ML-EWS: Machine Learning Early Warning System

The application of machine learning to predict in-hospital patient deterioration

Dr Vishal Nangalia

UCL

This dissertation is submitted for the degree of

Doctor of Philosophy

I, Vishal Nangalia confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank Dr David Barber for his guidance on machine learning and its application, and from which I have learnt a great deal. Dr Reecha Sofat for her unequivocal support in the final lap. I must also thank the Medical Research Council and NHS England for funding my research. Thanks also go to the Division of Medicine for supporting my project and helping smooth over the many bumps along the way, especially to Prof Raymond Macallister, Dr Dave Spratt and Prof Robin McAnulty.

I would like thank numerous people who helped me during this research. Dr David Prytherch, Prof Gary Smith and Dr Tony Wolf, who inspired me to embark on this research. Trevor Peacock for his guidance on all things related to data protection and information governance. Tabitha Kavoi, for her patience in helping me navigate the ethics process and her ongoing support for the institutional management of my research. John Adams, for setting up and sorting out my unique IT requirements. Dr Anne Dawnay, for her expertise on everything related to blood results and acute kidney injury.

My local principal investigators, for believing in this research and enrolling their institutions into the study. The research and information teams at all the hospitals that took part in the study, without whom I would not have had the data for this project. Simon Brown, Ani Dwarakanath and Prashant Lele, for believing in my vision of Patient Rescue, helping to mould it, and working well beyond the call of duty to help bring it to fruition. Looking forward to continuing this journey together.

I would also like to thank the NHS England Clinical Entrepreneurship Programme for the ongoing opportunity to help translate this research into clinical practice. It is truly an honour to be part of this inaugural cohort of Clinician Entrepreneurs.

I could not have accomplished this project without the unwavering support of my friends. Thanks to my 'Exponential Medicine' friends, for their inspiration, optimism, energy and belief that we can and will positively transform healthcare, in particular Jack, Laurence, and Milind for being the crazy human beings that you are. Ambrita, David, Kirsten, Maxine, Navin, Lee, Li, James, Victoria, Yulia for being awesome, fun and reminding me to focus on life and not just work. Chris for all those out of the box discussions. Damien, Paul, Julien and especially Peter, for their legal prowess. Roman, for all the squash sessions and Vidhya, for all those concerts.

Finally, this work would not have been possible without the unremitting support of my family. Mum, Dad, Jyoti, thank you for your love and for putting things in perspective.

Abstract

Preventing hospitalised patients from suffering adverse event (AEs) (unexpected cardiac, arrest, intensive care unit admission, surgery or death) is a priority in healthcare. Almost 50% of these AEs, caused by mistakes/poor standards of care, are thought to be preventable. The identification and referral of a patient at risk of an AE to a dedicated rapid response team is a key mechanism for their reduction.

Focussing on variables that are routinely collected and electronically stored (blood test data, and administrative data: demographics, date and method of admission, and co-morbidities), along with their trends, I have collected data on ~8 million admissions. I have explained how to navigate the complex ethical and legal landscape of performing such an ambitious data linkage and collection project.

Analysing data on ~2 million hospital admissions with an in-hospital blood test result, I have

1. described how these variables (particularly urea and creatinine blood tests, method of admission, and date of admission) influence in-hospital mortality rate in different groups of patient.
2. created four machine learning (ML) models that have the highest accuracy yet described for identifying a patient at risk of an SAE, while at the same time capturing the majority of patients likely to die (high sensitivity). These models ML-Dehydration, ML-AKI, ML-Admission, and ML-Two-Tests, can be applied to admissions with limited data, specific syndromes, or on all patients in hospital at different time points in their hospital trajectory respectively. Their area under the receiver operator curves are 79.6%, 85.9%, 93% and 90.6% respectively.
3. built and deployed a technology platform *Patient Rescue* that allows for the automated application of any model in any hospital, as well as the communication of rich patient level reports to clinicians, all in real-time.

The ML models and the *Patient Rescue* platform together form the ML – Early Warning System.

Table of Contents

List of Figures

Figure 6.2 Example Patient Rescue: AKI report 151

List of Tables

Abbreviations

AE	Adverse Events
AKI	Acute Kidney Injury
ALB	Albumin
AUROC	Area Under the Receiver Operator Curve
BCr	Baseline Creatinine
BHOM	Biochemistry and Haematology Outcome Model
Bili	Bilirubin
BUN	Blood Urea Nitrogen
CCMDS	Critical Care Minimum Data Set
CDL	Clinical Decision Limit
CDS	Commissioning Data Set
CDS.APC	Commissioning Dataset for Admitted Patient Care
CKD	Chronic Kidney Disease
COPD	Chronic Obstructive Pulmonary Disease
CQC	Care Quality Commission
CQUIN	Commissioning for Quality and Innovation
Cr	Creatinine
CRB	Criminal Records Bureau
CRN	Clinical Research Network
DBS	Disclosure and Barring Service
Drrt	Died or Renal Replacement Therapy
ECG	Electrocardiograph
eGFR	estimated glomerular filtration rate
EWS	Early Warning System
EWSC	Early Warning Score
GF	Gradient Boosting
GPs	General Practitioners
Hb	Haemoglobin
Hct	Haematocrit
HES	Hospital Episode Statistics
HRA	Health Research Authority
HRG	Healthcare Resource Group
HSCIC	Health and Social Care Information Centre

ICD10	International Classification of Diseases Version 10
ICO	Information Commissioners Office
ICU	Intensive Care Unit
IQR	Inter Quartile Range
IT	Information Technology
K	Potassium
KDIGO	Kidney Disease Improving Global Outcomes
LIMS	Laboratory Information Management System
LOINC	Logical Observation Identifiers Names and Codes
LPI	Local Patient Identifier
MARS	Medical Admissions Risk System
Max	Maximum
MCV	Mean Cell Volume
MET	Medical Emergency Team
Min	Minimum
ML	Machine Learning
ML-EWS	Machine Learning Early Warning System
MRN	Medical Record Number
Na	Sodium
NCEPOD	National Confidential Enquiry into Patient Outcome and Death
NEWS	National Early Warning Score
NHS	National Health Service
NHSE	National Health Service of England
NHSE-AKI	National Health Service of England Acute Kidney Injury
NICE	National Institute of Health and Care Excellence
NLMC	National Laboratory Medicines Catalogue
NPEx	National Pathology Exchange
NWIS	National Wales Informatics Service
OPCS	Office of Population Censuses and Surveys
OPCS-4	OPCS Classification of Interventions and Procedures, version 4
PDF	Portable Document Format
Plts	Platelets
P-POSSUM	Portsmouth Physiology and Operative Severity Score for enUmeration of Mortality

PPV	Positive Predictive Value
PSA	Patient Safety Alert
QAPI	Quality Assessment and Performance Improvement
RCP	Royal College of Physicians
REC	Research Ethics Committee
RF	Random Forest
RI	Reference Interval
RRS	Rapid Response System
RRT	Rapid Response Team
RV	Reference Value
SHA	Secure Hash Algorithm
SQL	Structured Query Language
SSRS	SQL Server Reporting Services
UCL	University College London
UK	United Kingdom
ULRI	Upper Limit of Reference Interval
Ur	Urea
Ur:Cr	Urea to Creatinine Ratio
USA	United States of America
USB	Universal Serial Bus
ViEWS	Vitalpac Early Warning Score
WCC	White Cell Count

Chapter 1: Introduction

1.1 Introduction

My responsibilities as an anaesthesiologist apart from intraoperative anaesthesia include the management of patients who have had a cardiac arrest (as a member of the cardiac arrest team), the optimisation of patients who need emergency surgery, and assisting in the management of patients who have severely deteriorated and are peri-arrest (as a member of the outreach team). When I am referred these patients I almost always notice, a) their physiological deterioration had been going on for some time prior to them being referred to me, and b) important interventions, whether they were additional tests or treatments, were either not done or were delayed. I believe that early identification, enabling immediate appropriate intervention, would prevent such patient deterioration. This early identification needs to occur in the resource constrained environment of the healthcare system (e.g. it must not require expensive additional tests), and must fit in with existing modes and processes of clinical care (i.e. there must be simple mechanisms for its implementation). The challenges are therefore to demonstrate a) that early identification is possible using existing data, and 2) a system to enable this early identification can be implemented that fits into both existing clinical workflows, and integrates with existing healthcare information technology (IT). This belief and drive to improve patient care are the motivations behind my doctoral research. My hypothesis is- "Better understanding of already captured clinical data, using machine learning, and the application of advanced information technology can enable early, automatic and real time identification of all hospitalised patients at risk of future deterioration".

Adverse events (AE) are defined as harm to a patient as a result of medical care or those that occur within a health care setting[1]. These include unexpected prolongation of hospitalisation, cardiac arrest, emergency surgery, emergency intensive care admission, and patient death. They may also include any untoward medical occurrence that result in persistent or significant disability, congenital abnormality, or birth defect. The primary focus of this research will be on the AE of in-hospital patient death.

An AE indicates that the care resulted in an undesirable clinical outcome and may have involved errors, negligence, or poor-quality care. Although this does not necessarily mean that the events were preventable[2], a significant number appear to be of this type. For example, a 2010 United States of America (USA) Government Investigation found that '44% of AEs were clearly or likely preventable'[3]. Similarly, a 2011 United Kingdom (UK) investigation reported less than half (48%) of high-risk surgical

patients received good care[4]. Disease-specific reports support these findings: only 50% of patients who died of acute kidney injury (AKI) received 'good' care[5]. 20–30% of the total number of AKI cases (fatal and non-fatal) were regarded as being preventable. Intervention-specific reports, in this case regarding the administration of intravenous fluids, found that a significant number of hospitalised patients were dying as a result of the infusion of too much or too little fluid[6].

The potential for improving the quality of in-hospital care, and for the reduction of adverse outcomes, is clear. These goals have been incorporated in national clinical guidelines in the UK[6,7], highlighted in government and patient reports[4,5,8,9,10,11,12], and linked to reimbursement via quality improvement metrics[13,14]. The UK, National Institute for Health and Care Excellence (NICE), National Confidential Enquiry into Patient Outcome and Death (NCEPOD) and the Royal College of Physicians (RCP) all advocate programmes to reduce AE frequency through the early recognition of patient decline. In the USA, as a condition of participation in the Medicare and Medicaid programmes, Federal regulations require that hospitals develop and maintain Quality Assessment and Performance Improvement (QAPI) Programmes[13]. As a part of their QAPI programmes, hospitals must 'track medical errors and adverse patient events, analyse their causes, and implement preventative actions'. In the UK, commissioning for quality and innovation (CQUIN) payments[14] are similar vehicles for prioritising AE reduction in hospitals.

Fulfilling these requirements of reducing AEs is a challenge. The substantial and continuously rising cost of healthcare dictates, in my opinion, that it is economically unfeasible to address the problems of AEs by simply increasing funding for the healthcare services. Instead, improved quality of care must accompany increased healthcare efficiency.

The key components of a programme of cardiac arrest reduction, as an example of an AE, have been described as a 'Chain of Prevention'[15], requiring the successful implementation of all the following:

- **Education:** Recognition of the signs of deterioration; appreciating clinical urgency; use of an early-warning score; when and how to use simple interventions; successful teamwork and organization; and end-of-life care.
- **Monitoring:** Includes patient assessment and measurement, and recording of vital signs.
- **Recognition** of all patients likely to deteriorate in the near future, or currently declining. This is difficult to achieve, and failure to do so is one of the most common features of AEs.
- **Call for help (communication):** Communication of this deterioration to the appropriate clinical team (e.g. the rapid response team (RRT)).
- **Response (intervention)** that successfully stabilises the patient or prevents the deterioration from occurring. Many countries have introduced RRTs to which deteriorating patients can be referred.

For my doctoral research, I have focussed on 'Recognition' of the deteriorating patient, using a different modality of monitoring the patient. Existing programmes of 'Monitoring' of patients rely on the measurement of vital signs[7]. Utilisation of laboratory blood test results, with patient administrative data, could be a novel alternative. It is possible that an early warning score (EWSC) based on multiple variables (including but not limited to patient demographics, co-morbidities, and admission dates) may have a high accuracy in identifying those who are at an increased risk to suffer AEs. Because these data are quality-controlled and stored electronically, theoretically this should make them easy to access, and amenable to complex computational analysis. In this thesis I have first applied advanced analytical techniques, specifically machine learning, to laboratory and administrative data, to build models that predict which hospitalised patients are likely to suffer an AE (with a particular focus on mortality). Second, I have built an advanced computer system ('Patient Rescue') that enables the application of such models to any healthcare provider in the world.

1.2 The Problem

A significant proportion of hospitalised patients receive sub-optimal care[7]. This leads to increased mortality and morbidity. Key factors in this sub-optimal care of patients are: *1) the lack of recognition of the seriousness of a patient's condition on first presentation to a hospital; 2) lack of recognition of a patient's subsequent clinical deterioration while in hospital; and 3) despite indications of clinical deterioration being present, it is neither appropriately recognised nor acted upon*[7]. Multiple reports from the UK NCEPOD have focussed on these issues and highlighted that many cases may have been preventable. A summary of each of these influential reports is described below.

- 'An Acute Problem', published in 2005[8], focussed on patients admitted to the Intensive Care Unit (ICU). Among its key findings it found that 'Patients often had prolonged periods of physiological instability prior to admission to ICU. In patients who had been in hospital more than 24 hours prior to ICU admission, 66% exhibited physiological instability for more than 12 hours.' The report also highlighted the inadequacy of physiological monitoring, and the lack of both early warning systems (EWS) and critical care outreach teams. Two of the key recommendations of the report were: 1) increased attention to patients exhibiting physiological abnormalities, as this is a marker of increased mortality; and 2) Deployment of EWSCs to cover all hospitalised patients. These EWSCs should be linked to a RRT that was appropriately skilled to manage the deteriorating patient.
- 'Emergency Admissions: a journey in the right direction?', published in 2007[9], focussed on emergency admissions to hospitals. It highlighted that 34.8% of patients received substandard care, and that the initial assessment and management plan for these patients was inadequate, with a failure to perform appropriate investigations and to recognise critically ill patients. Among

its principal recommendations were training to enable recognition of critically ill patients and a clear physiological monitoring plan for each patient, commensurate with their clinical condition. This physiological monitoring plan should detail what is to be monitored, the desirable parameters and the frequency of observations. It should also include an explicit statement of parameters that would prompt escalation of care, and review by an expert multidisciplinary team; i.e. an EWSC linked to a RRT.

- 'Adding Insult to Injury. A review of the care of patients who died in hospital with a primary diagnosis of acute kidney injury (acute renal failure)', published in 2009[5], focussed on patients who developed AKI. Among its key findings were: a) there was an unacceptable delay in recognising hospital-acquired AKI in 43% of patients; b) 20% of hospital-acquired AKI cases were avoidable; c) complications of AKI were missed in 13% of cases, and avoidable in 17%; and d) 50% of patients received substandard care. Among the report's many recommendations was the robust implementation of 'NICE Clinical Guidance 50', which are systems for recognising acutely ill patients.
- 'Caring to the End? A review of the care of patients who died in hospital within four days of admission', published in 2009[10], highlighted issues of inadequate early assessment, investigation, diagnosis, and management of patients on admission. It also focussed on the failure to urgently escalate care of these patients to a senior clinician or consultant.
- 'Knowing the Risk. A review of the peri-operative care of surgical patients', published in 2011[4] focussed on the care of high-risk surgical patients. Among its key findings were: a) only 48% of high-risk patients received care that was 'good'; b) there was a lack of consensus as to what constitutes high peri-operative risk (a recognition problem); c) 24% of patients were not monitored by an early warning system; and d) there were inadequate pre-operative interventions to optimise a patient's nutritional and fluid status, resulting in poor outcomes for those cases where optimisation did not occur. The key recommendations included better assessment of high-risk patients and optimisation of their nutritional and fluid status, along with the escalation of their care to an enhanced recovery pathway.
- 'Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of an in-hospital cardiorespiratory arrest', published in 2012[11], focussed on patients who had a cardiac arrest in hospital. It found that 64% of cardiac arrests were predictable, with warning signs for imminent cardiac arrest present in 75% of cases. The report also claimed 38% of cardiac arrests were avoidable. Key factors in this poor antecedent patient care were identified as the lack of 'recognition' and failure to escalate the care of these deteriorating patients. Again, the implementation of 'NICE Clinical Guidance 50' was one of the key recommendations.
- Finally, in 'Just Say Sepsis! A review of the process of care received by patients with sepsis', published in 2015[12], the principal recommendations included: a) the implementation of a formal

protocol to enable the early identification and management of patients with sepsis; b) the use of an EWSC, such as the National Early Warning Score (NEWS); c) adequate vital sign monitoring; d) adequate staffing and resources; and e) implementation of 'care bundles'.

Two overriding themes emerge, first poor identification of patients at high risk of deterioration, or those who are already deteriorating; and second inadequate urgency in the escalation of care and review of these patients, by an appropriate senior clinician (consultant or specialist nursing team).

1.3 The Current Solution: Rapid Response Systems

Addressing the problems of poor or delayed identification of high-risk or acutely ill patients is complex, and apart from rapid identification and effective clinical intervention, this involves: education, training and resourcing. Resourcing applies to adequacy in staff and facilities e.g. critical care beds.

The key components of a programme of AE reduction- the 'Chain of Prevention'[15], are described in detail in Section 1.1. To deal specifically with 1) monitoring and identification of the high-risk/deteriorating patient, and 2) the call for help and response components, the most widely implemented solution has been the rapid response system (RRS). The RRS is composed of an afferent limb of monitoring, recognition and alerting (call for help), namely the EWSC, and an efferent limb of intervention (response), which is the RRT. These are described below (in reverse order):

1.3.1 The Efferent Limb: Rapid Response Team

RRTs are referred to by a variety of names, including 'medical emergency team', 'outreach team', 'critical care outreach team', 'emergency response team', or 'patient emergency response team'. The RRT is a designated group of healthcare professionals with a mixture of skills. According to Jones et al, 'These teams are key components of rapid-response systems, which have been put in place because of evidence of "failure to rescue" with available clinical services, leading to AEs'[16]. As a minimum requirement, each RRT includes a senior clinician (acute medicine physician, anaesthetist or intensivist) and a critical care nurse, though the composition varies by institution. The RRT differs from the traditional 'cardiac arrest' team ('code team' in the USA) in a number of ways (Figure 1.1)[16]. The RRT assesses a larger number of hospitalised patients at an earlier stage of their clinical deterioration. The RRT's role is to immediately assess and treat a patient showing objective or subjective signs of clinical deterioration. The RRT's goal is to reverse the patient's deterioration and prevent an AE. The RRT responds to emergencies, proactively evaluates high-risk ward patients, educates and acts in liaison with ward staff. The RRT may also follow up on patients discharged from an ICU. RRTs have

been introduced in many countries, including for example Australia, Canada, Denmark, the Netherlands, New Zealand, Sweden, UK and USA.

