1); 69-81.
19. Teasdale GM and Jennett B. Assessment of coma and impaired consciousness. *Lancet* 1974; 2:81-84.
20. Bouch D.C. and Thompson J.P. Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia, Critical Care & Pain* 2000;8(5):181-185.
21. Knaus W.A., Draper E.A., Wagner D.P. and Zimmerman J.E. APACHE II: a severity of disease classification system. *Critical Care Medicine* 1985; 13(10):818-29.
22. Knaus W.A., Zimmerman J.E., Wagner D.P. et al. APACHE - acute physiology and chronic health evaluation: A physiologically based classification system. *Crit Care Med* 1981; 9:951.
23. Knaus W.A., Le Gall J.R., Wagner D.P. et al. Evaluating outcome from intensive care: A preliminary multihospital comparison. *Crit Care Med* 1982; 10:491.
24. Gustafson D.H., Fryback D., Rose J. et al. A decision theoretic methodology for severity index development. *Med Decis Making* 1986; 6(1): 27-35.
25. Knaus W.A., Wagner D.P., Draper E.A. et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalised adults. *Chest* 1991; 100:1619-1636.
26. Woods A.W., MacKirdy F.N., Livingstone B.M., Norrie J. and Howie J.C. Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system. *Anaesthesia* 2000; 55:1058-65.
27. Livingston B.M., Mackirdy F.N., Howie J.C., Jones R. and Norrie J.D. Assessment of the performance of five intensive care scoring models within a large Scottish database. *Crit Care Med* 2000; 28(6):1820-1827.
28. Recalibration of APACHE II. Scottish Intensive Care Society Audit Group. http://www.sicsag.scot.nhs.uk/Index.html.

29. Le Gall J.R., Loirat P., Alperovitch A., Glaser P., Granthil C., Mathieu D., Mercier P., Thomas R. and Villers D. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12(11):975-7.
30. Le Gall J.R., Lemeshow S and Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American Multicenter study. *JAMA* 1993; 270(24):2957-63.
31. Lemeshow S., Teres D., Pastides H., Avrunin J.S. and Steingrub J.S. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Critical Care Medicine* 1985; 13(7):519-25.
32. Lemeshow S., Teres D., Klar J., Avrunin J.S., Gehlbach S.H. and Rapoport J. Mortality Probability Models (MPM II) based upon an international cohort of intensive care unit patients. *Journal of the American Medical Association* 1993; 270: 2478-86.
33. Copeland G.P., Jones D. and Walters M. POSSUM: a scoring system for surgical audit. *Br. J. Surg.* 1991; 78(3):355-60.
34. Prytherch D.R., Whiteley M.S., Higgins B., Weaver P.C., Prout W.G. and Powell S.J. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enumeration of mortality and morbidity. *Br. J. Surg.* 1998; 85(9):1217-1220.
35. Cullen D.J., Civetta JM, Briggs BA and Ferrara LC. Therapeutic Intervention Scoring System: a method for quantitative comparison of patient care. *Crit Care Med* 1974; 2:57-60.
36. Reis M.D., de Rijk A. and Schaufeli W. Simplified Intervention Scoring System: The TISS-28 items – Results from a multicenter study. *Crit. Care. Med.* 1996; 24: 64-73.
37. Lefering R., Zart M. and Neugebauer E.A.M. Retrospective evaluation of the simplified Therapeutic Intervention Scoring System (TISS-28) in a surgical intensive care unit. *Intensive Care Medicine* 2000; 26(12): 1794-1802.
38. Vincent J.L., Moreno R., Takala J. et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; 22:707-710.
39. Marshall J.C., Cook D., Nicolas Christou et al. Multiple Organ Dysfunction Score: A reliable descriptor of a complex clinical outcome. *Crit. Care Med.* 1995; 23(10): 1638-1652.