Figure 1.1[16]

Comparison between a Traditional Code Team and a Rapid-Response Team.*

Feature	Traditional Code Team	Rapid-Response Team
Typical criteria for calling the team	No recordable pulse, no recordable blood pressure, absence of respiratory effort, unresponsive	Low blood pressure, rapid heart rate, respiratory distress, altered consciousness
Typical conditions that the team assesses and treats	Cardiac arrest, respiratory arrest, airway obstruction	Sepsis, pulmonary edema, arrhythmias, respiratory failure
Typical team composition	Anesthesia fellow, ICU fellow, internal-medicine house staff, ICU nurse	ICU fellow, ICU nurse, respiratory therapist, internal-medicine house staff
Typical call rate (no./1000 admissions)	0.5–5	20–40
Typical in-hospital mortality (%)	70–90	0–20

* ICU denotes intensive care unit.

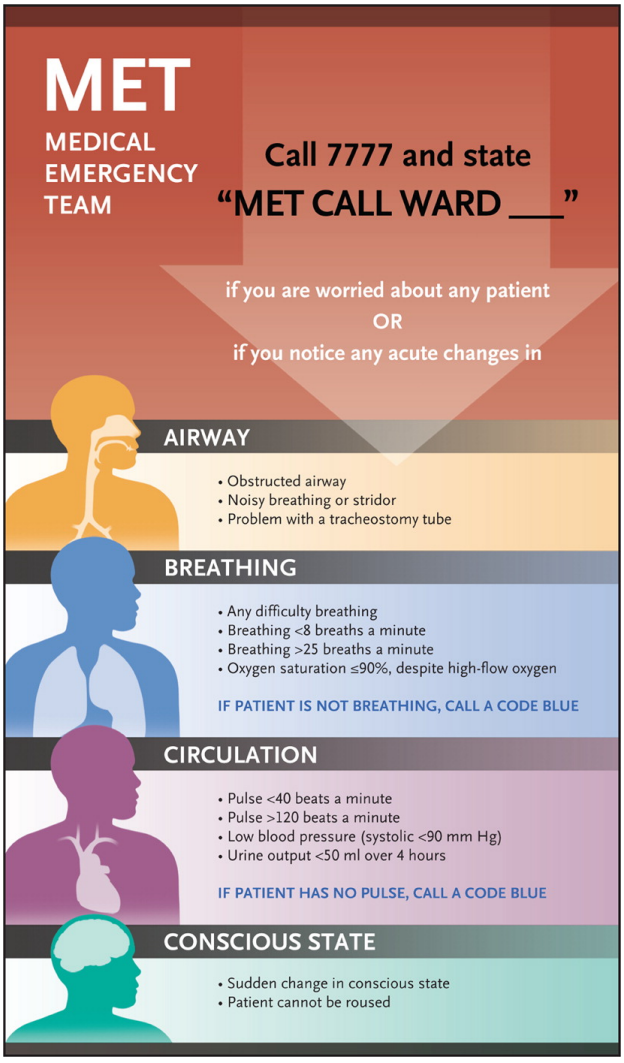
1.3.2 The Afferent Limb: Vital Sign EWSCs

It is well recognised that, prior to an in-hospital patient having a cardiac arrest, being admitted to an ICU, or dying; they exhibit a period of physiological deterioration that is potentially recognisable and reversible[4,8,9,10,11,12]. However, frontline medical staff often ignore this. As a result, the use of an EWSC has been mandated in health care settings, in order to aid the detection and escalation of care of the high-risk/deteriorating patient.

EWSC were developed to facilitate early detection of deterioration by categorising a patient's severity of illness and prompting frontline medical staff to request a medical review at specific trigger points[7]. EWSC are sometimes referred to as 'Track and Trigger Score'.[7] Each of a patient's vital signs (blood pressure, heart rate, respiratory rate, etc.), is allocated a numerical score depending on its absolute value. Individual scores are added to give the total score for that patient, which is their EWSC. A high or rising EWSC value indicates a patient has or is deteriorating, while falling values indicate their improvement. Thresholds are set, where if a patient's score is equal to or above a certain value, specific actions are recommended, from increasing the frequency of recording the vital signs to immediately calling the RRT. It is also not always necessary to aggregate the individual vital signs together, and some countries, particularly the USA, have instituted MET (medical emergency team) calling criteria, which are dependent only on the absolute values of individual vital signs. Examples of MET and EWSC calling criteria are shown in Figures 1.2 and 1.3. At some hospitals, patients themselves and family members may trigger/call the rapid response team.

In the UK, in 2007 the NICE issued guidance that all patients in hospital should be monitored using an EWSC with appropriate escalation.[7] In 2012, The RCP published the 'National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS' report.[17] In this, it recommended that a new EWSC, known as 'NEWS', be implemented across all NHS hospitals. In 2014, the Irish National Clinical Effectiveness Committee issued its National Clinical Guideline No. 1, namely the National Early Warning Score (NEWS), and recommended its introduction in Ireland.[18] Figure 1.4 details the assigned values for each vital sign parameter of NEWS, and Figure 1.5 shows the thresholds and triggers of the aggregated values. NEWS was evaluated against a range of other EWSC, and was shown to have an area under the receiver operator curve (AUROC) of 89% for discriminating for in-hospital mortality within 24 hours of NEWS assessment, for patients admitted to a medical admissions unit (medical emergency admissions). As a result, rapid response systems (RRS) are now standard practice in the UK and in most developed countries.

Figure 1.2: Single-parameter MET calling criteria[16]



A hospital poster listing criteria for activation of a rapid response team. Such posters are displayed on the walls of hospitals to remind caregivers of abnormalities in vital signs that are considered to require intervention. This poster is based on one displayed at Austin Hospital, Heidelberg, Victoria, Australia [16]

Figure 1.3: Multi-parameter EWSC[19]

Modified Early Warning Score

Score	3	2	1	0	1	2	3
Respiratory rate (min^{-1})		≤ 8		9–14	15–20	21–29	> 29
Heart rate (min^{-1})		≤ 40	41–50	51–100	101–110	111–129	> 129
Systolic BP (mmHg)	≤ 70	71–80	81–100	101–199		≥ 200	
Urine output (ml/kg/h)	Nil	< 0.5					
Temperature (°C)		≤ 35	35.1–36	36.1–38	38.1–38.5	≥ 38.6	
Neurological				Alert	Reacting to voice	Reacting to pain	Unresponsive

The scores for each parameter are recorded at the time that observations are taken. If the total is 4 or more then the ward doctor is informed.

Figure 1.4: National Early Warning Score (NEWS)

PHYSIOLOGICAL PARAMETERS	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51 - 90	91 - 110	111 - 130	≥131
Level of Consciousness				A			V, P, or U

*The NEWS initiative flowed from the Royal College of Physicians' NEWSDIG, and was jointly developed and funded in collaboration with the Royal College of Physicians, Royal College of Nursing, National Outreach Forum and NHS Training for Innovation.

Figure 1.5: National Early Warning Score (NEWS) escalation criteria

NEW scores	Clinical risk
0	Low
Aggregate 1–4	
RED score* (Individual parameter scoring 3)	Medium
Aggregate 5–6	
Aggregate 7 or more	High

The NEWS trigger system aligned to the scale of clinical risk

***RED score** refers to an extreme variation in a single physiological parameter (ie a score of 3 on the NEWS chart, coloured RED to aid identification and represents an extreme variation in a single physiological parameter). The consensus of the NEWSDIG was that extreme values in one physiological parameter (eg heart rate ≤40 beats per minute, or a respiratory rate of ≤8 per minute or a temperature of ≤35°C) could not be ignored and on its own required urgent clinical evaluation.

1.3.3 Issues with Vital Sign EWSCs

Before, embarking on creating an EWSC based on blood results, administrative data was undertaken, an understanding of the key problems with vital sign based EWSCs was carried out. I did this to mitigate any of these issues arising in my ML based approach. Although, vital-sign-based EWSCs have been widely implemented. I believe there remain significant issues with their use, which limits their effectiveness in adequately identifying the high-risk/deteriorating patient.

In my opinion, the two most significant problems are

- Limitations on the measurement of vital signs
- Poor positive predictive value in the identification of the high-risk/deteriorating patient

Both of which are explored in below.

1.3.3.1 Limitations on the measurement of vital signs

Measuring vital signs is routine. Vital signs are used to help diagnose and monitor the health status of the patient, and their response to clinical interventions. However, there are numerous studies that have shown that such measurements may not be performed predictably, accurately or completely.[16] In addition, the frequency of measuring vital signs has been shown to be inadequate and a cause for the delayed detection of a deteriorating patient.[8,9,10,12] Thus, if monitoring is not being carried out, or the values recorded from it are inaccurate, then any escalation of care that relies on the interpretation of the vital signs is bound to be detrimentally affected, reducing the effectiveness of a vital-sign-based EWSC.

1.3.3.2 Poor positive predictive value in the identification of the high-risk/deteriorating patient

There is significant variation in the predictive performance of a vital-sign-based EWSC to identify patients at risk of an AE[20, 21].

Single-parameter EWSCs comprise escalation criteria based on the values of single variables only, such as 'heart rate < 40 beats per minute (bpm) or > 130 bpm', 'respiratory rate < 8 breaths per minute (brpm) or > 30 brpm', etc. Smith et al. compared 80 single-parameter EWSCs in 9,987 emergency medical patients, and demonstrated variations in their positive predictive values (PPVs) to identify death within hospital, which ranged from 13.5% to 26.1%.[20] Sensitivities varied from 7.3% to 52.5%, with specificities ranging from 69.1% to 98.1%. The maximum positive predictive value (PPV) of identifying a patient dying in hospital was 26.1%, which equates to one patient death in every 3.8 (1:3.8) predicted patient deaths. However, this particular EWSC had a sensitivity of only 7.3%; i.e. it identified only 7.3% of all patients who died in hospital. A slightly more effective single-parameter EWSC was that of Salamonson[22], which had a marginally lower PPV of 25.7% (1:3.9), but identified only

19% of patients likely to die. The single-parameter EWSC with the highest sensitivity of 52.8% had a PPV of only 13.5% (1:7.4).[23]

Multiple-parameter EWSCs perform substantially better. For the same vital sign database used to evaluate the single-parameter EWSCs above, the area under the receiver operator curve (AUROC) for death within 24 hours of the vital signs being measured varied from 65.7% to 78.2%.[21] This comparative analysis was repeated on a larger dataset (35,585 admissions, 198,755 vital sign observation sets) with the NEWS.[24] NEWS achieved an AUROC for identifying death within 24 hours of 89.4%, with the other EWSCs' AUROCs ranging from 81.3% to 85.8%. No positive predictive values, sensitivities or specificities for any threshold levels (the numerical value of the EWSC) were explicitly stated in the paper. However, for NEWS, it was indicated that when the value was 5, the sensitivity of capturing patients who died within 24 hours was 75%. In a separate study, two EWSCs, ViEWS (Vitalpac Early Warning Score, the basis for NEWS) and MEWS (Modified Early Warning Score), were tested on a dataset of 1.15 million vital signs obtained from 42,230 admissions.[25] The mortality rate in this study was 1.79% (756 deaths). ViEWS achieved an AUROC of 86.2%, while MEWS's AUROC was 86.5%. However, for ViEWS, the positive predictive value in discriminating whether the patient was likely to die, even at a relatively high threshold value of 11 (the recommended threshold for a severe patient is 7), was only 1:4.5 (22.1%), and this was achieved with a poor sensitivity of only 30.7%. Another study, by Romero-Brufau et al.[26], used a different dataset of almost seven million sets of vital sign observations from 46,366 admissions, and compared various EWSCs for their predictive accuracy in correctly classifying a composite outcome. The composite outcome was positive if it occurred within 36 hours, and included cardiac arrest, unplanned transfer to ICU, resuscitation call, or RRT activation. This study relied not just on admission data, but also on data collected throughout a patient's hospital stay. The maximum PPV attained by an EWSC was 21%, but this was paired with a poor sensitivity of only 8%. NEWS seemed to consistently have a PPV of less than 10% (the sensitivity was not reported). The exact threshold value for NEWS was not stated, but it can be estimated to be either 5 or 7, corresponding to the moderate or severe thresholds respectively.

In summary, existing vital sign based EWSCs do not offer good predictive value for RRT activation, and better EWSCs need to be developed and validated.

1.4 An Alternative Afferent Limb: Use of Laboratory Data

An alternative to using an EWSC based on vital signs would be to use existing quality-controlled data stored electronically in hospital. This would mitigate issues affecting data quality and calculations.

Medical professionals already use the results of blood tests to help them diagnose, prognosticate and monitor interventions. Although not regarded as an EWSC, blood test results are already used to identify high-risk/deteriorating patients. Diagnoses are normally made by interpreting individual blood results and their trends, ratios of different blood results, simple combinations of a few blood results, and in the case of blood gas measurements: to aid the diagnosis of metabolic vs respiratory acidosis/alkalosis. Scoring systems also exist which use blood result data to aid diagnosis and severity grading of diseases, for example Ranson's criteria for predicting the severity of acute pancreatitis.[27]

In my opinion, four specific methods by which blood results are interpreted, their use in clinical practice, and their potential for use as an EWSC, are given below (ranked in increasing complexity):

1. **Simple single-parameter: Reference intervals.**
2. **Simple dual-parameter: Dehydration (urea to creatinine ratio).**
3. **Dynamic single-parameter: AKI.**
4. **Aggregated multi-parameter: potential laboratory-based EWSCs.**

In this thesis, I have explored each of the above methods. I have specifically in Chapters 3, 4 and 5, used the methods, of 2, 3 and 4 above.

In each of these chapters and for each use case I have demonstrated

1. the inadequacy of that method in discriminating between patients likely to die or not
2. the substantial improvement in discrimination an ML approach achieves

I have shown this by the application of three techniques

1 adding additional variables

2 incorporating dynamic change

3 applying ML techniques

The background behind each of these existing methods of interpretation follows below.

1.4.1 Simple single parameter: Reference intervals

Most blood results are reported with a reference interval (RI). The purpose of the RI is to help the clinician to interpret the blood result. The display of these RIs is a legal requirement in Europe (EU Directive 98/79/EC) and the rest of the world[28], and each laboratory calculates and presents the RI for each blood test performed[28]. These RIs may vary slightly between laboratories. RIs are usually derived from a 95% interval of a reference distribution of values present in that population[28]. The population used to calculate the reference range of a test should be representative of the healthy population on

whom the test will used. However, frequently the sample population is not representative and is based on young Caucasian males, rather than the diverse ethnicities and ages of the population who present for medical treatment[28]. Disease states themselves may shift the appropriate RIs for patients for example chronic kidney disease patients having known and stable creatinine results of between 100 and 200 micromol/L.

RI's date back to 1968.[29] The use of RI's is supported by the International Federation of Clinical Chemistry, which in 1991 recommended that each laboratory follow defined procedures to produce its own reference values for tests carried out.[28] The latest, most significant step in the development and harmonisation of reference ranges is the 2010 EP28-A3C guideline.[30]

Frequently confused with RIs are clinical decision limits (CDLs). According to Yesim Ozarda in his recent review of RIs[28],

> *'CDLs are thresholds above or below which a specific medical decision is recommended. CDLs are based on the diagnostic question and are obtained from specific clinical studies to define the probability of the presence of a certain disease or a different outcome. These limits lead to the decision that individuals with values above or below the decision limit should be treated differently'.*

Examples of CDLs include cholesterol, troponin, HBA1C, and blood transfusion thresholds[28]. Although RIs and CDLs are distinct, they are commonly reported in the same reference interval/range field when a result is displayed. The consequence of this is that, RI's maybe be inappropriately regarded as a CDL, rather than just a statistical range, especially by junior clinicians. This may cause them to inappropriately believe the patient is free from disease.

High or low values for a range of blood tests, such as for sodium, glucose, white cell count, urea and Ur:Cr ratio, have been shown to correlate with poor outcomes in various hospital populations.[31] Figures 1.6 and 1.7 are two examples of univariate analysis on individual blood results, which highlight increased patient mortality when their blood test result is outside its RI. Using individual blood results, the odds ratio for mortality for results outside the RI range from 0.6 to 13.8, for lymphocytes > 3 and age > 65 respectively (Figure 1.6). Similarly, the AUROC for predicting mortality for individual blood results range from 0.5788 to 0.8069 for haematocrit and pH respectively (Figure 1.7). As for single-parameter vital-sign-based EWSCs, univariate analysis on individual blood tests provides poor discrimination for predicting in-hospital mortality. Thus, although essential for diagnosis and monitoring or patient health, an EWSC based only on individual blood tests results has little utility.

Figure 1.6: Individual blood test results: Odds ratio for in-hospital mortality [31]

Variables	Ranges	Number		Univariate analysis		
		Dead $n = 550$	Alive $n = 1100$	OR	95% CI	P
Age (years)	20–49	27	437		Reference range	
	50–64	35	223	3.9	2.7–5.7	0.0001
	≥65	487	392	13.5	6.4–28	0.0001
Sodium (mmol/L)	>145	37	25	4.3	2.5–7.5	0.0001
	135–145	313	914		Reference range	
	130–134	104	110	2.7	2.0–3.7	0.0001
	125–129	69	47	4.2	2.8–6.4	0.0001
	<125	27	14	5.6	2.8–11.3	0.0001
Potassium (mmol/L)	>5.0	69	34	5.2	3.3–8.2	0.0001
	3.5–5.0	335	860		Reference range	
	<3.5	95	156	1.8	1.3–2.4	0.0001
Chloride (mmol/L)	>109	21	22	3.1	1.6–5.9	0.0003
	99–109	262	864		Reference range	
	<99	267	232	3.7	3.0–4.7	0.0001
Bicarbonate (mmol/L)	>33	37	27	3.5	2.0–5.9	0.0001
	22–33	324	816		Reference range	
	<22	189	257	1.9	1.5–2.3	0.0001
Urea (mmol/L)	>7.0	397	284	7.4	5.8–9.4	0.0001
	2.5–7.0	143	758		Reference range	
	<2.5	10	58	0.9	–	0.9
Creatinine (μmol/L)	>130	235	143	5.0	3.9–6.4	0.0001
	50–130	315	958		Reference range	
	<50	1	4	0.8	–	1†
Glucose (mmol/L)	>11.0	57	67	2.3	1.5–3.4	0.0001
	7.1–11	148	186	2.1	1.6–2.8	0.0001
	5.0–7.0	195	521		Reference range	
	<5.0	38	150	1.5	1.0–2.2	0.06
Leucocytes ($\times 10^9$/L)	>10	335	469	2.2	1.8–2.7	0.0001
	4–9	200	617		Reference range	
	<4	14	12	3.6	1.5–8.5	0.002
Neutrophil ($\times 10^9$/L)	>7	377	485	2.9	2.3–3.7	0.0001
	2–7	153	576		Reference range	
	<2	10	33	1.1	–	0.9
Lymphocyte ($\times 10^9$/L)	>3	33	139	0.6	0.4–0.9	0.04
	1–3	275	745		Reference range	
	<1	232	211	3.0	2.4–3.8	0.0001
Platelet ($\times 10^9$/L)	>450	56	41	3.2	2.1–5.0	0.0001
	150–450	423	998		Reference range	
	<150	68	55	2.9	2.0–4.3	0.0001
Haemoglobin (g/dL)	>17	12	30	1.2	–	0.7
	12–17	283	873		Reference range	
	<12	255	195	4.0	3.2–5.1	0.0001

†Fisher exact test result. CI, confidence interval; OR, odds ratio; P, probability.

Figure 1.7: Individual blood test results: Area under the receiver operator curve for in-hospital mortality [32]

Variable	MET call AUC-ROC [95% CI]	ICU admission AUC-ROC [95% CI]	Death AUC-ROC [95% CI]
Haemoglobin	0.610 [0.580–0.643]	0.5845 [0.5387–0.6299]	0.6330 [0.6133–0.6532]
Haematocrit	0.587 [0.557–0.621]	0.5489 [0.4996–0.5993]	0.5788 [0.5562–0.6004]
Total bicarbonate	0.557 [0.527–0.593]	**0.7820 [0.7439–0.8212]**	**0.7318 [0.7126–0.7515]**
WCC	0.625 [0.595–0.657]	0.6304 [0.5838–0.6750]	0.6913 [0.6711–0.7099]
Albumin	0.660 [0.626–0.693]	**0.7244 [0.6775–0.7683]**	**0.7791 [0.7614–0.7966]**
pH	0.631 [0.595–0.669]	**0.7926 [0.7659–0.8170]**	**0.8069 [0.7913–0.8211]**
Bilirubin	0.580 [0.543–0.618]	0.5661 [0.5161–0.6147]	0.5799 [0.5574–0.6020]
Creatinine	0.655 [0.624–0.685]	**0.7461 [0.7116–0.7840]**	**0.7645 [0.7494–0.7803]**
Urea	0.674 [0.644–0.705]	0.6976 [0.6565–0.7369]	**0.7905 [0.7766–0.8059]**

Hb, haemoglobin concentration (g/L); Hct, haematocrit concentration (L/L); Total bicarbonate, bicarbonate + carbonic acid (mmol/L); ALB, albumin (g/L); Bili, bilirubin (μmol/L); WCC, white cell count ($\times 10^9$/L); cr, creatinine (mmol/L); U, urea (mmol/L). Values >0.70 highlighted in bold. AUC-ROC, area under the receiver operating characteristic curve; ICU, intensive care unit; MET, medical emergency team.

1.4.2 Simple dual-parameter: Dehydration (urea to creatinine ratio):

One of the simplest ways two blood result variables have been used in conjunction has been the urea to creatinine ratio (Ur:Cr) as a non specific surrogate for dehydration.[33]

Dehydration is a reduction in total body water. This can occur due to reduced intake, or increased losses. A reduction in the intravascular volume without a reduction in total body water, which occurs when intravascular water moves into the third spaces (interstitial, peritoneum, etc.), is intravascular hypovolaemia. Most clinical measurements of dehydration rely on sampling of the intravascular plasma and thus cannot normally differentiate between the two.[34] Understanding the specific type of dehydration is critical to initiating the appropriate treatment. This section however, focusses on providing an overview of the deleterious effects on the body of dehydration (total body water and hypovolaemia), when measured primarily using Ur:Cr.

Water accounts for some 60% of adult body mass and is essential for human life: it is the environment in which most biochemical reactions occur, and is necessary for the convective transport of nutrients, oxygen and metabolic waste products. Dehydration disturbs these functions, reducing cardiac output and cell volume, increasing plasma osmolality and blood viscosity, and driving visceral blood flow redistribution (including a reduction in cutaneous blood flow).[35,36]

A laboratory-based marker for classifying whether or not a patient is dehydrated is the serum Ur:Cr. Ur:Cr is an historical means of diagnosing dehydration.[33] Urea is the end product of nitrogen-containing amino-acid metabolism. In the kidneys, urea is freely filtered by the glomerulus, as well as both resorbed and secreted by renal tubules. Creatinine is derived from the metabolism of muscle creatine. In the kidneys, creatinine is freely filtered and secreted by the proximal renal tubular. Ur:Cr changes when there is either a disproportionate fall in creatinine vs urea, or a disproportionate rise of urea vs creatinine. In individuals who are dehydrated, urea concentrations in the renal medulla (and thus plasma) rise[37] whilst creatinine continues to be freely filtered. This results in a rise in the Ur:Cr. A Ur:Cr >80 mmol/L:mmol/L (BUN:CR >20 mg/l:mg/l)) has been traditionally considered a marker of dehydration (or intravascular volume depletion).[38,39] However, Ur:Cr may rise for other reasons: urea for instance, also rises in hypercatabolic states (sepsis, major surgery, starvation)[40]; with the large 'blood protein meal' of an upper gastrointestinal bleed[41,42]; or with high-dose glucocorticoid administration.[43] Alternatively, in the context of a low skeletal muscle mass (e.g. in the elderly, cachectic or chronically malnourished)[40,44,45], where a rise in Ur:Cr may be due to a fall in creatinine.

Hospitalised patients are at risk of dehydration, due to impaired intake (resulting from cerebral, musculoskeletal or gastrointestinal pathology) or excessive losses (e.g. enteric, renal or insensible). Even modest dehydration (<2% loss of body mass) may impair cognitive[46] and physical performance[47], mood[48] which can manifest as delirium in the elderly[49], and impair thermoregulation.[50] It can cause constipation, and is associated with impaired wound healing[51], urolithiasis and urinary tract infection.[36] Dehydration also leads to renal hypoperfusion, and plays an important pathogenic role in the development of AKI.[5,52] Specifically, a raised Ur:Cr is associated with an increased risk of death in patients with AKI[53] and in those on long-term dialysis.[54,55] In stroke, it is associated with early clinical deterioration[56], impaired functional outcome[57,58], thromboembolism[37] and mortality[57,58]. Elevated Ur:Cr is also an independent marker of mortality in critical care[40,59], heart failure[60], myocardial infarction[61] and gastrointestinal cancer[39], and also indicates functional impairment[62] and poor rehabilitation in the elderly.[63]

Independently, both low creatinine and raised urea levels are also associated with poor outcomes. Low Cr marks a poor outcome in critical illness[45], and raised Ur indicates a poor outcome in pneumonia[64,65], acute pancreatitis[66], coronary artery bypass grafting[67], myocardial infarction[61], decompensated[68] and chronic[60] heart failure, and critical illness.[40]

The impact of dehydration on healthcare costs and outcomes is causing increasing concern to England's Care Quality Commission[69], patient associations[70] and the Parliamentary Ombudsman[71], with such concerns being echoed in recent independent inquiries[72] and in the media.[73,74,75,76,77] The NICE has recently issued clinical guidelines in an effort to improve the prevention and management of dehydration and AKI.[6,78] However, despite concerns over the clinical impact of dehydration and the awareness that it may be more commonplace than expected, the true current prevalence of dehydration in hospital patients is not known. In 1991, 6.7% of 731,695 USA Medicare admissions recorded dehydration as a cause for admission.[79] From 1994 to 2000, admission rates for dehydration ranged from 130 to 134 admissions per 100,000 population.[69] In England in 2012–2013, some 11,417 (0.075%) of all hospital admissions (n=15.1 million), and 16,928 (0.095%) of all consultant episodes (n=17.7 million) were coded for primary dehydration (International Classification of Diseases, 10th edition, code: E86.X).[80] This rose to 243,161 (1.7%) of all consultant episodes, if dehydration was coded either as a primary or an associated diagnosis. When Ur:Cr is used to define dehydration, estimates of the prevalence of dehydration rise dramatically: 54% of all elderly orthopaedic rehabilitation patients (n=39) were dehydrated at the point of admission.[63] Likewise, between a quarter and a half of stroke patients were dehydrated on admission[56,81], and 62% suffered this condition at some point during their hospital stay (n=2591).[58]

Thus, two routinely performed blood tests, whether viewed independently or interpreted in conjunction, provide significant insight into the condition of the majority of hospitalised patients. The results of these tests, and in particular their ratio, are strongly associated with in-hospital mortality. It is, however, yet to be established whether an EWS, applicable to all hospitalised patients, that incorporates the Ur:Cr of a patient and tracks its change, could be developed and used to identify high-risk/deteriorating patients. This is because all reported studies to date have been single-site, and either disease- or speciality-specific.

1.4.3 Dynamic single-parameter: AKI

The trends of blood results, i.e. their rise or fall, have always been extremely valuable to clinicians. A recent and highly publicised interpretation of a very specific dynamic change in just one specific blood result (creatinine) is that of AKI.[82]

AKI is an abrupt impairment in kidney function that results in a rise in serum creatinine concentration or a fall in urine output. AKI is a broad clinical syndrome that encompasses a number of aetiologies, including but not limited to kidney diseases (e.g. acute interstitial nephritis; acute glomerular and vasculitic renal diseases; ischaemia, toxic injury), extra-renal pathology (for example, pre-renal azotaemia, acute post-renal obstructive nephropathy, sepsis). More than one of these conditions may coexist in the same patient. In 2012, to harmonise the detection and treatment of AKI, a rise in patient's serum creatinine and a fall in their urine output measures were used by the Kidney Disease: Improving Global Outcomes (KDIGO) group, to standardise definitions of its presence and stage of severity (stages 1–3).[83]

One in every five hospitalised adults (21%) suffer AKI[84], worldwide; with a prevalence of 14% reported amongst UK hospital admissions.[79] AKI is also associated with increased risk of death[85], prolonged hospitalisation[86], requirement for renal replacement therapy[87], or the development of chronic kidney disease.[88] The associated health care costs of AKI exceed £1 billion per year in the UK.[79] Over 50% of AKI-associated morbidity and mortality may be preventable with early detection and appropriate intervention.[5] A more systematic approach that does not rely on any one individual checking or acting on results, rather a process that used existing data to alert health care staff that a patient is at high risk of an poor outcome based on their results is now achievable, given the way we store and report blood test results.

In response, initiatives throughout the world have attempted to promote AKI recognition and encourage timely interventions to hasten its resolution.[82] In 2015, the National Health Service of England (NHSE) mandated the implementation of a standardised AKI detection algorithm (based on

KDIGO, Table 1.1) in hospital laboratory information-management systems (NHSE-AKI algorithm, Figure 1.8). The NHSE-AKI algorithm compares the current creatinine result of a patient with their previous results, to determine whether a significant rise has occurred. Specifically, the current result is compared to a 'baseline creatinine' value, which is calculated as either the patient's minimum creatinine result in the previous seven days, or the median of all their creatinine results in the preceding 8–365 days, whichever is lower.

Table 1.1: Kidney Disease Improving Global Outcomes (KDIGO): Staging of AKI

Stage	Serum Creatinine	Urine output
1	1.5–1.9 times baseline or >0.3 mg/dl (>26.5 mol/l) increase	< 0.5ml/kg/hr for 6-12 hours
2	2.0–2.9 times baseline	< 0.5 ml/kg/hr for > 12 hours
3	3.0 times baseline or Increase in serum creatinine to >4.0 mg/dl (>353.6 mol/l) or Initiation of renal replacement therapy or In patients <18 years, decrease in eGFR to <35 ml/min per 1.73 m2	< 0.3 ml/kg/hr for > 24 hours or Anuria for > 12 hours

The NHSE-AKI algorithm is intended to alert clinicians to potential AKI cases and their likely severity stage, and thus identify those at risk of subsequent AEs, such as death or renal replacement therapy (Drrt).[89] Clinical assessment following such an AKI alert may lead to an escalation of intervention according to the AKI stage (Figure 1.9). Thus, although not explicitly labelled as such, the NHSE-AKI algorithm could be regarded as the afferent limb of a 'rapid response system' (RRS), where the efferent limb 'rapid response team' (RRT) includes a nephrologist.

However, whilst clinicians may consider more severe AKI stages to be more dangerous (and worthy of prioritisation of care), the degree to which this is correct is unclear. Patient heterogeneity or clinical state (e.g. the presence of dehydration or of co-morbidities) might influence the degree of risk within any one AKI stage, or across stages. Likewise, multiple trigger routes can lead to the same AKI stage (Figure 1.8), but might be associated with different outcomes. None of these issues have been sufficiently evaluated on a large multisite dataset, and thus we have yet to establish the benefit of implementing an existing AKI algorithm, which does not account for all of these factors, as the afferent limb of a referral system.

Figure 1.8: The NHS England Acute Kidney Injury Algorithm (NHSE-AKI algorithm)

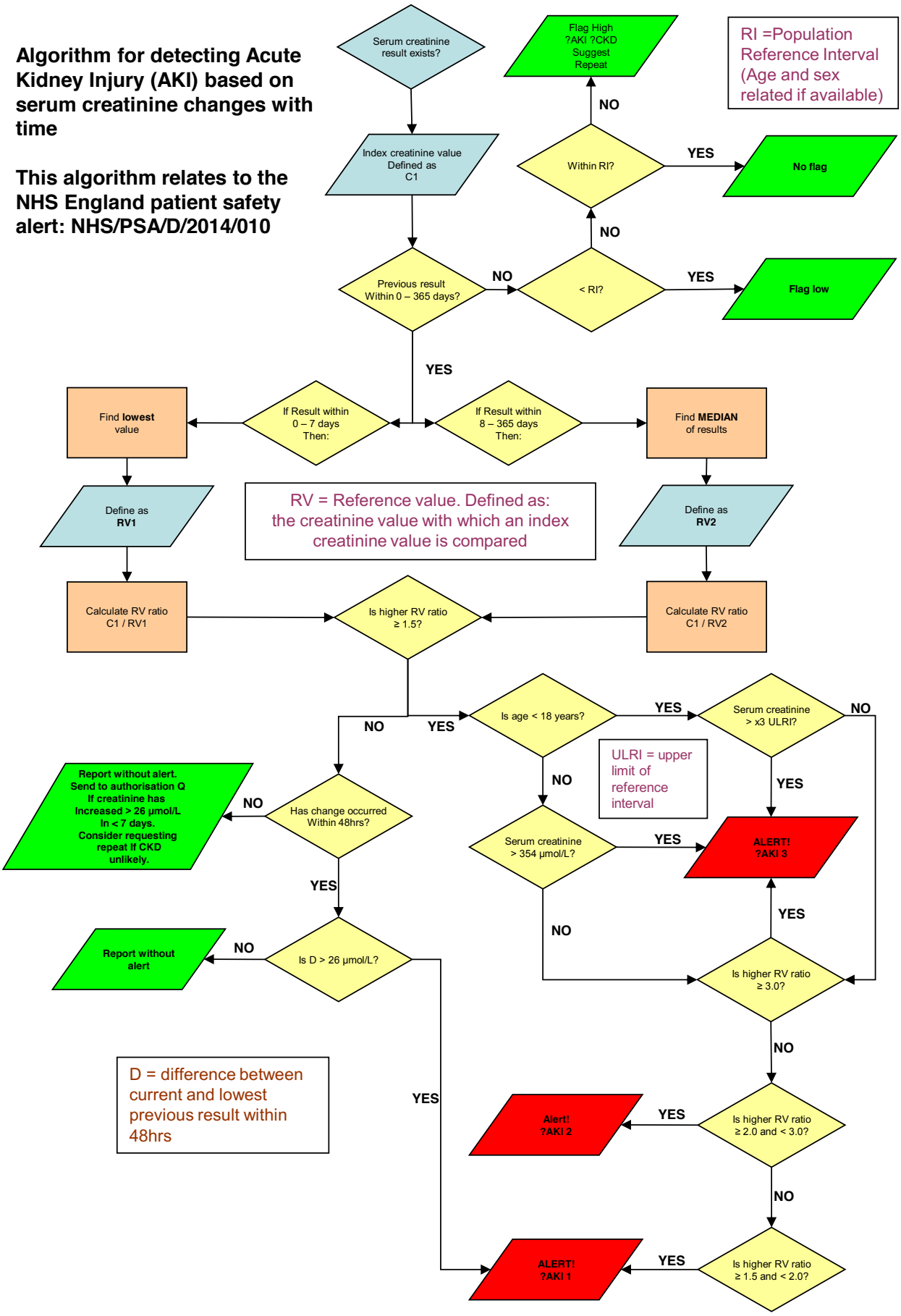
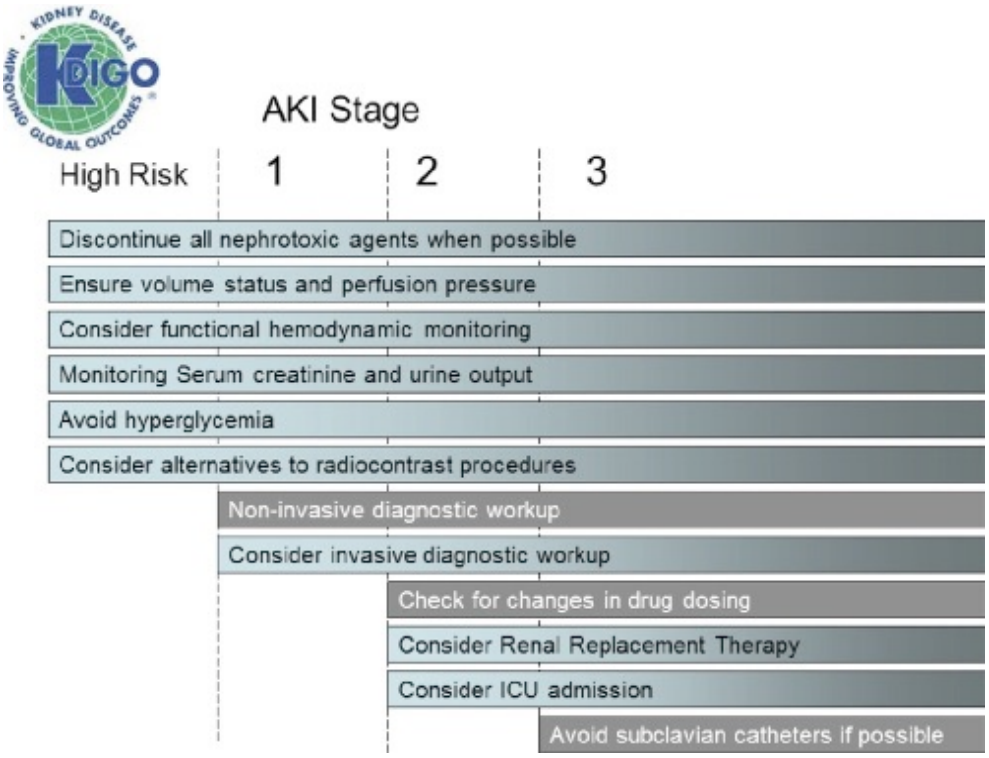


Figure 1.9: Stage-based management of AKI



Shading of boxes indicates priority of action: solid shading indicates actions that are equally appropriate at all stages whereas graded shading indicates increasing priority as intensity increases. AKI: acute kidney injury; ICU: intensive care unit.