40. Bota D.P., Melot C., Ferreira F.L., Ba V.N. and Vicent J.L. The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure (SOFA) score in outcome prediction. *Intensive Care Med* 2002; 28:1619-1624.
41. Boyd C.R., Tolson M.A. and Copes W.S. Evaluating Trauma Care: The TRISS method. *J. Trauma* 1987; 27: 370-378.
42. Lim W.S., van der Eerden M.M., Laing R. et al. Defining community acquired pneumonia severity on presentation to hospital: an interventional derivation and validation study. *Thorax* 2003; 58(5): 377-82.
43. Wallace A.B. The exposure and treatment of burns. *Lancet* 1951; 1(6653): 501-4.
44. Baux S. Contribution a l'Etude du traitement local des brulures thermigues etendues. 1961; Paris: These.
45. Osler T., Glance L. and Hosmer D.W. Simplified estimates of the probability of death after burns injuries; extending and updating the Baux score. *J. Trauma* 2010; 68: 690-97.
46. Ridley S. Severity of illness scoring systems and performance appraisal. *Anaesthesia* 1998; 53: 1185-1194.
47. Kononenko, I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 2001;23(1):89-109.
48. Magoulas, G and Prentza, A. Machine learning in medical applications. ACA199 workshop, chapter 1.
49. Mena L., Orozco E., Felix V and Ostos R. et al. Machine learning approach to extract diagnostic and prognostic thresholds: Applications in prognosis of cardiovascular mortality. *Computational and Mathematical Methods in Medicine* 2012.
50. Clifton D.A., Gibbons J., Davies J. and Tarassenko L. *Machine Learning and Software Engineering in Health Informatics* 2012. RAISE Workshop. IEEE International Conference on Software Engineering, Zurich. 37-41.
51. Elasy T. and Gaddy G. Measuring subjective outcomes. *J. Gen. Intern. Med.* 1998; 13(11): 757-761.
52. Carmines EG, Zeller RA. Reliability and validity assessment. 1st ed. Beverly Hills / London: SAGE Publications; 1979.
53. Bachman, L. F. 1990.Fundamental Considerations in Language Testing. Oxford: Oxford University Press.
54. Fulcher, G. Assessment in English for academic purposes: putting content validity in its place. *Applied Linguistics* 1999; 20(2): 221-236.

55. Versi E. “Gold standard” is an appropriate term. *British Medical Journal* 1992 Jul 18; 305(6846): 187.
56. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet. 1986*;1(8476):307-10.
57. Davis D., Quinn R.O., Whiteman C.T., Williams J.D. and Young C.R. Concurrent validity of four clinical tests used to measure hamstring flexibility. *J. Strength. Cond. Res*. 2008;22(2):583-8.
58. Mosby’s medical dictionary, 8th edition, 2009. Elsevier.
59. Ely W., Truman, B., Shintani, A. et al. Reliability and Validity of the Richmond Agitation-Sedation Scale (RASS); *JAMA* 2003;289(22):2983-2991.
60. Testa M.A., Simpson D.C., Assessment of quality-of-life outcomes. *N Engl J Med.* 1994;334:835-40.
61. Srisawat N, Kellum J. Acute kidney injury: definition, epidemiology, and outcome. *Current Opinion in Critical Care* 2011; 17: 548-555.
62. Bellomo R, Ronco C, Kellum JA, *et al*. Acute renal failure: definition, outcome measures, animal models, fluid therapy and information technology needs – the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care* 2004; 8: R204-R212.
63. Mehta RL, Kellum JA, Shah SV, *et al*. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care* 2007; 11:R31-R39.
64. Scottish Intensive Care Society Audit Group. Audit of Critical Care in Scotland 2012 reporting on 2011.
65. Henderson LW, Besard A, Michaels A, Bluemle LW. Blood Purification by Ultrafiltration and Fluid Replacement (Diafiltration). *Trans. Amer. Soc. Artif. Int. Organs* 1967; 13(1): 216-222.
66. Kramer P, Wigger W, Rieger J, Matthaei D, Scheler F. Arteriovenous hemofiltration: a new and simple method for treatment of over-hydrated patients resistant to diuretics. *Klin Wochensch* 1977; 55: 1121-1122.
67. Gatward JJ, Gibbon GJ, Wrathall G, *et al*. Renal replacement therapy for acute renal failure: a survey of practice in adult intensive care units in the United Kingdom. *Anaesthesia* 2008; 63(9): 959-66.
68. Bywaters El, Joekes AM. The artificial kidney: Its Clinical Application in the Treatment of Traumatic Anuria. *Proc. R. Soc. Med.* 1948; 41(7): 420-426.
69. Srisawat N, Lawsin L, Uchino S, Bellomo R and Kellum J. Cost of acute renal replacement therapy in the intensive care unit: results from The Beginning and

Ending Supportive Therapy for the Kidney (BEST Kidney) Study. *Crit Care* 2010; 14(2): R46.