1.4.4 Aggregated multi-parameter

The use of multiple laboratory based variables (parameters) in a formal score is not commonplace in-hospital. However, one field of medicine where occasionally an aggregated scoring system is used to modify the care a patient receives, is that of perioperative medicine. For example, for patients undergoing major surgery, some doctors calculate a score that helps them determine whether the patient should be transferred to intensive care post-operatively, and arrangements are made prior to the commencement of surgery. A recent review article by Moonesinghe et al.[90] summarises the most common of these surgical scoring systems. The models in this review use a mixture of data, comprising demographic, co-morbidities, lifestyle questions, surgical questions, vital-sign and laboratory information, in both the pre-operative and post-operative periods. In this review, P-POSSUM (Portsmouth Physiology and Operative Severity Score for enUmeration of Mortality) and the Surgical Risk Scale were highlighted as being the most validated and accurate.

Outside of perioperative medicine, the use of multi-parameter laboratory-based models for assessing the risk of mortality in *all* hospital patients has not been implemented individually or at scale. However, as individual blood test results are known to correlate with in-hospital mortality (Section 1.4.1), researchers have explored combining these with basic demographic data (i.e. age, sex), to predict risk of death. Researchers have developed a number of such models (Table 1.2).

Table 1.2: Existing aggregated multi-parameter EWSC

Model	Patient Group	No. of Sites	No. of Patients	Predictor Variables	Modelling Technique	AUROC
Biochemistry and Haematology Outcome Model (BHOM), 2005 [91]	All medical admissions	1	9,497 Train: 2,257 Test: 7,240	Sodium Potassium Urea Creatinine Ur:Cr Albumin Haemoglobin White cell count Sex Method of admission	Logistic regression	76.7 %
Hucker, 2005 [92]	Accident and emergency admissions	1	1,424 Train: 681 Test: 743	Albumin Phosphate Heart rate Age	Logistic regression	82 %
Medical Admissions Risk System (MARS), 2010 [93]	Emergency medical admissions	2	16,779 Train: 13,182 Test: 3,597	Potassium Urea Haematocrit White cell count Heart rate Mean arterial pressure Respiratory rate Temperature Age	Logistic regression	90 %
Asadollahi 'detailed model', 2011 [31]	All admissions	1	6,478 Train 1,650 Test 4,828	Sodium Potassium Urea Creatinine Chloride Glucose Bicarbonate Haemoglobin Platelets White cell count Neutrophils Lymphocytes Age	Logistic regression	86.1 %
Asadollahi 'simple model' variable, 2011 [31]				Sodium Urea Glucose Haemoglobin Platelets White cell count Age		84.8 %

Loekito, 2012 [94]	All admissions with a LoS >24 hours	2	55,818 Train 42,701 Test 13,317	Urea Creatinine Albumin Bilirubin Bicarbonate Haemoglobin Haematocrit White cell count pH	Logistic regression	88 %
Mohammed Reference Intervals, 2012 [95]	All emergency admissions	2	87,014 Train: 10,050 Test 76,964	Sodium Potassium Urea Creatinine Albumin Haemoglobin White cell count	Decision trees	86.6 %
Mohammed Raw Values, 2012 [95]						88.2 %
Laboratory Decision Tree Early Warning Score, 2013 [96]	Emergency medical admissions	1	86,472 Train: 3,496 Test: 82,976	Sodium Potassium Urea Creatinine Albumin Haemoglobin White cell count Sex	Decision trees	75.5 % - 80.1 %
Train: size of the training data set; Test: size of the testing data set; Colour Blue: vital sign parameters						

The AUROC for these multi-parameter models range from 76.7%[91] to 90%[93]. However, when only laboratory and demographic variables are used as predictor variables (i.e. vital signs are excluded), the maximum AUROC achieved falls to 88.2%. Despite some of these models having a reasonably high AUROC, the positive predictive value (when calculated) of identifying a patient likely to die is quite low. The Loekito model (AUROC 88%)[94] only attained a maximum positive predictive value (PPV) of 19% (1:5.3) on a sensitivity of 14.9% (i.e. only 14.9% of patients who died were identified). Using a different threshold with a sensitivity of 40%, the model's PPV dropped to only 10% (1:10).

Overall, a range of multi-parameter models have been built, some of which have reasonable discrimination (AUROC > 80%). However, there are a range of issues concerning these laboratory-based EWSCs, as follows.

- Most of the models were derived from single-hospital datasets and were used in sub-groups of patients (for example, emergency medical admissions or medical patients only), and thus may not be applicable to the entire patient population.
- No model takes into account the trends of any blood results; i.e. whether they have risen, fallen or stayed the same since they were last tested.
- No model analyses the time when the blood test was performed (e.g. 2 p.m. or 2 a.m.), or the time between blood tests (1 hour or 24 hours).

- None of these models include additional a priori known administrative data about the patient co-morbidities, or day and month of admission. All of these factors may have an influence on a patient's health status.
- The majority of algorithms built for surgical perioperative outcomes, and the various EWSCs, whether they are vital-sign-based or laboratory-based, use variants of logistic regression to create their respective models. The key problem with this approach is that the models assume that the dependent variables (the variables used to predict the outcome, such as heart rate, etc.) are completely independent from each other. However, this is not the case; for example, age and creatinine could be inversely correlated, as increasing age is associated with decreasing muscle mass[97], and thus with falls in creatinine.[98] Similarly, many variables rise or fall together in certain pathological processes, such as dehydration (urea, creatinine and possibly sodium) and severe malnutrition (falls in both haemoglobin and albumin, and rises or falls in mean cell volume, depending on the specific nutrients that are lacking). Although outcome prediction may be maintained despite multicollinearity in logistic models, the actual variable coefficients would be meaningless as markers for variable influence.
- If the outcome of a model is used to determine to determine escalation of care, such as activation of a RRT, the number of patients referred for every patient likely to die is low (i.e. low PPVs).
- There are newer, more appropriate analytical techniques to create models from data: specifically, advanced machine learning methods.

Thus, there is at present no laboratory-based model that has a sufficiently high PPV (along with high sensitivity and specificity) to be used as the afferent limb for a rapid response system. Advances in machine learning offer an opportunity to create an EWSC using existing laboratory and administrative data.

1.5 Machine Learning and application to health care

Machine learning (ML) is a method of automating model building using iterative algorithms which learn from the data. This method allows a machine to find insights without being explicitly programmed to look. The first reference to ML was in 1959, Arthur Samuel stated 'programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort.', this concept eventually became known as ML.[99] Tom Mitchell in his book *Machine Learning* stated that 'The field of machine learning seeks to answer the question: How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?'.[100] ML lies at the intersection of computer science, engineering, mathematics and statistics.

In the context of my research, ML can be described as a group of techniques which enable the automatic creation of models from data; models which can then be applied to new unseen data to predict characteristics/outcomes.

ML is widely applied in the modern connected world. Examples include:

- Cameras which automatically highlight and focus on a person's face in a photo or video
- Recommendations of which movie to watch next on Netflix
- Recommendations of which product to buy on Amazon
- Automatic credit card fraud detection
- Self-driving cars
- Computer voice assistants: SIRI (Apple) and Alexa (Amazon)
- Algorithmic equity (stocks/shares) trading by computers
- Automatic spellchecking and correction software
- Handwriting recognition, optical character recognition systems
- Automatic language translation, both voice (now available on Skype) and text

Even in medicine, ML is in some niche cases being applied, such as to diagnose arrhythmias from electrocardiographs (ECGs) and highlight fractures in radiographs (x-rays). [101]

Broadly, ML tasks can be divided into three categories, as follows (though in reality there are overlaps in the computational theory, and multiple methodologies may be applied to solve a particular task).

- **Supervised learning:** The computer system is provided with data that is labelled for a particular outcome; for example, blood results and mortality outcomes. Supervised learning tasks are categorised in 'classification' and 'regression' problems. In a classification problem, the goal is to predict results with a discrete output, such as whether it will rain tomorrow. In a regression task, the goal is to predict an actual value, such as how many centimetres of rain will fall tomorrow.
- **Unsupervised learning:** A computer system is provided with data, but neither a structure nor labels are defined. For example, blood results alone are provided. The task of unsupervised learning is to describe the hidden structure of the data.
- **Reinforcement learning:** This refers to learning behaviour; how to map situations to actions, so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of ML, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and-

error search and delayed reward, are the two most important distinguishing features of reinforcement learning.

The task of building an EWSC based on blood results and administrative data falls into the category of supervised learning. It is in addition a categorisation problem, as the prediction goal is whether or not a patient is likely to die in hospital.

The two most popular and accurate algorithms for the supervised learning of multivariate structured data are the random forest (RF) and gradient boosting (GB). Both RF and GB are ensemble techniques that combine multiple learning methods to achieve better predictive performance.

1.5.1 Random Forest

RF is an ensemble technique that, at its simplest, combines random decision trees with bootstrap aggregating (bagging). In this method multiple decision trees are built, each tree in this ensemble only uses a randomly drawn subset of the training data. This forest of trees then votes on the classification outcome. RF is a low-bias, high-variance method that reduces bias by increasing the number of trees, and averaging (for regression), or using the mode for classification, the predictive output of each individual tree. RFs were first described by Ho in 1995.[102] He established that if multiple decision trees were built, but each tree used only a random selection of the total number of variables, then the depth of each tree could increase, and thus increase prediction accuracy, without suffering from overtraining. RFs as they are implemented today are based on work by Leo Breiman[103], who described a method of building decision trees that combined randomised optimisation at each node with bootstrap aggregating (bagging).

1.5.2 Gradient Boosting

Gradient boosting is an ensemble technique that combines weak prediction models (high bias, low variance), typically decision trees. A model is built in a step wise fashion, weak classification algorithms are sequentially applied to a dataset, all misclassified data is used for the next iteration. Thus producing a series, in this case, of decision trees. The model is generalised by optimisation of a differentiable loss function (gradient).[104] The concept of gradient boosting was introduced in the form recognised today by Friedman in 2001. Gradient boosting is currently the best performing algorithm for classification tasks on structured data.

1.5.3 Software toolkits

There are a number of software toolkits that have optimised the implementation of GBs and RFs, as well as many other ML algorithms. The key optimisations are that the software can run on different

operating systems, or be controlled using different programming languages; but most important of all, that the implementation of the ML algorithm can be distributed across multiple cores and processors (multiple cores make up a processor), parallelising the computation and thus reducing the time it takes to complete. One popular software toolkit is 'H2O'.[105] It is implemented in the JAVA programming language, and has a version that works on the two most common technologies in the 'big data' space- 'Hadoop' and 'Spark', thus making it 'futureproof' for at least the near future. H2O has implemented a wide range of algorithms, is open-source, and is freely available to use. It has excellent documentation, and the company that develops and maintains it is quick to respond to any queries.[105]

1.5.4 Key concepts

The process of building and deploying ML models broadly involves four steps:

1. **Data collection:** Identifying and collecting the data to be used for model creation.
2. **Data pre-processing:** The process of ensuring that the collected data is clean and consistent. It includes tasks such as integrating multiple data, handling missing or inconsistent data, and converting data types to a canonical (standard) format.
3. **Model creation:** The process of using appropriate software to generate models from the data. This usually involves
    - **I.** feature engineering and selection
    - **II.** splitting the dataset into training and testing datasets (an additional validation dataset could also be created, but a better method of validation is n-1 cross-validation)
    - **III.** building the model using the training dataset (the optimal model is selected by tuning the hyper-parameters of the model and measuring the prediction errors using either cross-validation or on the validation dataset)
    - **IV.** testing the model's performance on the test dataset
4. **Model deployment: A final** model can be operationalised on an appropriate computer system, enabling live or batch predictions.

Some important additional ML concepts are:

- **Feature Engineering** is the process of extracting or selecting features (variables) from a dataset in order to enhance the dataset and improve the creation of accurate predictive models. For example, patient admission data could be enhanced by creating a feature (variable) that is the day of admission (Mon–Sun), and blood result data could be enhanced by calculating the absolute difference between consecutive blood results.
- **Generalisation** usually refers to a ML model's ability to perform well on new unseen data. It is related to the concept of overfitting.

- **Hyper-Parameter Tuning – Grid search:** ML algorithms have a variety of inputs (hyper-parameters) in addition to the training data. These inputs provide instructions for how the algorithm should model the data. These parameters determine 'higher level' properties of the model such as its complexity and how fast it should 'learn'. Grid search is the traditional method of hyper-parameter optimisation; sometimes referred to as parameter sweep, it simply refers to building multiple models from the same dataset by iterating through multiple different hyper-parameters of an algorithm. The model which has the highest performance based on a specific performance metric (logloss or AUROC), either via cross-validation or on a validation dataset (not on the test dataset), is the selected machine-learning model; the performance of this model can then be tested on the test dataset.
- **N-fold Cross-Validation:** To enable optimisation of a model and enhance its predictive accuracy, it needs to be iteratively tested on an independent dataset, modified and then re-tested. However, because the independent test dataset is used for final performance testing, an additional dataset distinct from the train dataset is required. Rather than just splitting the train dataset into two parts, n-fold cross-validation offers an alternative. For n-fold cross-validation where n = 5, the train dataset is split into five parts, four of which are used to train the model, while the fifth part is used to test it. This process is repeated a minimum of five times, with the model iteratively being improved, especially for generalisation. The mean performance metrics (AUROC, logloss) of all the models provide an indication of the final model's theoretical performance on an independent dataset.
- **Overfitting** occurs when an algorithm models random error or noise, rather than an actual relationship within the data; thus, when the resultant model is applied to an independent dataset, it performs poorly.
- **Tree depth, and leaf:** the depth of a tree (in a ML algorithm) is the number of edges from the node to the tree's root node. A tree can have multiple depths depending on the distribution of leaves. Maximum, minimum, and mean are the usual ways of communicating tree depth e.g. minimum tree depth would be the shortest number of edges from a leaf node to the root node. A leaf is a vertex of degree one in a tree (decision, boosted, etc.).
- **Train and Test Datasets:** The pre-processed dataset is split, ideally using randomisation, into two prior to model training. Ensuring that both the train and test datasets have similar outcomes and underlying characteristics. There is no universally agreed percentage split, but a 70:30 split for train:test is common.

1.6 Aims and Objectives

1.6.1 Aim

To develop a system for predicting AEs in all hospitalised patients, using ML of routinely collected blood test results and existing electronically held patient data relating to co-morbidities and demographics.

1.6.2 Objectives

1. To identify the appropriate universally accessible datasets and their specific variables to build and implement ML-EWS.
2. To map both the ethical and legal landscape required to undertake a multi-site 'big data' study.
3. To create a large dataset of ~1 million patients, and their blood results and administrative data, from multiple acute hospitals in different geographic locations in the UK.
4. To develop expertise in programming, to undertake large-scale data capture and manipulation, and to implement ML models.
5. To investigate mortality associated with known models of disease. Specifically:
    a. For Dehydration,
        i. To understand the effect of dehydration (Ur:Cr) on outcome
        ii. To understand changes in dehydration (Ur:Cr), combined with AKI, on outcome
        iii. To create a model that incorporates urea and creatinine results with simple demographic data to identify those at risk of poor outcome
    b. For AKI,
        i. the epidemiology of patients admitted to hospital who are diagnosed with AKI
        ii. the relationship between the NHSE-AKI algorithm defined AKI stage and in-hospital outcome (Death or Renal Replacement Therapy (Drrt)), and whether this relationship differs according to method of admission and existing co-morbidities
        iii. whether the Drrt risk differs depending on the route by which the AKI stage is defined
        iv. whether the NHSE-AKI algorithm fails to identify patients who continue to have AKI
        v. whether a ML approach can better stratify risk than the current NHSE-AKI algorithm.
6. To build ML models that can be used on all patients in hospital in order to identify their risk of dying in hospital, both on admission and subsequently.
7. To build a proof-of-concept computer system that is agnostic to the internal IT infrastructure of a hospital, but which can deploy advanced ML models by ingesting hospital data and relaying results to the hospital/clinician in real time.

Chapter 2: The Data, Ethical and Legal Frame

To undertake an investigation using electronically stored data, the ideal source for these data must be identified and understood. These sources of data must be both accessible for this investigation but also amenable to real time access, so they may be used for the real time implementation of the models developed. Issues with data access and use must be explored, specifically the legal and ethical framework as well as the technical issues of acquisition and transformation. Investigation into these issues, have highlighted their complexity and frequently conflicting framework.

In this chapter, I will describe in detail

1. **the data,**
2. **the ethical and legal framework, and**
3. **data acquisition and transformation,**

providing their context, and highlighting the challenges I faced and how I overcame them.

2.1 The Data

The data I required for my study included administrative information: patient demographics, admission and discharge dates, methods of admission, as well as information relating to patient diagnoses and treatments in-hospital. There are many electronic sources in a hospital that store patient information, ranging from the electronic health record (EHR), patient administration system (PAS), theatre booking system, dedicated diagnostic coding software, etc., all of which vary significantly between hospitals. As my investigation was requesting data from multiple NHS Trusts, a common dataset and a standard extraction method of requesting the data needed to be found. Rather than creating my custom requirements, I chose to align my data request to match one of the national mandatory data submissions that all hospitals are required to perform- the Commissioning Dataset for Admitted Patient Care (CDS.APC).

In addition, I needed access to blood test results, unfortunately there is no common standard for their use, and custom extractions needed to be performed at each NHS Trust in my study. As opposed to the CDS.APC data, blood results are used almost exclusively for clinical purposes.

Thus, the two existing data sources I used to capture patient trajectory, and thus develop models to predict patient trajectory, were:

1. The Commissioning Dataset for Admitted Patient Care (CDS.APC)
2. Blood results

These datasets have in common that they are both created, stored, transmitted and interrogated or viewed via electronic means, and that the entirety of the data is stored in multiple databases. Theoretically, this enables easy access to the entirety of the data. The specific characteristics of each dataset are described below.

2.1.1 CDS.APC (Commissioning Data Set for Admitted Patient Care)

The CDS.APC dataset is a national UK dataset; such datasets 'define a standard set of information that is generated from care records, from any organisation or system that captures the base data. They are structured lists of individual data items, each with a clear label, definition and set of permissible values, codes and classifications'.[106]

Commissioning Data Sets (CDS) are patient-level datasets intended to deliver robust, comprehensive, nationally consistent and comparable person-based information on activity, in order to support a variety of secondary use purposes (i.e. not for the direct care of the patient). The CDS are the primary mechanism for the reporting of secondary care (in-hospital) activity that is either NHS funded and/or provided by NHS organisations. They support a number of high-profile national NHS requirements, including the current HealthCare Resource Group (HRG) version for the calculation of payment to trusts, and monitoring or other initiatives such as Referral to Treatment (RTT), as well as the national reporting of activity through Hospital Episode Statistics (HES). The dataset I have focussed on is the 'Admitted Patient Care – Finished General Episode CDS' (CDS.APC). This covers all NHS and private Admitted Patient Care (day case and inpatient) activity taking place in any of the following: acute, community or mental health NHS trusts, other NHS hospitals, non-NHS hospitals, and non-NHS hospitals or institutions where the care delivered is NHS-funded. The complete list of CDS datasets is provided in Appendix 1[106], highlighting the complexity of the various datasets available in terms of number as well as similarity.

The CDS.APC dataset is created by multiple teams within an organisation. The nature and composition of these teams varies by organisation, and I did not find a standard organisational framework that dictates who collects, validates, transforms and submits the data to Health and Social Care Information Centre (HSCIC, recently restructured into a new organisation NHS Digital). Frequently, different tasks are undertaken by different departments within the organisation, with some tasks also being outsourced to commercial entities. This results in multiple different workflows within and between different organisations, for what is a mandatory and standardised national data submission. In one organisation, a common example would be as follows: post discharge of a patient from hospital, the patient's notes would be transferred to the clinical coding team, who would examine these to

determine not only what diseases the patient suffers from chronically, but also any specific acute conditions that were treated in their most recent hospitalisation. This information would form the basis of the patient's ICD10 (International Classification of Diseases Version 10), Treatment Function and Main Specialty codes. However, in another organisation, the admitting or discharging doctor would be responsible for listing all the relevant patient's diseases, enabling the coding team to easily assign ICD codes based on a clinical summary. OPCS-4 (OPCS Classification of Interventions and Procedures, version 4) codes are similarly transcribed by different teams, from a variety of sources, which include operating theatre lists, radiology intervention lists, day case procedure lists, and again, from the examination of raw text from the patient's notes. The potential for mislabelling and omitting relevant ICD10 and OPCS-4 codes is significant; and indeed, analysis of Hospital Episode Statistics for CDS.APC data has highlighted such errors[107] : e.g. every year, a number of male patients have been coded to have had a hysterectomy (removal of the uterus).

The data for each patient episode (row) in this dataset is usually compiled after the patient has completed their 'episode' of care in their Secondary Care Organisation. In the context of this research project, this will always be an NHS Trust. It is important to note that a patient may have multiple 'episodes' of care during one hospital admission. The combination of all applicable patient 'episodes' makes up a patient's 'spell', which is equivalent to a single hospital admission. Episodes are meant to represent contained healthcare activity under a single named consultant, and when a patient's care is transferred to another consultant, a new episode begins. The data elements that comprise the CDS.APC dataset are extremely comprehensive, though not all the fields present are necessarily filled. The full list of all CDS.APC variables is available online at: 'http://www.datadictionary.nhs.uk/data_dictionary/messages/cds_v6-2/data_sets/cds_v6-2_type_130_-_admitted_patient_care_-_finished_general_episode_cds_fr.asp?shownav=1'.
The key variables I required from organisations, and their attributes, are described in the following tables. These are a description of the raw data, and where mentioned, the equivalent anonymised/grouped data that was provided by the collaborating NHS Trusts for this project.

- **Organisation details:** The organisation code of the secondary care provider: a series of 3–4 alphanumeric characters denoting the organisation code.
- **Patient characteristics** (Table 2.1)
- **Hospital provider spell characteristics** (Table 2.2)
- **Treatment function code** is recorded to report the specialised service within which the patient is treated. It is based on 'Main Specialty', but also includes approved sub-specialities and treatment specialties used by lead 'Care Professionals'.[106]
- **ICD10 codes:** These are listed as comprising one primary diagnosis, and up to fifty secondary diagnoses. The order of the diagnoses is not usually based on the guidance by healthcare

professionals (based on discussion with the clinical coders), but is usually ordered so as to maximise the Health Care Resource Group Code, and thus generate maximum income for the Trust. The International Statistical Classification of Diseases and Related Health Problems (ICD) is a comprehensive classification of causes of morbidity and mortality. All inpatient episodes and attendances that contain diagnoses must be recorded to the mandated version of ICD. 'ICD-10' refers to the tenth revision. The World Health Organisation (WHO) publishes and distributes the ICD-10 classification. WHO is the copyright holder of ICD-10, which is used under licence for United Kingdom government purposes. The full list of ICD-10 codes is available at: 'http://apps.who.int/classifications/icd10/browse/2010/en'.

- **OPCS-4 codes:** Upto fifty OPCS-4 codes are listed, along with the date on which said procedure/intervention was performed. The Classification of Surgical Operations and Procedures was originally issued by the Office of Population Censuses and Surveys (OPCS). The fourth revision was first implemented in hospital information systems in 1987. This was subject to a significant number of amendments, and a consolidated version was reproduced in 1990. The OPCS Classification of Interventions and Procedures (OPCS-4.2) was substantially enhanced to ensure that modern clinical practice was represented appropriately within the classification, and a new version was implemented in 2006, titled 'OPCS Classification of Interventions and Procedures' (OPCS-4.3), with a commitment to undertake an annual review and potential update. The current version is OPCS-4.7. The classification comprises a list of alphanumeric codes with mainly anatomically based chapters, most of which relate to the whole or part of a body system. Each chapter is designated alphabetically; e.g. Chapter A covers the nervous system and Chapter K is assigned to the heart. The alphabetic character for each chapter forms the prefix of the 3- and 4-digit codes within it. The strict link between chapters and body systems, with specific procedures for individual organs, was broken in OPCS-4.3 because of limited capacity. A full list of OPVSv4 codes is available at 'http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4'.
- **CCMDS:** The data in the Critical Care Minimum Data Set primarily relates to any part of the patient's hospital spell that requires care in a designated critical care bed. Although this information was requested, the amount of missing data and poor quality of the data received from the participating NHS Trusts led to its exclusion from my research.

The key issues of accessing these data are the difficulty of using varying local data schemas, which would have to be transformed post-collection to the national standard, and also of determining which team in each collaborating Trust could provide the data.

Table 2.1: Patient characteristics

Raw Data	Attributes		Requested	New Attributes
NHSNumber	10 digit numerical, with internal validation structure		Anonymised NHSNumber	Variable length and variable character encoding
Local Patient Identifier (LPI) / Medical Record Number (MRN)	Custom per organisation variable length with both alphanumeric characters		Anonymised LPI/MRN	Variable length and variable character encoding
Postcode of usual address	UK postcodes		Only the first half of postcode (3/4 characters)	Max length of 4 alphanumeric characters
Patient Birth Date	Day, month and year of birth		Only the Month and Year of Birth	Month and year of birth- numeric format
Person Gender	0: not known, 1: Male, 2: Female		Same	0,1,2 but also arrives as M:Male, F:Female or U:unknown
Ethnic Category	**Caucasian**		Same	Same
	A	British		
	B	Irish		
	C	Any other white background		
	Mixed			
	D	White and Black Caribbean		
	E	White and black African		
	F	White and Asian		
	G	Any other mixed background		
	Asian or Asian British			
	H	Indian		
	J	Pakistani		
	K	Bangladeshi		
	L	Any other Asian background		
	Black or Black British			
	M	Caribbean		
	N	African		
	P	Any other black background		
	Other Ethnic Groups			
	R	Chinese		
	S	Any other ethnic group		
	Z	Not stated		

Table 2.2: Hospital provider spell characteristics		
Variable	**Attributes**	
	National Code	**Description**
Admission Method	**Elective Admission**	
	11	Waiting list
	12	Booked
	13	Planned
	Emergency Admission	
	21	Accident and emergency or dental casualty department of the Health Care Provider
	22	GENERAL PRACTITIONER: after a request for immediate admission has been made direct to a Hospital Provider, i.e. not through a Bed bureau, by a GENERAL PRACTITIONER or deputy
	23	Bed bureau
	24	Consultant Clinic, of this or another Health Care Provider
	25	Admission via Mental Health Crisis Resolution Team
	2A	Accident and Emergency Department of another provider where the PATIENT had not been admitted
	2B	Transfer of an admitted PATIENT from another Hospital Provider in an emergency
	2C	Baby born at home as intended
	2D	Other emergency admission
	28	Other means
	Maternity Admission	
	31	Admitted ante-partum
	32	Admitted post-partum
	Other Admission	
	82	The birth of a baby in this Health Care Provider
	83	Baby born outside the Health Care Provider except when born at home as intended.
	81	Transfer of any admitted PATIENT from other Hospital Provider other than in an emergency
Source of Admission	19	Usual place of residence unless listed below, for example, a private dwelling whether owner occupied or owned by Local Authority, housing association or other landlord. This includes wardened accommodation but not residential accommodation where health care is provided. It also includes PATIENTS with no fixed abode.
	29	Temporary place of residence when usually resident elsewhere (e.g. hotels,

		residential Educational Establishments)
	39	Penal establishment, Court, or Police Station / Police Custody Suite
	49	NHS other Hospital Provider - high security psychiatric accommodation in an NHS Hospital Provider (NHS Trust or NHS Foundation Trust)
	51	NHS other Hospital Provider - WARD for general PATIENTS or the younger physically disabled or A & E department
	52	NHS other Hospital Provider - WARD for maternity PATIENTS or Neonates
	53	NHS other Hospital Provider - WARD for PATIENTS who are mentally ill or have Learning Disabilities
	54	NHS run Care Home
	55	Hospital site within the same Trust - ward for general patients or young physically disabled or A & E department
	56	Hospital site within the same Trust - ward for maternity patients or neonates
	57	Hospital site within the same Trust - ward for patients who are mentally ill or have learning disabilities
	65	Local Authority residential accommodation i.e. where care is provided
	66	Local Authority foster care
	79	Babies born in or on the way to hospital
	85	Non-NHS (other than Local Authority) run Care Home
	86	Non-NHS (other than Local Authority) run Nursing Home
	87	Non NHS run hospital
	88	Non-NHS (other than Local Authority) run Hospice
	99	Not known: a validation error
Discharge Method	1	PATIENT discharged on clinical advice or with clinical consent
	2	PATIENT discharged him/herself or was discharged by a relative or advocate
	3	PATIENT discharged by mental health review tribunal, Home Secretary or Court
	4	PATIENT died
	5	Stillbirth
	8	Not applicable - hospital provider spell not finished at episode end (i.e. not discharged, or current episode unfinished)
	9	Not known: a validation error
Discharge destination	19	Usual place of residence unless listed below, for example, a private dwelling whether owner occupied or owned by Local Authority, housing association or other landlord. This includes wardened accommodation but not residential accommodation where health care is provided. It also includes PATIENTS with no fixed abode.
	29	Temporary place of residence when usually resident elsewhere (includes hotel, residential Educational Establishment)

	30	Repatriation from high security psychiatric accommodation in an NHS Hospital Provider (NHS Trust or NHS Foundation Trust)
	37	Court
	38	Penal establishment or police station
	48	High Security Psychiatric Hospital, Scotland
	49	NHS other Hospital Provider - high security psychiatric accommodation
	50	NHS other Hospital Provider - medium secure unit
	51	NHS other Hospital Provider - WARD for general PATIENTS or the younger physically disabled
	52	NHS other Hospital Provider - WARD for maternity PATIENTS or Neonates
	53	NHS other Hospital Provider - WARD for PATIENTS who are mentally ill or have Learning Disabilities
	54	NHS run Care Home
	55	Hospital site within the same Trust - ward for general patients or young physically disabled or A & E department
	56	Hospital site within the same Trust - ward for maternity patients or neonates
	57	Hospital site within the same Trust - ward for patients who are mentally ill or have learning disabilities
	65	Local Authority residential accommodation i.e. where care is provided
	66	Local Authority foster care
	79	Not applicable - PATIENT died or still birth
	84	Non-NHS run hospital - medium secure unit
	85	Non-NHS (other than Local Authority) run Care Home
	86	Non-NHS (other than Local Authority) run Nursing Home
	87	Non-NHS run hospital
	88	Non-NHS (other than Local Authority) run Hospice
	98	Not applicable - hospital provider spell not finished at episode end (i.e. not discharged, or current episode unfinished)
	99	Not known: a validation error
Start Date Hospital provider Spell	is the date of admission: the CONSULTANT or MIDWIFE has assumed responsibility for care following the DECISION TO ADMIT the PATIENT.	
Age on Admission	is derived as the number of completed years between the PERSON BIRTH DATE of the PATIENT and the START DATE (HOSPITAL PROVIDER SPELL)	
Discharge Date (HOSPITAL PROVIDER	DISCHARGE DATE (HOSPITAL PROVIDER SPELL) is the date a PATIENT was discharged from a Hospital Provider Spell.	

SPELL)	
Episode number	is the same as attribute ACTIVITY IDENTIFIER and is used to uniquely identify episodes, and is a sequence number for each Consultant Episode (Hospital Provider) in a Hospital Provider Spell. The first episode of each new Hospital Provider Spell (including re-admitted PATIENTS) commences at 01. A known episode number can be between 01 - 87
LAST EPISODE IN SPELL INDICATOR CODE	is a derived data element which identifies whether the consultant episode is the final episode in the Hospital Provider Spell Data attributes: 1: This episode is the last episode in the Hospital Provider Spell. 2: The episode is not the last episode in the Hospital Provider Spell.
Episode Start Date	start date of the patient episode
Episode End Date	date of completion of this specific patient episode

2.1.2 Blood results

Blood test results form a core component in over seventy percent of all clinical interventions. They are used in determining diagnosis, choice of treatment, and efficacy of treatment. In England and Wales alone, over 300,000 patients undergo a blood test every working day, equating to nearly 800 million tests performed annually. Of these 800 million, 500 million are biochemistry tests and 130 million haematology tests.[108]

For this thesis I have focussed on routine biochemistry and haematology tests, and how these routine measures can be used in AE prediction. Biochemistry and haematology tests are ordered for almost every patient, either immediately prior to admission, or on admission to hospital. In addition, hospital laboratories frequently provide pathology services to patients in the community (rather than admitted patients), and these results are also available. The specific data variables I requested for blood results are described in Table 2.3. I decided on these variables after examining the various potential sources of blood result data, the most common being the laboratory information management system (LIMS).

Table 2.3: Blood Result Characteristics

Variable Name	Attributes	Requested	New Attributes
TestName	Name of the blood result test performed e.g. haemoglobin	Same	Same
Result value	Full result	Same	Same
Result numeric	Numerical component of the result value	Same	Same
Units	e.g. mg/L	Same	Same
Sample source	Source of the sample e.g. venous blood	Same	Same

Location the test was performed	The physical location (hospital site, or ward)	Same	Deleted
Upper reference range	Numerical value	Same	Majority missing
Lower reference range	Numerical value	Same	Majority missing
Sample collected date and time	Date and time the sample was collected from the patient	Same	Deleted
Sample received in the laboratory date and time	Date and time the sample was logged as being processed in the laboratory	Same	Deleted
Validated date and time	Date and time the result of the test was validated/available	Same	Same
NHSNumber	10 digit numerical, with internal validation structure	Anonymised NHSNumber	Variable length and variable character encoding
Local Patient Identifier(LPI) / Medical Record Number(MRN)	Custom per organisation variable length with both alphanumeric characters	Anonymised LPI/MRN	Variable length and variable character encoding

LIMS is the software that underpins the pathology laboratory to perform its operations. The LIMS is a regulated system, specifically licensed and approved for the processing of pathology samples. The LIMS enables:

- Processing the order for a sample from the order communication system of the requesting entity.
- Reception and log-in of the sample and its associated patient data.
- Assignment, scheduling and tracking of the sample and its subsequent analysis.
- The process and quality control associated with the sample and the equipment that was used for its analysis.
- Storage of all data associated with both the inputs and the outputs of the analysis.
- The inspection, approval and reporting of the results of the analysis.

Additional features are frequently added onto these systems, such as triggers for alarms based on the results of certain tests, clinical decision support, rules for ordering additional tests, and triggers for secondary verification, to name but a few.

Such LIMS have been ubiquitous in the NHS, and most developed healthcare systems. As such, there is a vast repository of pathology data theoretically available for analysis, however because 1) the LIMS are commercially proprietary closed systems, and 2) the contracts with the LIMS providers restrict any changes to the system setup, this potential has not, until now, been realised. The technology architecture powering the LIMS is not usually shared with customers, and even when known, access to the data within the LIMS is only possible through specific non-standard interfaces, which are not

designed for large-scale data extraction or querying. It is understandable that in a critical system, processes and queries that are not related to the immediate task of live sample and result management do not interfere with the running of a laboratory. However, the difficulty and cost of customising the LIMS for additional tasks impedes research, and ultimately limits the further integration of pathology results into clinical pathways. LIMS providers frequently charge substantial consultancy fees for these 'additional' services, such as data extraction. In fact, during the course of this project, I received quotes of £50,000–100,000 for accessing pathology results from different Trusts' LIMS (personal communication).

In addition to the technical and fiscal challenges to pathology data access, there is no robustly adhered-to information standard of nomenclature for the tests carried out, or for samples. This means that the same test performed on the same type of sample may be referred to differently in different organisations.

> *'Most laboratories and clinical services use HL7 (the medical messaging standard) to send their results electronically from their reporting systems to their care systems. However, the tests in these messages are identified by means of their internal, idiosyncratic code values. As a result, the receiving care system cannot fully "understand" and properly file the results they receive unless they either adopt the producer's test codes (which is impossible if they receive results from multiple sources), or invest in the work to map each result producer's code system to their internal code system'* [109].

In fact, HSCIC succinctly defined the problem as follows:

> *'Up until now, there has been no way of reporting pathology results in a consistent, standardised way across the country. Different names in different settings could have meant the same or different things. This has led to problems. A patient may have a test in one region which could be interpreted as a different test in another region and, consequently, might either have to repeat the same test or might not receive a test they need. Without official names for tests, or a national mechanism to agree whether or not a test was worth using, there has been potential for misinterpretation, confusion and waste'* [110].

In 1994, Clem McDonald, an investigator at the Regenstrief Institute, begun to address this problem, particularly because of the electronic transmission of blood results between organisations. He setup the LOINC committee in 1994, whose task was to develop a common terminology for laboratory results.[109] 'LOINC' refers to Logical Observation Identifiers Names and Codes[111] and is now the unofficial international standard. In the last decade, NHS England has taken steps to try to mitigate

this issue as well, with the development of the National Pathology Exchange (NPEx) and of the National Laboratory Medicines Catalogue (NLMC)[110] . NPEx connects NHS pathology laboratories across the UK to a national hub, so that test referrals and results can be sent between laboratories electronically in a few seconds. The NPEx service is available to all NHS pathology labs in the UK at no cost other than a modest support fee. Connecting to the national exchange allows a lab to use its lab system (LIMS) to send its test requests to any other lab using the exchange, and receive the results back electronically, directly into their LIMS. The NLMC is a list of pathology tests that have been validated for use within the NHS. A team of Pathology Discipline Leads, led by a National Clinical Lead, defines the tests, which are then reviewed by Subject Matter Experts. Each item in the list is coded based on a number of core attributes that make each definition unique. This means that hospital doctors, GPs, nurses and other health professionals can be certain they are requesting the right test every time. However, despite being a national standard and essentially free to use, these systems are not ubiquitous across the NHS. Individual Trusts still use their own standards and nomenclature, possibly due to the costs of upgrading their LIMS and all allied systems to this new standard. This was estimated at between 5–10 million pounds for one large multisite trust (personal communication). As the NLMC is still a work in progress, I therefore decided to use the international LOINC nomenclature for my canonical standard.

In summary, the two major technical challenges I needed to overcome in order to access and utilise pathology data were:

1. Access to the pathology data from various LIMS and/or other databases at each NHS Trust.
2. Interpretation of the data: conversion of the pathology data into a canonical standard.