70. Davenport A, Will E and Davidson A. Improved cardiovascular stability during continuous modes of renal replacement therapy in critically ill patients with acute hepatic and renal failure. *Crit Care Med* 1993; 21(3): 328-338.
71. Vinsonnea C, Camus C and Combes, *et al*. Continuous venovenous haemodiafiltration versus haemodialysis for acute renal failure in patients with multiple-organ dysfunction syndrome: a muticentre randomised trial. *The Lancet* 2006;368(9533):379-385.
72. Maggiore Q, Pizzarelli F, Dattolo P, Maggiore U and Cerrai T. Cardiovascular stability during haemodialysis, haemofiltration and haemo-diafiltration. *Nephrol. Dial. Transplant* 2000;15(suppl.1):68-73.
73. Lim W, Holinski P, Devereaux PJ, Cook D *et al*. Detecting myocardial infarction in critical illness using screening troponin measurements and ECG recordings. *Critical Care* 2008; 12:R36.
74. Lim W, Qushmaq I, Troponin T Trials Group *et al*. Elevated troponin and myocardial infarction in the intensive care unit: a prospective study. *Critical Care* 2005; 9: R636-644.
75. The joint European Society of Cardiology / American College of Cardiology Committee. Myocardial infarction redefined – A consensus document of the Joint European Society of Cardiology / American College of Cardiology Committee for the Redefinition of Myocardial Infarction. *European Heart Journal* 2000; 21:1502-1513; *Journal of the American College of Cardiology* 2000; 36: 959-969.
76. Thygesen K, Alpert JS and White HD on behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *European Heart Journal* 2007; 28: 2525-2538.
77. Thygesen K., Alpert J.S., Jaffe A.S. et al. Third universal definition of myocardial infarction. *Eur. Heart J.* 2012;33:2551-67.
78. White H.D., Thygesen K., Alpert J.S. and Jaffe A.S. Clinical Implications of the Third Universal Definition of Myocardial Infarction. *Heart* 2014;100:424-432.
79. Krenn L, Delle G. Myocardial infarction in critically ill patients: A diagnostic challenge. *Critical Care Medicine* 2010; 38(12)2304-9.
80. Babuin L, Vasile VC, Rio Perez JA *et al*. Elevated cardiac troponin is an independent risk factor for short- and long-term mortality in medical intensive care unit patients. *Critical Care Medicine* 2008; 36(3): 759-765.

81. Lim W., Holinski P., Devereaux P.J., Thaczyk A., McDonald E., Clarke F., Qushmaq I., Terrenato I., Schunemann H., Crowther M. and Cook D. Detecting myocardial infarction in critical illness using screening troponin measurements and ECG recordings. *Crit Care* 2008;12(2):R36.
82. Lim W, Qushmaq I, Troponin T Trials Group et al. Elevated troponin and myocardial infarction in the intensive care unit: a prospective study. Critical Care 2005; 9: R636-644.
83. Noble JS, Reid AM, Jordan LV, Glen AC, Davidson JA. Troponin I and myocardial injury in the ICU. Br J Anaesth 1999; 82(1):41-46.
84. Scottish Intensive Care Society Audit Group. Audit of Critical Care in Scotland 2012 reporting on 2011.
85. National Kidney Foundation KDOQI Guidelines. Clinical Practice Guidelines for Cardiovascular Disease in Dialysis Patients. http://www.kidney.org/professionals/kdoqi/guidelines_commentaries.cfm
86. sign.ac.uk/guidelines/fulltext/77/index.html.
87. http://www.healthcare.philips.com/pwc_hc/main/shared/assets/documents/patient*monitoring/icip/icip*concept_wp_10.pdf.
88. Moss L, Sleeman D, Kinsella J, Sim M.A.B. ACHE: an Architecture for Clinical Hypothesis Examination. *Proceedings of CBMS* 2008: 21st IEEE International Symposium on Computer-Based Medical Systems.
89. De Vriese AS, Colardyn FA, Philippe JJ, Vanholder RC, De Sutter JH and Lameire NH. Cytokine removal during continuous haemofiltration in septic patients. *J. Am. Soc. Nephrol.* 1999;10(4):846-53.
90. Kooman J, Basci A, Pizzarelli F *et al*. EBPG guideline on haemodynamic instability. *Nephrol. Dial. Transplant* 2007;22(suppl.2):22-44.
91. Canaud B and Lertdumrongluk P. Ultrapure Dialysis Fluid: A new standard for contemporary haemodialysis. *Nephro-urology Monthly* 2012;4(3):519-523.
92. Ridley S. Outcomes in Critical Care 2002. Butterworth Heinemann.
93. Spodick DH, Raju P, Bishop RL and Rifkin RD. Operational definition of normal sinus heart rate. *American Journal of Cardiology* 1992;69:1245-1246.
94. Humes DH and Kelley 1. *Kelley's Essentials of Internal Medicine* 2000, 4th edition. Lippincott Williams and Wilkins. Chapter 76:501.
95. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, KnausWA, Schein RM, Sibbald WJ. Definitions for sepsis and organ failure. *Chest* 1992; 101(6): 1644-55.