One key issue I have not addressed is whether pathology results can be compared between healthcare organisations. The answer to this query should undoubtedly be 'YES'. From the point of view of the end user, either the healthcare professional looking after the patient, or indeed the patient him/herself, they both interpret a pathology result as being absolute and comparable, regardless of where it was obtained. In reality, the result is only bound by the quality assurance standard of the local organisation, which does not guarantee interchange between organisations, though the National Pathology Programme is trying to address this. I have specifically decided to not address this issue, partly because it is a task that is unachievable in the scope of my research, but also because for an individual patient episode/spell, all pathology results would be obtained from the same Trust.

2.2 The Ethical and Legal Framework

Before any patient data could be requested and used, the appropriate ethical, legal and information governance permissions were needed. However, identifying the precise requirements of these permissions based on the work I wanted to undertake was not an easy task, as described below.

The ethics of conducting medical research are based upon the World Medical Association's Declaration of Helsinki.[112] This is a statement of the ethical principles for medical research involving human subjects, including research on identifiable material and data. In the UK, the research governance framework for health and social care[113] outlines the principles of good governance that apply to all research within the NHS. Both these documents emphasise that 'The dignity, rights, safety and wellbeing of participants must be the primary consideration in any research study', and that 'informed consent is at the heart of ethical research'.

For research focussing on data, there are two broad types of data available:

- Research data is created/collected solely for the purpose of the research study, and is not used for the direct care of the patient. This requires, in the vast majority of cases, informed consent, and specific legal safeguards must be in place for individuals who cannot give such consent.
- Healthcare data is primarily created for one specific purpose: the direct care of an individual patient. There are, however, a multitude of secondary uses (not for direct patient care) of this data. Capturing the details of a patient, their co-morbidities, their acute condition, the specific procedures and interventions carried out in their care, and various administrative details, allows a patient's medical interventions to be categorised, costed and appropriately reimbursed. Reimbursement to the healthcare establishment (and individuals) for providing medical care is in fact the leading use of such data, after its utilisation for direct patient care. Whether this reimbursement be via a government organisation, such as the NHS in the UK or via Medicare/Medicaid in the USA, via health insurance, or from the patient him/herself, the principle of the secondary use of this data is the same.

The legal and ethical issues arise when data collected for one purpose, in this case for the direct care of the patient, is utilised for another; specifically, research. There is, on the surface, conflicting guidance and advice from the UK Health Research Authority (HRA) (www.hra.nhs.uk). The HRA guidance states that, regarding NHS Research,

> *'Research limited to secondary use of information previously collected in the course of normal care (without an intention to use it for research at the time of collection) is generally excluded from REC (research ethics committee), provided that the patients or*

services users are not identifiable to the research team carrying out the research. This exception also applies to research undertaken by staff within a care team using information previously collected in the course of care for their own patients or client, provided that data is anonymised in conducting the research.' [114]

However, within the REC application process, there exists a subset of questions relating specifically to research on anonymised data, and recommending submission for REC approval. Specifically, an REC review is required if a specific research project involves:

'use of previously collected tissue or information from which individual past or present users of these services could be identified, either directly from that tissue or information, or from its combination with other tissue or information in, or likely to come into, the possession of someone to whom the tissue or information is made available.'

Also, both university and NHS Trust research offices advise that any research conducted on even the data of NHS patients must be submitted to a REC for approval before the research can be considered, either for sponsorship or for initiation as a study.

The Health Research Authority's own tool to determine whether a study utilising previously collected data needs REC approval (http://hra-decisiontools.org.uk/ethics/) makes its decision solely on the potential of whether the supposedly anonymised data obtained could, if combined with any other data, be used to identify any patients by members of the research team. There is thus a 'grey area', in both legal and ethical terms, when conducting such research. This is further compounded by the fact that the data required for research frequently does not exist in an anonymised form, and there is no existing formal mechanism to carry out this anonymisation prior to the research team's involvement. Thus, identifiable data must be accessed and manipulated by a team that is distinct from both the clinical and research teams, purely for the reason of carrying out research.

It is important to note here that REC approval is not a confirmation of the legal basis for conducting research or accessing and utilising data. As stated in the governance arrangements for research ethics committees[115]:

'It is not the role of the REC to offer a legal opinion on research proposals.

Researchers, sponsors and organisations where the research is carried out remain responsible for making sure the research is conducted in accordance with the requirements of law, relevant regulators and guidance, e.g. the Data Protection Act, the Codes of Practice issued under the Mental Capacity Act and Human Tissue Act, or recognised standards of Good Clinical Practice.'

Additional legal requirements that healthcare data used for research must comply with are 1) the European Union Directive 95/46/EC: The Data Protection Directive (which is in the process of being

replaced by the General Data Protection regulation adopted in April 2016); and 2) the UK Human Rights Act (based on the European Convention on Human Rights). However, it is unknown how long these European directives will continue to apply in the UK, following the UK's decision to leave the European Union.

2.2.1 The ethical and legal considerations for conducting research with 'grey area' anonymised data

First of all, we need to explore the precise definition of anonymised data. The word 'anonymous' comes via late Latin from the Greek *anōnumos* 'nameless' (from *an-* 'without' + *onoma* 'name'). It refers to a person, or in this case data, that cannot be identified or attributed to a specific name (person). However, the problem with the concept of anonymised data is that there must be sufficient obfuscation of the data to ensure that any individual in the dataset cannot be identified directly from the data itself, but it may not be possible to ensure this if the data is combined with any other information. The UK Information Commissioners Office (ICO) has produced a code of practice, 'Anonymisation: managing data protection risk' ('the code'), to govern the anonymisation of identifiable data.[116] The ICO acknowledges within 'the code' the risk of re-identification, and, dependent on the potential consequences of such re-identification, recommends either (1) limiting access to this 'grey data', along with implementing robust governance processes; (2) more rigorous risk analysis and anonymisation; or (3) seeking the data subjects' consent for the disclosure of this anonymised data, explaining its possible consequences.

Dame Fiona Caldicott, in both her original 1997 Caldicott Review[117], which led to the appointment of Caldicott Guardians, and her subsequent Information Governance Review[118] in 2013, also concurred with and drew on the ICO's Anonymisation Code of Practice.[116] She has suggested a simple framework that describes three different forms of data, and the conditions under which data can be processed and disclosed. These are:

- **Data for publication:** This is data that has been anonymised in line with the ICO anonymisation code to the point where determining individual identities from the data is unlikely, requiring unreasonable effort. The data does not require a legal or contractual basis for processing it, and can be publicly disclosed. This data is called 'de-identified data for publication'.
- **Personal confidential data:** This is data in which individuals are clearly identified, or are easily identifiable. This data should not be processed without a clear legal basis. Personal confidential data should only be disclosed with consent or under statute, and any disclosure must always be limited and accompanied by a contractual agreement that mitigates the risk of misuse and inappropriate disclosure. The contractual agreement needs to set out, as a minimum, the legal

basis for the data flow, the purposes to which the data can be put, the safeguards that should be in place to protect data, and how the public are informed about these purposes.

- **Data for limited disclosure:** This data is called 'de-identified data for limited disclosure or access'. This is data that has been through a process of anonymisation, such as removing formal personal identifiers, using coded references or pseudonyms in their place, or by aggregating data together so that it is not possible to identify individuals. However, it would be relatively straightforward for a third party to re-identify individuals or de-anonymise the data, especially if combined with other data. This represents the 'grey area' of data.

However, nowhere is the distinction clearly defined between what precisely constitutes the truly anonymised dataset and this grey area of data that could potentially, at some point in the future, be de-anonymised. There are numerous mechanisms to perform this process, one of which involves linking the 'anonymised' dataset to another, accessible dataset. There are numerous very large population-level datasets that are available, either free of charge or at a cost. These datasets include national data, such as the electoral register, London Oyster Card usage, and 'Boris Bike' use, which include a mixture of identifiable and pseudonymised data. In addition, 'cookies' on websites and 'apps' on smartphones can effectively track a user's location and browsing habits, and identify them by age, sex, ethnicity, and even by name. Social media provides intimate knowledge about many individuals, and local/national news has information about specific individuals. Any of these datasets could potentially be linked to supposedly anonymised medical data. An individual may him/herself overtly, though inadvertently, enable such linkages to take place, by their personal use of Twitter, which has minimal data protection, and which thrives on the 'openness' and transparency of its platform. I could, using Twitter's own API (application protocol interface), download all tweets from a certain location and from a certain time that state something specific, such as 'arrived in hospital'. Thus, if I pre-emptively want to find a data-point in an anonymised dataset that links to a known identifiable person using freely available and publicly shared data, this may potentially be possible.

In this environment of lack of clarity, and lack of communication with the ultimate providers of the data, i.e. patients, is the UK government's commitment to encouraging medical research. The National Data Guardian Consultation paper[119] clearly states:

> *'Medical research relies on people's health data to develop new medicines and treatments to transform and save people's lives. Therefore, the Government is committed to do all it can to encourage people to allow their data and information to be used to help realise the opportunities for further progress. There is a need to continue to build on the public's confidence in how data about individuals is used and safeguarded.'*

These are some of the key challenges in the 'big data' world, which research and medical providers have, unfortunately, yet to completely grasp. This shortfall in both guidance and mechanisms to ensure the principles of ethics and data protection for large patient datasets is still not being adequately addressed. These problems are impeding both medical research and the linkage of patient data across organisations. Nowhere is this more apparent, than in what could be regarded as NHS England's flagship large scale data project 'Care.Data'. Understanding the controversies surrounding Care.Data was necessary to prevent my study encountering the same issues.

2.2.1.1 Care.Data

NHS England's Care.Data programme[120] aim is to extract data from all general practitioners (GPs) and upload and link this information to the existing national databases maintained by the HSCIC (renamed 'NHS Digital' in June 2016). However, the implementation of this initiative has been dubbed a 'fiasco'[121] and denounced by the press[122], and privacy groups.[123] The issues that led to and continue to plague the Care.Data initiative are the problem of 'implied consent', and the potential sharing of this linked database with non clinical organisations not only within the NHS (e.g. clinical commissioning bodies), but also outside it (e.g. pharmaceutical companies, insurance companies and consulting companies, as well as universities, think-tanks and health charities). As previously mentioned, if data is not being used for the direct care of the patient, then it is considered as 'secondary use'. Normally, this would require consent. However, this requirement was bypassed by the Health and Social Care Act 2012, which legally obliges GPs to hand over patient data if NHS England directs HSCIC to request it. This Act put the GPs in an ethical and legal quandary, where they had to act in the best interests of their patients, but at the same time were forced to share their patients' data. As such, a large number of GPs have objected to the Care.Data initiative. Even the patient's choice to opt out of handing over this data is not a legal requirement, but due to the initial public outcry against this initiative, a mechanism was hastily put in place. Two additional perceptions are affecting the Care.Data initiative: firstly, a lack of trust in the government to actually do what it says it is doing with the data (this has been fuelled by the recent Edward Snowden revelations); and secondly, that creating such a central store of all the data makes it vulnerable to being 'hacked', and thus private patient data is no longer private and becomes freely available. For all of the above reasons, the Care.Data initiative remained in a pilot phase, with national rollout being repeatedly postponed. Finally, in July 2016 the Care.Data programme was officially suspended, however it is thought that the sharing of patient data would still take place, but be rebranded as another initiative. Failure to make the case for personal health data linkage is regarded, by NHS England (personal communication), as the primary reason for failure of the original Care.Data initiative.

2.2.2 My Approach

Considering the granularity and breadth of information I would be requesting, and the access that I myself as a doctor have to clinical information systems at many of the Trusts providing the data, it would be inappropriate, and indeed illegal, to treat the data as being completely anonymised. My study is different from the Care.Data initiative, in that although I would be requesting the linkage of two normally separate datasets (CDS.APC and Pathology), the linkages would occur within each NHS Trust. Anonymisation of these linked data would also take place within the confines of each NHS Trust, with no sharing of any identifiable data outside each organisation.

Hence, in accordance with guidance (ethical and legal), and based on both mine and that of my university's information governance lead's interpretation, of the relevant regulations/laws, I have regarded and communicated my study data as being of the category: 'data for limited disclosure'. I applied for and received NHS Research Ethics Committee approval (REC No: 13/WS/0243). I liaised with every Trust's research and information governance teams, not only to clarify the purpose of the project, but also to reassure them regarding the study's compliance with the principles of data protection and information governance. I clearly stated that the data they were providing was not going to be regarded as a freely usable anonymised dataset, and absolutely no attempt to de-anonymise the data would be made. As the Information Governance Review [118] was published as recently as March 2013, the recommended 'Accredited Safe Havens' for such 'data for limited disclosure' had yet to be established and accessible at the beginning of my study; and indeed, even now in 2016, my local 'Accredited Safe Haven' does not have sufficient storage or computing resources to allow resource-intensive (computing and storage) research to be carried out within it. I therefore followed the principles of the ICO's 'limited access safeguards' [116] to define the data management of my study, specifically:

- **Purpose limitation:** The purpose of the acquisition of the data was clearly defined and restricted to my study only. The protocol of my study was made available to all research sites (NHS Trusts) (Appendix 2).
- **Training of recipients' staff with access to data, especially on security and data minimisation principles:** I would be the only person with direct access to the dataset, and as such I not only undertook Good Clinical Practice certification and UCL Information Governance certification, but also liaised extensively with the Information Governance Lead for UCL on all aspects of the project. In addition, I consulted with both the IT teams at UCL to understand the IT infrastructure at UCL, as well as external entities that provide NHS storage and computing resources, to determine the appropriate way to store and access the data.

- **Personnel background checks for those granted access to the data:** As a medical doctor, I have regular Disclosure and Barring Service (DBS) (formerly known as CRB – Criminal Records Bureau) checks carried out, all of which have raised no concerns. In addition, I am bound by the General Medical Council standards of 'Good Medical Practice'.
- **Controls over the ability to bring other data into the environment, allowing the risk of re-identification by linkage or association to be managed:** A firm commitment was made to all study sites and written into all the study documents, guaranteeing that 'No attempt to link the data obtained from the various organisations with any other research database will be made'.
- **Limitation of the use of the data to a particular project:** A firm commitment was made to all study sites and written into all the study documents, guaranteeing that the data would only be used for my PhD project as defined by the protocol: 'The data will be used solely for this project'; and in addition, that 'The data will not be shared with any other investigators'.
- **Restriction on the disclosure of the data:** A firm commitment was made to all study sites and written into all the study documents, guaranteeing that 'The data will not be shared with any other investigators'.
- **Prohibition on any attempt at re-identification and measures for the destruction of any accidentally re-identified personal data:** A firm commitment was made to all study sites and written into all the study documents, guaranteeing that 'No attempt to link the data obtained from the various organisations with any other research database will be made'.
- **Arrangements for technical and organisational security:** Appropriate computational technologies would be used to store and handle the data, and all would be compliant (in terms of security and so forth) with data protection needs. Some of the specific measures stated in the various documents were: 'Dr Vishal Nangalia, the Chief Investigator, will analyse the data at University College London.', 'All the anonymised electronic data will be stored on secure, password protected and encrypted computers at UCL.', 'Access to the Dr Nangalia's office at UCL is only available via a security controlled gate and only to authorised members of staff. He will use secure, password protected and encrypted computers for the storage and analysis of the "anonymised" data'.
- **Encryption and key management to restrict access to data:** Only secure, password protected and encrypted computers at UCL would be used for data storage and analysis. The only person to have access to this system, and indeed the data from the study, would be Dr Vishal Nangalia.
- **Limiting the copying of, or the number of copies of the data:** Only local copies, on the same computer system, of the data would be made, apart from a single additional backup. The backup would also be encrypted, and only Dr Nangalia would have access to it.

In addition to the above, as part of the protocol of the study, specific measures were put in place, specifically for data collection:

- A local collaborator would be identified at each individual NHS Trust.
- The data extraction and anonymisation would be carried out by the information teams at the respective NHS Trusts, using existing hospital systems.
- It was explicitly stated that the following would be carried out with regard to specific data fields in the dataset:
    - Removal of patients' names
    - Removal of patients' addresses
    - Partial removal of postcodes. Only the first 4 characters (including spaces) would be requested, thus placing each partial postcode within a group of approximately 8,600 households. However, post collection, I made the decision to delete all postcode data completely to further the anonymisation of the dataset.
    - Removal of day of birth from date of birth. Only the month and year of birth would be requested.
- Each NHS Trust's own Research and Development, Information Governance and Caldicott Guardian teams would approve the 'anonymised' data extraction.
- No patient-identifiable data would leave the NHS Trust or be shared with the researcher.
- The data would be collected by the Chief Investigator from each individual trust, using an NHS-approved encrypted USB (Universal Serial Bus) stick 'IRONKEY'.

I also provided guidance to the local NHS Trust information teams on how the anonymisation of the data should be carried out, in addition to being available for advice on technical methods to remove identifiable data. Regarding the anonymisation of the NHS Number and Local Hospital Number (MRN-Medical Record Number), I advised the use of a one-way cryptographic hash function with the inclusion of a 'SALT'. The addition of a random SALT, which was not to be shared with me, the researcher, would guarantee the irreversibility of the encryption/anonymisation of the specific data field. According to Wikipedia: 'In cryptography, a SALT is random data that is used as an additional input to a one-way function that hashes a password or passphrase. The primary function of SALTs is to defend against dictionary attacks versus a list of password hashes and against pre-computed rainbow table attacks.' Unlike secret-key or public-key algorithms, hash functions use one-way encryption and have no key. A fixed-length hash value is computed based on the plaintext input, in such a way that it is impossible for either the length of the original plaintext or the contents to be determined. The algorithm I recommended was the SHA2_256 function, which is present in SQLServer 2012 onwards. The '256' in the SQLServer 'hashbytes' function refers to the blocksize that the SHA2_256 function employs, in this case 256-bit (32 bytes). It also utilizes 64 rounds of the algorithm to generate the final hash.

A 'dictionary attack' refers to 'a technique for defeating a cipher or authentication mechanism by trying to determine its decryption key or passphrase by trying hundreds or sometimes millions of likely possibilities, such as words in a dictionary'. A pre-computed rainbow table attack is similar to a dictionary attack, but in this case the dictionary is already associated with pre-computed hash values. This method is void if a SALT is added to the data prior to the hash function being employed. It is important to note, though, that my recommendations for anonymisation were no more than recommendations, and it was the responsibility of the local Trusts' information teams to decide the exact anonymisation method used, which I advised them not to share with me.

Finally, as 'Chief Investigator' of my study, I was the person with primary responsibility for the design, conduct and reporting of the study, despite not being an investigator at any particular site. I was responsible for:

- Developing proposals that were scientifically sound and ethical.
- Submitting the design for independent expert review.
- Submitting the study (or proposal) for independent ethical review.
- Conducting a study to the agreed protocol (or proposal), in accordance with legal requirements, guidance and accepted standards of good practice.
- Preparing and providing information for participants.
- Ensuring participants' welfare during the study.
- Arranging to make findings accessible, following expert review.
- Feeding back results of the research to participants.

The documents relating to the study included the Protocol (Appendix 2), the Research Ethics Committee Application Form, The NHS Research and Development Application Form, Site-Specific Application Forms, and Proof of Insurance Indemnity.