96. Davis M and Gore RW. Determinants of cardiac function: simulation of a dynamic cardiac pump for physiology instruction. *Advances in Physiology Education* 2001;25(1):13-35.

97. Lostis JB, Moreyra AE, Amendo MT, Pietro JD, Cosgrove N and Kuo PT. The effect of age on heart rate in subjects free of heart disease. Studies by ambulatory electrocardiography and maximal exercise stress test. *Circulation* 1982;65:141-145.

98. Webb AR, Shapiro MJ, Singer M, Suter PM. *Oxford Textbook of Critical Care* 1999, 1st Edition 1999. Oxford University Press. Chapter 16:1121.

99. Wolfson AB, Hendey GW, Hendry PL, Linden CH, Rosen CL, Schaider J, Sharieff GQ, Suchard JR. *Harwood-Nuss' Clinical Practice of Emergency Medicine* 2005, 4th Edition. Lippincott Williams and Wilkins Chapter 32:198.

100. Pinnock C, Lin T, Smith T. *Fundamentals of Anaesthesia* 2003, 2nd Edition. Greenwich Medical Media. Section 2.

101. Wachter RM, Goldman L, Hollander H. *Hospital Medicine* 2005, 2nd Edition. Lippincott Williams and Wilkins. Chapter 99:998.

102. Marino PL. *The ICU Book* 2007, 3rd Edition 2007. Chapter 38:697-712.

103. Goroll AH, Mulley AG, Albert G. *Primary Care Medicine* 2007, 5th Edition 2007. Chapter 11.

104. McArdle WD, Katch FI and Katch V. *Essentials of Exercise Physiology* 3rd edition 2006. Lippincott Williams and Wilkins. Chapter 10.

105. Parry-Jones AJD, Pittman JAL. Arterial pressure and stroke volume variability as measurements for cardiovascular optimisation. *International Journal of Intensive Care* 2003; 10(2):67-72.

106. Peters JI, Melo J. Low systemic vascular resistance: differential diagnosis and outcome. *Critical Care* 1999; 3(3): 71-77.

107. Peck TE, Hill SA, Williams M. *Pharmacology for Anaesthesia and Intensive Care* 2003, 2nd Edition. Cambridge University Press.

108. Mehta D. *British National Formulary*, 41st Edition, 2001.

109. Yate PM, Thomas D, Short SM, Sebel PS, Morton J. Comparison of infusions of alfentanil or pethidine for sedation of ventilated patients in the ITU. *British Journal of Anaesthesia* 1986; 58(10): 1091-1099.