I believe the ethicality of the use of medical data depends on who owns that data. The ownership of a patient's individual healthcare record is currently unclear, and unfortunately, it seems that minimal steps are being taken to address this issue. There is the risk that patients, as data owners, could withdraw their consent for the use of their data for any purpose other than their direct clinical care. Even when their record is used for their direct clinical care, patients may decide to restrict the access of healthcare providers/individuals to parts of their record. The ongoing paternalistic attitude of the medical profession, the healthcare providers, and the government, as well as the introduction of legislation such as the Health and Social Care Act 2012, has, in my opinion, so far prevented a meaningful and open debate of this particular issue. Until such a time as this issue is clearly resolved, my approach, as outlined above, could serve as a model for linking datasets at source, and carrying

out research which would comply with both ethical and legal requirements. This approach is of particular value in the context of resource-intensive (computing and storage) machine-learning projects that cannot be accommodated within existing 'Accredited Safe Havens'.

2.3 Data Acquisition and Pre-Processing

2.3.1 Data Acquisition

Ethical approval was just the beginning of the long process of acquiring the data. The process for initiating my study involved

1. Protocol approval and agreement with the sponsor, in this case my university.
2. NHS REC approval
3. Clinical Research Network (CRN) application for inclusion as a national portfolio study;
4. Approaching individual NHS Trusts to encourage adoption of the study in their respective organisations (which entailed discussions and approval of the local research and development, information governance, Caldicott Guardian, and the specific hospital teams who would be carrying out the study). I also identified a local site investigator who would take responsibility for their specific organisation.
5. Liaising with the local IT teams to discuss actual data extraction, modification, anonymisation, and ultimately transfer.

Processes 1 and 2 proceeded smoothly; however, due to an error made in the application process, automatic CRN and national portfolio inclusion did not occur. This process had to be repeated, with a separate formal application to the CRN justifying why CRN inclusion was applicable. The justification for my study's inclusion was that it was directly funded by the Medical Research Council, and that it had undergone a process of extensive external peer review before funding had been awarded. My study's CRN approval arrived approximately nine months after ethical clearance was received. The specific advantages of CRN and portfolio inclusion for my study were that these offered: 1) support for the local NHS sites to assess and ultimately approve the study by providing access to CRN-funded research delivery staff (research nurses, data managers, etc.); 2) access to a central document management system for all the documents and approvals relating to the study; 3) re-imbursement to the local NHS sites for participation in the study, which thus encouraged them to participate; and 4) site selection. After being made aware of my study via their local CRN, local researchers at several sites contacted me and requested to participate in the study.

Recently (as of Spring 2016), there has been a substantial change in the way research in the NHS is approved and managed. The HRA has shifted some of the responsibilities of research and development approval away from local NHS trusts to a central body. However, when my study was initiated, a full review was required at each study site (NHS Trust), prior to initiation of the study at that site.

The issues I encountered during local approval and data collection are summarised as follows:

1. **Information governance issues:**
    a. Discussing why patient consent was not required: i.e. this would be an anonymised extract used only for a 'limited' purpose, with strict access controls (i.e. only one person, myself, would have access to this anonymised dataset).
    b. Clarifying that the project already had ethical approval.
    c. Clarifying why initial extraction of the data (but not transfer) required certain patient identifiers to be retained, so that the two different datasets could be linked together. After this stage, removal of identifiers and further anonymisation could be carried out on the linked dataset prior to its transfer to me.
    d. Providing reassurance that the data would only be used for the project specified, and not shared with any other individual or group.
    e. Providing reassurance that each NHS Trust's data would not be compared with any that of other NHS Trust; i.e. the data would only be analysed as a grouped cohort.

2. **Technical issues**
    a. Identification of the relevant information teams at each NHS hospital, these being responsible for maintaining or providing data from each of the two data repositories (CDS.APC and blood results).
    b. Identifying the individual with the authority to authorise the task of extracting the data.
    c. Understanding the variable IT and data landscapes at each NHS Trust, and identifying the optimal systems by which to extract the required data.
    d. Convincing the various information team individuals that it was possible to extract the data I was requesting from either the databases or the relevant IT system.
    e. Convincing the various information team individuals that the work of extracting the data from either the database or the IT system would not require months of work.
    f. Convincing the various information team individuals that external IT consultants did not need to be hired, at a cost of approximately £50,000, to extract, link and anonymise the data I was requesting.

g. Convincing the various information team individuals to prioritise and actually carry out the data extraction. The individuals concerned were frequently overworked, in departments which were struggling to manage their routine workload.
h. Most NHS Trusts were unable to access their datasets from as long ago as 2005 (the requested data extraction start date), and thus shorter historic time periods had to be accepted.

To resolve the above issues, 1) I learnt about the workings of the clinical IT systems, frequently from the manufacturers of such systems, and provided technical support to the relevant individuals at each NHS Trust, to help them extract/link/anonymise the data; and 2) made multiple visits to every NHS trust enrolled in the study, along with making weekly/fortnightly phone calls.

The initial timeline for this phase (data collection) of my study was set at six months. However, the entire process, up to the point of receiving an anonymised dataset, took approximately two-and-a-half years. For all the NHS Trusts, this was the first time they had been involved in a study in which such large amounts of data were extracted, linked between different hospital systems, anonymised and transferred to an external investigator. The full list of NHS Trusts that enrolled in the study and provided data is shown in Table 2.4. A number of NHS Trusts either declined to participate in the study, or enrolled and passed information-governance approval, but were then unable to provide any data, due to an inability to resolve some of the technical issues previously outlined.

Table 2.4: List of all NHS Trusts participating in the ML-EWS Study

	NHS Trust
1	Barnet and Chase Farm Hospitals NHS Trust, England
2	Barts Health NHS Trust, England
3	Betsi Cadwaladr University Health Board, Wales
4	Cardiff and Vale Health Board, Wales
5	Colchester Hospital University NHS Trust, England
6	Epsom and St Helier University Hospital NHS Trust, England
7	George Eliot Hospital NHS Trust, England
8	Homerton University Hospital NHS Trust, England
9	Imperial College Healthcare NHS Trust, England
10	Royal Cornwall Hospitals NHS Trust, England
11	Royal Free NHS Trust, England

12	Royal Marsden Hospital NHS Trust, England
13	University College London NHS Trust, England
14	Whittington Hospital NHS Trust, England

2.3.1 Pre-Processing of the Collected Data

The challenges encountered with the data collection continued during the transfer of the anonymised datasets to me. In fact, the most technically challenging aspect of data acquisition was the pre-processing of the transferred data. Pre-processing involves 1) *ingestion* of the data to a database; 2) *transformation*: the conversion of the data to a standard format (for example male = '1' as opposed to male = 'M'); 3) *cleaning*: identification of incomplete, inaccurate, and irrelevant data and conversion or removal of such data; 4) *feature engineering*: creation of new variables (features) from the raw data; and 5) *variable selection*: selecting the specific variables needed to build a model. The feature engineering and variable selection for each model will be described in the relevant chapters for each model (Chapters 3, 4 and 5). Here, I describe the challenges and steps involved in performing the ingestion, transformation and cleaning aspects of pre-processing.

2.3.1.1 Data Ingestion

The data from the various NHS Trusts were inserted into a SQL Server 2014 database, using the SQL Server 'import wizard'. The database was initially housed on a secure and dedicated workstation, and then transferred to a secure dedicated server; both computers were only used for this project, and had all appropriate security measures deployed (as described in my ethics submission and research protocol). The data themselves were occasionally provided by the NHS Trusts as a database backup, but most commonly as a number of simple text files ('flat files'). The 'flat files' consisted of plain text with a delimiter between each variable, such as: '-' (dash), '|' (pipe), or '||' (double pipe). Occasionally the variables within the files were either placed in fixed-width slots (i.e. variable one occupied the space for characters 1–10; variable two, characters 11–20, etc.), or were separated by fixed-width empty spaces ('tabs'). The initial import of the data was performed using a 'catch-all' philosophy, where all the data were imported into a text format that allowed for various types of characters ('varchar'). Thus dates, text and numbers were all regarded as strings of characters. Despite this catch-all approach with no initial quality control, there were still issues with the data ingestion. Most commonly, 1) the delimiters within the data changed part-way through a file, or 2) there were additions of comments within the file that were not in the same format as the rest of the data. To correct these and other errors, I utilised special software tools such as 'R' and 'Notepad++' to manually inspect the data files and fix/remove corrupted segments.

The above procedure was carried out for every data file received from every NHS Trust. This was a long and laborious process, as not only did the data files differ between NHS Trusts, but when a Trust consisted of multiple hospitals, the data was commonly provided on a per hospital basis, each with its own data structures. In addition, as data from multiple years was requested (2005 onwards), the majority of NHS Trusts decided to extract their data on a yearly basis, and rather than consolidating it all into one dataset, they provided multiple files for a specific dataset, sometimes covering only one year at a time. The data structure within these files also changed every few years, for every NHS trust. A factor that further complicated this entire process was that there was frequently no documentation available for any of this historic data. Thus, data transformation had to be carried out by visual inspection of the data contents.

2.3.1.2 Data Transformation

Despite requesting data that is shared nationally using two national standards (CDS.APC and NLMC), and requesting the data in these national formats, none of the NHS Trusts who participated in my study provided me with this data in the respective national formats. In fact, every NHS Trust provided the data using its own customised format. These customised formats used different names not only for the same variables (for example, 'sex' could be 'persongender' or 'persongendercurrent', or 'gender'), but also for the data items within the variables (for example: male could be: '1', 'M' or 'male'; hemoglobin could be: 'Hb', 'hb', 'haemoglobin', 'HB', 'serum_HB', 'venous_HB', etc). Surprising as this may seem, the creation and maintenance of local customised data models (as opposed to the adoption of a national standard), seems to be the path of least resistance within NHS Trusts. Consider the CDS.APC dataset: although it is a national requirement to submit these data monthly to HSCIC in England (NWIS in Wales), not a single NHS Trust in my cohort submitted this data directly to its relevant national body. Instead, each NHS Trust submitted its data to a commercial data processing company (most commonly Indigo Ltd). Indigo would then 1) perform quality checks on the data, 2) transform it to the HSCIC's required XML (eXtensible Mark-up Language) format, and 3) submit it on behalf of the NHS Trust to the HSCIC. In the majority of cases, the NHS Trust would not even hold a copy of this final XML submission. Thus, the national standard for the data generates work for commercial data analytics companies, rather than encouraging every NHS organisation to standardise data capture and processing, and to create workflows that align with national requirements.

There are, I believe, multiple reasons for this:

1. *Clinical IT systems can be highly specialised and proprietary*. The majority of healthcare IT systems in NHS Trusts today are relatively old. They are based on state-of-the-art technology from at least a decade or two ago. The majority have also not been upgraded to the latest available system release, due to the cost of modifying the newer systems to communicate with

other legacy IT systems. Thus, these systems have limited functionality, and even less ability to be modified for new data capture and extraction requirements.

2. *Limited access to the software that powers the clinical IT systems.* The commercial agreements that NHS Trusts have signed with their IT providers frequently do not allow the NHS Trust (the customer) to directly access the raw clinical or administrative data stored within these systems. The only access that the NHS trusts have are via customised 'portals' that allow only a very narrow range of queries to be carried out on the underlying databases that store the data. The queries themselves are not based on a standard database querying language, such as SQL (standard query language), but have customised and proprietary syntax, adding an additional barrier to data access and extraction.
3. *The IT expertise within an NHS Trust is limited, as the IT departments are generally underfunded, with a high staff turnover rate*. The salaries for information analysts within the NHS are lower than equivalent salaries in the commercial sector. In addition, it is always more politically expedient to spend money on hiring more managers, nurses, or doctors, than on an IT system or staff. During the course of my project, I frequently found that my IT liaison at an NHS Trust had moved to another organisation.
4. Finally, *procedures to procure and maintain the myriad clinical information systems have traditionally been conducted haphazardly (in my opinion) within the NHS*. As a result, although the money spent on these systems is substantial, the system obtained is either not fit for purpose, or the contracts signed do not allow any flexibility without massive additional costs. Therefore, when clinical requirements change, the IT systems are not appropriately modified; instead, inappropriate (and cheap) modifications are employed to enable increased functionality. This creates an IT environment within an NHS Trust that includes multiple systems, all diverging from standard manufacturer specification and industry best practice guidance. Over time, this complexity increases, and it becomes even more difficult, and costly, to change and upgrade the systems.

To return to my data: post data ingestion, I performed a number of steps, both to check the data quality and to transform the data to my standard format. I designed a series of queries ('scripts') to perform these tasks. These tasks were:

1. *To confirm the anonymisation of the identifiable fields*. If anonymisation was in doubt, another round of 'hashing' was carried out on these fields, across the linked data.
2. *To confirm the deletion of non-requested fields (especially identifiable fields)*. If the presence of any identifiers was noticed, these were immediately deleted (in both the original raw file and the database) and discussed with the providing NHS Trust. No significant identifiable data was received in this way.

3. *To confirm the linkages of the CDS.APC data to the blood results data*. Occasionally, no linkages across these two were found. After liaising with the NHS Trusts concerned, the most common cause was inappropriate anonymisation: i.e. different anonymisation algorithms were employed to anonymise or 'hash' the same identifiers in the two datasets (CDS.APC and blood results), resulting in different hashed values which did not match. Once identified, the NHS Trusts appropriately anonymised the datasets and provided them to me.
4. *A database 'View' (method of transforming data and visualising the new format), was created for each NHS Trust for the CDS.APC and blood results data*. These 'Views' transformed both the variable names from the original 'flat files', as well as the data types (string, numeric, datetime etc.) into my standard format. Thus, two Views (CDS.APC and blood results) were created for each NHS Trust. This step required multiple iterations, as the transformation frequently failed due to corrupted data, non-standard characters, or the assignation of data elements to the wrong category (for example, the date of admission being placed in the 'method of admission' field, with the date being stored as a custom data type (datetime), while method of admission data were numerical).
5. Once these Views were created, the identifiers across the CDS.APC and blood results were rehashed and then replaced sequentially by an index number, beginning at '1'. Finally, a coded prefix unique to that NHS trust was added to each anonymised identifier. Thus, identifier '1' became 'abc-1'. This was to allow merging of the data across NHS Trusts and prevent duplicate hashed identifiers from occurring; i.e. if two separate NHS Trusts used different SALTs, it is possible that different NHS numbers could have been converted to the same anonymised 'hash'.
6. Once these Views and new identifiers were finalised for each NHS Trust, the data was transferred (via the Views) to a new database, where all the data (across NHS Trusts) were merged together, into two separate tables: 1) CDS.APC, and 2) ORU.Pathology (blood results).
7. New tables were now created that extracted key variables from the CDS.APC table into additional linked (relational) tables. These were: ICD10 (which consisted of the hashed patient identifier, Spell date of admission and Spell date of discharge, and a single ICD10 code. Thus, one CDS.APC row could result in fifty rows in the ICD10 table, as each ICD10 code would have its own row. A similar table was created for OPCS-4 codes. This process of converting data into singular rows is referred to as conversion from 'wide' to 'tall' format. Additional tables were created that held only patient demographic data. Rather than permanently link each row in each table to rows in other tables using unique row identifiers (a relational structure), I chose to include key data in each of the tables (a flat structure). The advantage of this 'flat' approach is that each table could be independently queried for data; however, the disadvantage is that some data would be duplicated.

2.3.1.3 Data Cleaning

Once all the data was inserted into the final database 'RHub', inspection and cleaning of the contents of the data variables could begin. The size of the dataset at this time was immense. The ORU.Pathology (blood results) table alone consisted of over 1.1 billion rows, and the CDS.APC table had over 16 million rows. However, this was the 'uncleaned' data, and was currently unfit for use in my analysis. The steps I performed to 'clean' this data were:

1. The first task was to check the quality of the linked identifiers. Frequently, when a patient arrives in hospital, although they have been admitted/registered on the hospital systems before, a new hospital number (MRN) is assigned to them. Thus, a patient may have more than one MRN. The resultant problem is that all the patient's historic data becomes inaccessible if the new MRN is used to search for such data (for example, blood results). To remedy this situation in my dataset, I searched for multiple MRNs that were linked to the same NHS Number (I had requested the hashed MRN and NHS numbers for each patient). For each of these, I then checked whether the date of birth (month and year of birth) and sex matched; if so, all these identifiers were consolidated into one, and the old identifiers deleted. This was checked across both the CDS.APC and ORU.Pathology tables. If there were clashes, such as if the linked MRNs and NHS numbers had different birth dates and sexes associated with them, then these data were deleted. Eventually, a final identifier list was created and updated across all the tables.
2. I then, for each of the variables listed in Tables 2.1 and 2.2, sequentially checked how much of the data within each variable conformed to the expected attributes. All data that did not perfectly match the expected attributes were either converted to match these where possible (for example, ethnic category 'British' changed to 'A'), or deleted. When data was missing for a variable, this was checked against the ORU.Pathology and demographics tables, and if present were used.
3. Admission and discharge dates prior to 1 January 2004 and 1 October 2016 were also deleted, as were rows of data where either of these values was missing, or where the discharge date was before the admission date. In addition, all rows of data where the discharge date was >183 days (more than six months) from the admission date were also deleted. This was because I found lengths of stays of five years or more for hundreds of patients, many with the same discharge date (implying that a default discharge date had been inserted when none was present; this is poor but common practice in databases that have been designed not to accept empty values). I acknowledge that this approach does exclude genuine patients with lengths of stays of greater than six months, but these patients form a tiny proportion of the total number of admissions.
4. Date of birth parameters were also further pruned. The remit of the study was to investigate only adult patients; however, as the dataset spans over a decade, patients who are now adults may have been admitted when they were younger than eighteen years old. Although these admissions

would not be investigated directly, the past medical history of adult patients is crucially important in understanding a patient's condition and their likelihood of deterioration. Thus, to comply with the remit of my project, but also to retain past medical history data, I deleted the data for all patients born on or after September 1995; eighteen years before the start of my project. Thus, only adult patients, according to the approved data ethics for my project, were included in the investigation. However, I also took the additional step of including only patients who were eighteen years or over on the date of admission to hospital in my subsequent analyses (Chapters 3 to 5). Most NHS Trusts were not able to perform this filtering of the data prior to transfer.

5. For each blood result test, multiple dates and times were available: i.e. for sample collected, sample received in laboratory, and result validated. However, all three categories were sparsely populated, with the most common variable being 'validated datetime'. Therefore, I used 'validated datetime' for my analyses. Out-of-range values were also deleted.
6. For the 'TestNames' variable, there were thousands of unique blood result tests, and sorting and categorising all of these would not have been possible during the course of my investigation. Therefore, I focussed on the most common blood tests performed in hospital, specifically full blood count (FBC), urea and electrolytes (U's and E's), and albumin. Each of these tests was represented under multiple names. Customised searches were carried out on common variations of each of the possible names, and when it was confirmed that they did indeed represent the test in question, all the differing names were changed to the database standard. All other tests were left untouched for 'cleaning' at a later date. I also received result data from samples other than blood; however, I analysed only venous or arterial blood results.
7. One would expect the units in which common blood tests are reported to be the same; however, this was not the case. For each blood test, I confirmed the range of its actual results along with the pairing of its units, and if they conformed to reported quality standards, approved them. However, when the ranges and units did not match, further investigation was carried out, and where possible the result values were converted to the appropriate units.
8. Finally, I also carried out a range of additional quality checks for the distribution of the data.