110. Osborne A. Applied Imagination - Principles and Procedures of Creative Writing. ISBN 1447417100.

111. Dalkey, N. C., and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*;9(3):458-467.

112. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, Teres D, ICU Scoring Group: The logistic organ dysfunction system: A new way to assess organ dysfunction in the intensive care unit. *JAMA* 1996, 276:802-810.
113. Linton RAF, Band DM and Haire KM. A new method of measuring cardiac output in man using lithium dilution. *British Journal of Anaesthesia* 1993; 71: 262-266.
114. deGroot, A. Thought and choice in chess. The Hague, The Netherlands: Mouton. 1946.
115. Ericsson. K. A. and Lehmann A. C. Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology* 1996, 47;273-305.
116. Camerer, C. E and Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson & J. Smith (Eds.), Towards a general theory of expertise (pp. 195-217). Cambridge: Cambridge University Press.
117. Lewandowsky S. and Kirsner K. Knowledge partitioning: Context-dependent use of expertise. *Memory and Cognition* 2000; 28(2): 295-305.
118. Ericsson K.A. and Simon HA. Protocol analysis; verbal reports as data. 1993. Cambridge MA: Bradford books.
119. Johnson P.E. What kind of expert should a system be? *Journal of Medicine and Philosophy* 1983;8:77-97.
120. Klein G. and Militello L. Some guidelines for conducting a cognitive task analysis. *Advances in Human Performance and Cognitive Engineering Research* 2001;1:161-199.
121. Hoffman R., Shadbolt N., Burton A. and Klein G. Eliciting knowledge from experts: A methodological analysis. *Organizational Behaviour and Human Decision Processes* 1995;62(2):129-158.
122. Klein G.A., Calderwood R. and MacGregor D. Critical decision method for eliciting knowledge. IEE*E Trans Syst Man Cybern* 1989;19:462-472.
123. Brule J.F. and Blount A. *Knowledge Acquisition* 1989.McGraw-Hill (New York).
124. Neale I.M. First generation expert systems: A review of knowledge acquisition methodologies. *Knowledge Engineering Review* 1988;3:105-146.
125. Cooke N. from Knowledge Elicitation.
126. Sleeman D., Aiken A., Moss L., Kinsella J. and Sim M.A.B. A system to detect inconsistencies between a domain expert's different perspectives on

(classification) tasks. *Recent Advances in Machine Learning* (Studies in Computational Intelligence) Volume 283/2010: 293-314 (Springer).

127. www.random.org
128. Nikolaos M, Leonidas G and Myrianthefs P. Increased blood troponin levels in ICU patients. *Current Opinion in Critical Care* 2011; 17:454-463.
129. Witten IH and Eibe F. Credibility: Evaluating what's been learned. In *Data Mining.* 2nd edition. Edited by Gray G. Morgan Kaufman; 2005: Chapter 5.
130. Sleeman D, Moss L, Sim MAB and Kinsella J. Predicting Adverse Events: Detecting Myocardial Damage in Intensive Care Unit (ICU) Patients. In *Proceedings of KCAP Conference*: 25-29 June 2011; Banff. ACM; 2012:73-79.
131. Moss L., Sleeman D., Sim M.A.B. and Kinsella J. Using Cardiovascular Derangements to Predict Raised Troponin Levels. 1st International Workshop on Capturing and Refining Knowledge in the Medical Domain (KMED 2012) – In publication.
132. Parshall M.B. Unpacking the 2x2 table. *Heart and Lung: The Journal of Acute and Critical Care* 2013;42(3):221-226.
133. www.medcalc.org
134. Haraldsson H, Edenbrandt L and Ohlsson M. Detecting acute myocardial infarction in the 12-lead ECG using Hermite expansions and neural networks. *Artificial Intelligence in Medicine* 2004; 32: 127-136.
135. Maeder M, Fehr T, Rickli H and Ammann P. Sepsis associate myocardial dysfunction. Diagnostic and prognostic impact of cardiac troponins and natriuretic peptides. *Chest* 2006; 129(5):1349-1366.
136. Constantino J and Murillo S. Myocardial dysfunction in sepsis: a large, unsolved puzzle. Critical Care Research and Practice 2012.
137. Webb I and Coutts J. Myocardial infarction on the ICU: can we do better? *Critical Care* 2008; 12:129.
138. Reynolds T, Cecconi M, Collinson P, Rhodes A, Grounds RM and Hamilton MA. Raised serum cardiac troponin I concentrations predict hospital mortality in intensive care unit patients. British Journal of Anaesthesia 2012; 109(2): 219-24.
139. Velasquez A, Ghassemi M, Osorio JS, Park S, Dejam A and Celi LA. Long-term outcome s of minor troponin elevations in the intensive care unit. *Critical Care Medicine* 2012; 40(12):846.
140. Sleeman D, Muniesa M, Sim MAB, Docking R and Kinsella J. Correlation between mean score of cardiovascular instability and patient outcome. (Abstract presented at the Scottish Intensive Care Society Annual scientific Meeting 2012).

141. Goldhill DR and Sumner A. APACHE II, data accuracy and outcome prediction. *Anaesthesia* 1998; 53(10): 937-943.
142. Féry-Lemonnier E., Landais P., Loirat P, Kleinknecht D. and Brivet F. Evaluation of severity scoring systems in ICUs – translation, conversion and definition ambiguities as a source of inter-observer variability in APACHE II, SAPS and OSF. *Intensive Care Medicine* 1995; 21(4):356-60.

Publications

The research presented in this thesis led to the following publications:

Moss L., Sleeman D., Sim M.A.B. and Kinsella J. Using Cardiovascular Derangements to Predict Raised Troponin Levels. 1st International Workshop on Capturing and Refining Knowledge in the Medical Domain (KMED 2012) – In publication.

Docking R., Moss L., Sim M.A.B., Sleeman D. and Kinsella, J. Investigation into Haemodynamic Stability during Intermittent Haemodialysis in the Critically Ill. *Critical Care* 2012; 16 (Suppl 1):371.

Docking R., Moss L., Sim M.A.B., Sleeman D. and Kinsella, J. Investigation into the Effects of Commencing Haemodialysis in the Critically Ill. *Critical Care* 2012; 16 (Suppl 1):359.

Sleeman D., Moss L., Aitken A., Hughes M., Sim M.A.B. and Kinsella J. Detecting and resolving inconsistencies between domain experts' different perspectives on (classification) tasks. *Artificial Intelligence in Medicine* 2012; 55(2): 71-86.

Moss L., Sleeman D., Quasim T., Sim M.A.B., Booth M., Puxty A. and Kinsella, J. Identifying Myocardial Damage from Routinely Recorded Data in the Intensive Care Unit (ICU). *Intensive Care Medicine* 2011; 37 (Supp 1):213.

Sim M.A.B., Moss L., Docking R., Henderson K., McCallum J., Sleeman D. and Kinsella J. A novel system for detecting myocardial damage in the critically ill patient. *Critical Care Medicine* 2011. 29(12).

Grando, A., Moss, L., Glasspool, D., Sleeman, D., Sim M.A.B. and Kinsella, J. Argumentation-Logic for Explaining Anomalous Patient Responses to Treatments. *Artificial Intelligence in Medicine, 13th International Conference on Artificial Intelligence in Medicine (AIME)* 2011 (Bled, Slovenia). Springer: 35-44.

Sleeman D., Moss L., Sim M.A.B. and Kinsella, J. Predicting Adverse Events: Detecting Myocardial Damage in Intensive Care Unit (ICU) Patients. *The Sixth International Conference on Knowledge Capture (KCAP)* 2011 (Banff, Alberta, Canada). ACM Press, New York: 73-79.

Moss L., Sleeman D., Sim M.A.B. Reasoning by Analogy in the Generation Domain Acceptable Ontology Refinements. *Proc. EKAW* 2010: 534-543.

Moss L., Sleeman D., Sim M.A.B. Booth M., Daniel M. and Donaldson. L. Ontology-Driven Hypothesis Generation to Explain Anomalous Patient Responses to Treatment. *Knowledge-Based Systems* 2010; 23(4): 309-315.

Sleeman D., Aiken A., Moss L., Kinsella J. and Sim M.A.B. A system to detect inconsistencies between a domain expert's different perspectives on (classification) tasks. *Recent Advances in Machine Learning* (Studies in Computational Intelligence) Volume 283/2010: 293-314 (Springer).

Corsar D., Moss L., Sleeman D. and Sim M.A.B. Supporting the Development of Medical Ontologies. *Proc. FOMI* 2009: 114-125.

Moss L., Sleeman D., Sim M.A.B., Booth M., Daniel M., Donaldson L., Gilhooly C., Hughes M. and Kinsella J. Ontology-Driven Hypothesis Generation to Explain Anomalous Patient Responses to Treatment. *Proc. A.I.* 2009 (The Twenty-ninth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK). Springer: 63-76.

Sim M.A.B., Aiken A., Moss L., Sleeman D. and Kinsella J. Confusion matrices to refine a novel scoring system for cardiovascular instability in intensive care. *Scottish Medical Journal* 2009; 54(2): 56.

Moss L., Sleeman D., Booth M., Daniel M., Donaldson L. Gilhooly C., Hughes M., Sim M.A.B. and Kinsella J. Explaining Anomalous Responses to Treatment in the Intensive Care Unit. *Proc AIME* 2009; 250-254.

Sleeman D., Moss L., Sim M.A.B., and Kinsella J. ACHE: Architecture for Clinical Hypothesis Examination. *Transactions 21st IEEE Symposium on Computer-Based Systems* (CBMS 2008); 158-160.

Conference Posters

The research presented in this thesis led to the following posters at meetings:

Docking R., Moss L., Sim M.A.B., Sleeman D. and Kinsella, J. An investigation into haemodynamic stability during intermittent haemodialysis in the critically ill. 32nd International Symposium on Intensive Care and Emergency Medicine, Brussels, Belgium, March 2012.

Docking R., Moss L., Sim M.A.B., Sleeman D. and Kinsella, J. An investigation into the effects of commencing haemodialysis in the critically ill. 32nd International Symposium on Intensive Care and Emergency Medicine, Brussels, Belgium, March 2012.

Sim M.A.B., Moss L., Sleeman D. and Kinsella J. A novel system for detecting myocardial damage in the critically ill patient. Society of Critical Care Medicine (SCCM) Annual Congress, Houston, Texas, USA, February 2012.

Sleeman D., Muniesa M., Moss L., Sim M.A.B., Docking R. and Kinsella, J. Correlation between mean score of cardiovascular instability and patient outcome. Scottish Intensive Care Society Annual Meeting, St.Andrew's, UK, January 2012.

Moss L., Sleeman D., Quasim T., Sim M.A.B., Booth M., Puxty A. and Kinsella, J. Identifying Myocardial Damage from Routinely Recorded Data in the Intensive Care Unit (ICU). European Society of Intensive Care Medicine (ESICM) LIVES Annual Congress, Berlin, Germany, October 2011.

Moss L., Grando M.A., Sleeman D., Sim M.A.B., Gilhooly G. and Kinsella J. Formalising and understanding collaborative decision making in the Intensive Care Unit (ICU). Scottish Intensive Care Society Annual Scientific Meeting, St. Andrews, 2011.

Moss L., Sleeman D., Sim M.A.B., Booth M., Donaldson L., Gilhooly C., Hughes M. and Kinsella J. Development of EIRA, a knowledge-based system to explain anomalous patient responses to treatment. BCS Health Scotland Conference, Glasgow, UK, September 2010.

Sim M.A.B., Booth M.G. Sleeman D., O'Reilly D and Kinsella J. Identification of Troponin positive events in Intensive Care. Scottish Intensive Care Society Annual Scientific Meeting, St Andrews, 2010.

Sim, M.A.B., Moss L., Aiken A., Kinsella J. and Sleeman D. Intermittent haemodialysis may be associated with increased haemodynamic stability. Scottish Intensive Care Society Annual Scientific Meeting, St. Andrews, 2010.

Moss L., Sleeman D., Sim M.A.B., Booth M., Daniel M., Donaldson L., Gilhooly C., Hughes M. and Kinsella J. Transforming Clinical Anomalies into Clinical Insights: Developing a Knowledge-based system which explains a patient's unexpected reaction to treatment. Scottish Intensive Care Society Annual Scientific Meeting, St. Andrews, 2010.

Sim M.A.B., Aiken A., Sleeman D., Moss L. and Kinsella J. Confusion Matrices to Refine a Novel Scoring System for Cardiovascular Instability in Intensive Care. Scottish Intensive Care Society Annual Scientific Meeting, Cumbernauld, January 2009 and Winner of Poster Prize.

Sim M.A.B., Moss L., Sleeman D. and Kinsella J. Assessing cardiovascular status in the ICU. Advances in Anaesthesia and Critical Care Symposium, Glasgow, September 2007.

Presentations

The research described in this thesis led to the following presentations other than the poster presentations listed previously:

"ICU update." Glasgow Western Infirmary anaesthetic departmental meeting, June 2011.

"Update of Studies taking place at the Western Infirmary." Glasgow Western Infirmary anaesthetic departmental meeting, January 2011

"Myocardial Infarction in ICU." Glasgow Royal Infirmary Continuing Professional and Personal Development Meeting May 2009

"Myocardial Infarction in ICU." Royal College of Anaesthetist's Anaesthetics Emergencies Meeting, Glasgow 2009.