Apart from various specific data cleaning tasks described above, I also deleted all blood results for which where there were no matching patient identifiers in the CDS.APC dataset. These blood tests were probably performed by the hospital laboratory on patients who had never been admitted to hospital, and were thus not relevant for my analyses. Duplicate blood results were also deleted.

The size of the dataset following these extensive changes was, for CDS.APC, ~8 million rows; and for ORU.Pathology, ~500 million rows. Further details of the data are provided in each of the analysis chapters (Chapters 3 to 5).

2.4 Lessons learned

I believe the reason for my success for carrying out such a large-scale data project, hinge on a combination of my clinical and technical expertise, and following the principles of these key lessons.

A. The ethical and legal requirements are not barriers to make life difficult for researchers, but necessary safeguards to maintain trust. These ethical/legal requirements are complicated, and sometimes conflicting, and thus frequently require discussion and clarification with individuals responsible for approving the data access

B. Technological advancements, especially in the field of encryption and anonymisation are taking place faster than regulations and guidance can keep up. It is essential to follow the principles of the ethical/legal framework rather than simply the guidance text. Thus, when appropriate additional steps would enhance/maintain these principles, then further data anonymisation, deletion, and limitations in access, need to be implemented.

C. Individuals working in a healthcare environment, especially those involved with data and IT are busy and have limited time to both engage with researchers and to extract the data. They have skills specifically suited to their job of maintaining the IT systems for clinical use, and these skills do not always enable large scale or live data access. However, they are extremely motivated. Thus to succeed in a project that requires medical data access, three key ingredients are required 1) patience, 2) technical support, and 3) management support to prioritise the data access workload.

D. Healthcare IT systems are generally proprietary systems, which are based both on older technology frameworks and have limited access to the end user (customer). This makes accessing the data that resides in them difficult. Engagement with the software provider is frequently necessary to understand these proprietary systems, and to enable technical support to the healthcare IT teams. Data access from certain IT systems will be unfeasible either due to technical issues, or because the software provider will demand a large payment to enable such data access.

E. Healthcare data resides in multiple IT systems, and is frequently collected in duplicate from multiple sources. In cases where data access from a certain healthcare IT system is unfeasible, this data may be accessible from another data repository within the hospital.

F. Healthcare data does not adhere to published international, national or even local standards. Demanding data in a format (schema and quality) suitable for the researcher is quick route to alienating IT staff. Raw data is better than heavily modified data. A significant amount of time needs to be spent in examining and understanding this data. Both clinical and technical expertise is essential in enabling data transformation to a canonical standard.

Chapter 3 Dehydration: simple dual parameter to ML

One of the simplest ways two blood result variables have been used in conjunction has been the urea to creatinine ratio (Ur:Cr) as a surrogate for dehydration. In this chapter, I will explore the extent to which this ratio is associated to poor outcome. I will expand on the historical snapshot interpretation of the ratio and its association with outcome by investigating 1) its dynamic change, 2) the effect the addition of simple demographic data modifies outcome prediction and 3) the application of a simple ML technique.

Made up of two of the blood tests most commonly performed in hospital are for urea (Ur) and creatinine (Cr). Individual results higher than the normal range, as well as an elevated Ur:Cr ($\geq$80 mmol/L:mmol/L), are known to be associated with poor outcome, (Chapter 1, Section 1.4.2). More specifically, Urea and creatinine results are also used to define serious medical problems prevalent amongst hospitalised patients, as follows: 1) a significant rise in creatinine is used to define AKI; and 2) a Ur:Cr $\geq$ 80 has historically been used to define dehydration. The availability of these two test results amongst data for hospitalised patients makes them obvious candidates for initial use in developing a risk stratification model for predicting AEs in hospitalised patients.

To explore the utility of Ur and Cr test results in predicting in-hospital decline, I performed my initial analyses on the first three NHS Trusts from which I obtained data. I focused on emergency admissions to hospital, as these have a high risk of in-hospital mortality.[124]

3.1 Objectives

To investigate outcome (mortality and length of stay) associated with known models of disease. Specifically, for dehydration, to explore

I. The significance of dehydration as measured by Ur:Cr on outcome
II. The significance of dynamic changes in dehydration (Ur:Cr) status, combined with AKI, on outcome
III. The application of an ML technique that incorporates urea and creatinine results with simple demographic data to identify those at risk of poor outcome

3.2 Methods

3.2.1 Data

Data used included in-hospital blood results and administrative data (demographics and ICD10 codes) (Chapter 2, Section 2.1), from adult (aged 18 years or over) emergency admissions to three UK NHS Hospital Trusts over 156 months (from early 2007 to mid-2013, dates varying slightly between hospitals). All three of these trusts were rated as 'good' or 'excellent' by the Care Quality Commission rating in 2009,[125] and had a Standardised Mortality Ratio below the national average as of January 2014[126]. Only data for patients with complete administrative records and ≥2 paired urea and creatinine blood test results were used for analysis. Excluded were maternity patients, those who self-discharged, and patients for whom a blood test was not performed within 24 hours of hospital admission (in whom a change from admission could not be assessed). Patients with ICD10 codes associated with upper gastrointestinal bleeding were also excluded from the analysis, due to the well-established correlation between changes in the Ur:Cr ratio and this pathology (Chapter 1, Section 1.4.2), thus making the assessment of AKI or dehydration unreliable.

The variables used for analysis were: 1) hospital administrative data (age, length of stay (LoS), diagnosis fields (ICD-10 International Classification of Diseases, 10th revision), method of admission and discharge); and 2) blood result data relating to urea (mmol/L) and creatinine (µmol/L). Data available to a clinician at the point of assessing a patient informed variable selection.

3.2.2 Analysis

To determine the association of admission Ur:Cr (and its subsequent change) with AKI, LoS and mortality, I performed two broad analyses:

1. Binary analysis: Proportions of patients with an admission Ur:Cr <80 (normal, N) or ≥80 (high, H), and whose admission status was either preserved (normal to normal: N->N, high to high: H->H) or changed (N->H, H->N) at any time during their hospital stay, were calculated. Fluctuating Ur:Cr was not analysed separately. Development of AKI was defined using the 'Kidney Disease: Improving Global Outcomes' (KDIGO) criteria [83] (Table 1.1). Each patient's admission creatinine defined their baseline.[33] Association between Ur:Cr (or its change) with outcome was sought in predefined age groups (<65, ≥65 years; traditionally the retirement age). Together, AKI and Ur:Cr yielded 16 sub-groups (Table 3.3 and 3.4) where in-hospital mortality (percentage) and, for survivors, median LoS, were calculated.
2. Continuous analysis: Ur:Cr was plotted against in-hospital mortality, for each set of a patient's blood tests (admission to last blood test before discharge or death). The percentage change in

Ur:Cr from admission to last test was calculated. Heatmaps of age vs Ur:Cr vs mortality were plotted for both the admission and the last blood tests, to visualise the influence of age. The last set of blood results were used ($Ur:Cr_{last}$) as these could be considered to be either the optimum, when a patient was discharged; or the worst, just prior to death. To highlight the complex relationship between all the available variables (Table 3.1), and their change (upto the second blood test) and in-hospital mortality, I built a RF ML model (ML-Dehydration). The ML-Dehydration model was built using 5-fold cross-validation. The hyper-parameters of the model were determined via a grid search. This analytical framework was applied to 70% of patients (training dataset). The final model was tested on the remaining 30% (testing dataset). Area under the receiver operator curve (AUROC) was calculated for the whole model. From the trained model, a threshold was set to classify patients as either alive or dead on discharge. I defined two classifiers (ML20 and ML33), based on two different thresholds selected to produce Positive Predictive Values of 1:5 and 1:3 respectively for the training dataset.

Table 3.1: Predictor Variables for the Machine Learning Model (ML-Dehydration)

Variables	**Type of variable**
Age	Numerical
Sex	Categorical
Admission Day (Mon-Sun)	Categorical
Admission Month (Jan-Dec)	Categorical
Admission method subset[127]	Categorical
1st Urea	Numerical
1st Creatinine	Numerical
2nd Urea	Numerical
2nd Creatinine	Numerical
1st Ur:Cr	Numerical
2nd Ur:Cr	Numerical
Difference of 1st Urea and 2nd Urea	Numerical
Difference of 1st Creatinine and 2nd Creatinine	Numerical
2nd Urea / 1st Urea	Numerical
2nd Creatinine / 1st Creatinine	Numerical
Difference of 1st Ur:Cr and 2nd Ur:Cr	Numerical
1st Ur:Cr / 2nd Ur:Cr	Numerical

3.2.2.1 Rationale for ML analysis

Traditional regression analyses were not used, due to the following issues of multicollinearity [128] (which refers to direct correlation between dependent variables) that existed within the dataset:

1. The Ur:Cr ratio is directly and inversely proportional to urea and creatinine respectively.
2. AKI is defined by a rise in creatinine [83] and is thus correlated with Cr.
3. Creatinine levels are linked to skeletal muscle mass, [98] which is non-linearly inversely related to age. [97]
4. Both urea and creatinine values may rise in a related (albeit non-linear fashion) with specific disease processes, such as renal injury.

A RF model allows for such collinearity, and is a powerful analytic technique for classification analyses.

3.2.3 Statistical Analysis

Fisher's Exact Method and the Mann Whitney U-Test were used to establish the significance of the differences between each age sub-group (<65, ≥65 y) and their respective group baselines, as well as the various interactions of admission Ur:Cr and its subsequent change during hospitalisation for the relative risks. Data are described as a percentage of the total (actual numbers) or the median (interquartile range). Jeffrey's method[129] was used to calculate confidence intervals. Sensitivity, specificity, PPV and AUROC were calculated to determine the performance of the RF predictive model.

The software used to perform the analysis comprised MATLAB R2014b (Mathworks, MA, USA), R (R Foundation, Austria), and h2o (h2o.ai, Mountain View, CA, USA).

3.3 Results

3.3.1 Basic characteristics

Overall, 79,949 admissions met the analysis inclusion criteria (Figure 3.1; total number of tests analysed: 402,733). Table 3.2 shows the baseline characteristics, and Figure 3.2 shows the histogram of age at admission.

3.3.2 Prevalence of raised Ur:Cr ratio and AKI

A high Ur:Cr was seen on admission in 45.5% (36,339) patients, of whom 83.3% (30,268) also had a high Ur:Cr at some other point during their hospital stay. Of the 54.5% (43,610) admitted with a normal Ur:Cr, this rose to ≥80 in 28.3% (12,337). The frequency distribution of admission Ur:Cr is shown in Figure 3.3. Overall, 10.7% of patients (8,558) developed AKI in hospital.

3.3.3 Length of stay of survivors (Table 3.3)

The LoS of those surviving to discharge was lowest (3.5 days (2–6)) in the <65y group with a sustained normal Ur:Cr from admission (N -> N), and highest (19 days, (10–33.5)) in the ≥65y group whose normal admission Ur:Cr subsequently became elevated and who also developed AKI (N->H + AKI). The differences in LoS for each subgroup, when compared to their respective group baselines (<65, ≥65 y), were all statistically significant ($p<1x10^{-5}$). An exception to this tendency was found for patients who were admitted with an elevated Ur:Cr that was subsequently corrected, but who yet developed AKI (H->N + AKI: p=0.33 and 0.06, zval= -0.98 and 1.85, ranksum= 4.9x108 and 9.5x108, for <65 and ≥65 respectively). The reason for this exception is probably due to the relatively small number of patients (0.6% of the total) along with their wide range of LoS (2.8-14 days).

Figure 3.1: Consort Diagram

Figure 3.2: Age at Admission

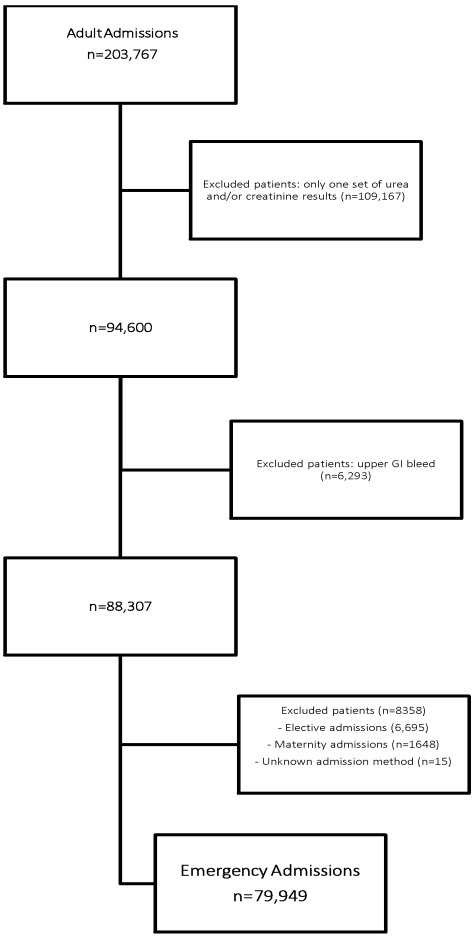
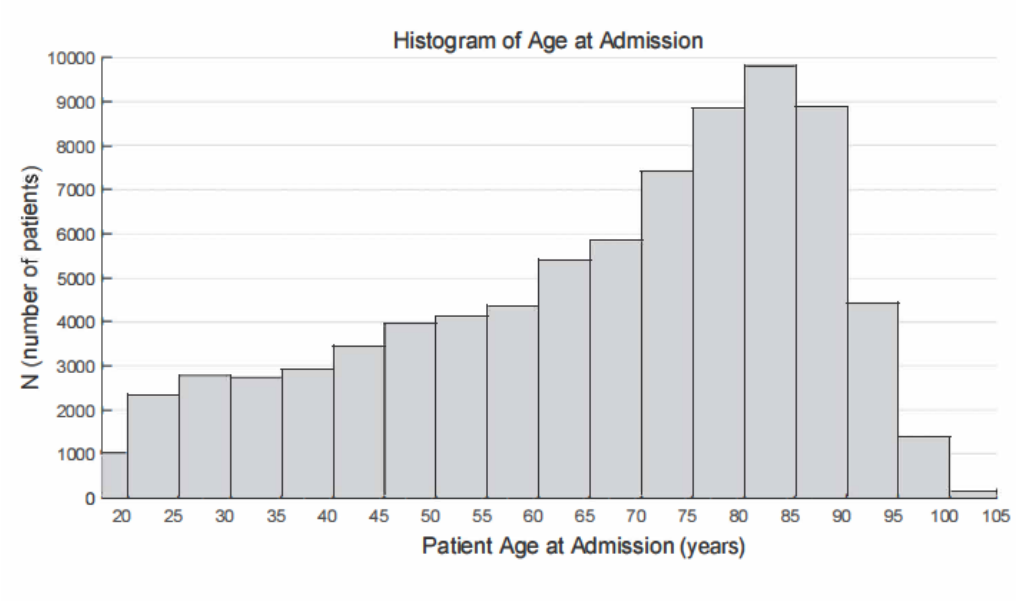
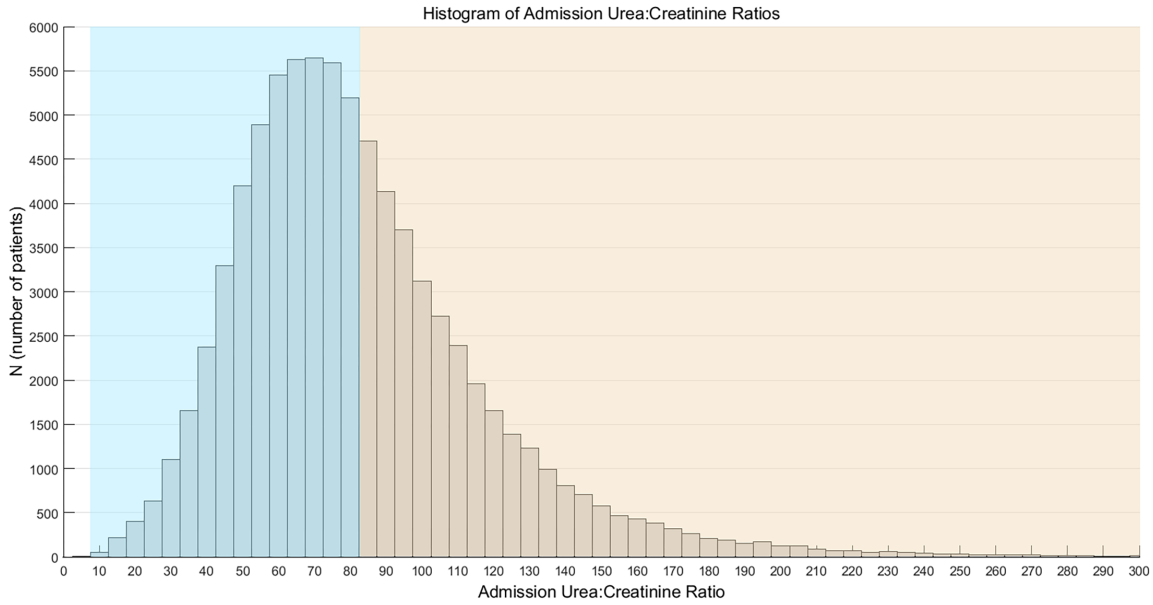


Table 3.2: Baseline Characteristics	
	Emergency admissions (n=79,949)
Age (Median (IQR (Inter Quartile Range) years	71 (51–83)
Male (%)	38,168 (47.7%)
Length of Stay (Median (IQR)), days	6.8 (3–13.2)
In-patient Mortality % (number)	6.4 % (5,080)



Histogram of age (years) of patients on admissions

Figure 3.3



Blue: Ur:Cr < 80 ; Orange: Ur:Cr ≥ 80

Table 3.3: Length of Stay of Survivors: Emergency admissions

Patient age group n=total number	Admission Blood Test	-> Subsequent Blood Tests	
Length of stay median days (interquartile range)	**Ur:Cr status at admission median days (interquartile range)**	**Ur:Cr and AKI status (percentage of baseline)**	**Length of Stay median days (interquartile range)**
<65 y n = 31,451 4 days (2–8)	High = H 5 (3–11)	H -> H (16.4%)	6.2* (3–13)
		H -> H + AKI (1.1%)	16.7* (7–31.1)
		H -> N (7.8%)	4* (2–6)
		H -> N + AKI (0.2%)	5[ns1] (2.8–10)
	Normal = N 4 (2–7.8)	N -> H (12.1%)	8* (4.2– 6)
		N -> H + AKI (1.1%)	18.8* (9.5–36)
		N -> N (56.3%)	3.5* (2–6)
		N -> N + AKI (4.8%)	6.1* (3.2–12)
≥65 y n = 43,418 8 days (4–16)	High = H 9 (5–18)	H -> H (45.2%)	10* (5–18)
		H -> H + AKI (4.1%)	18* (9–32.4)
		H -> N (7.5%)	6* (3–9)
		H -> N + AKI (0.4%)	7[ns2] (4–14)
	Normal = N 7 (4–14)	N -> H (14.8%)	11* (6–20)
		N -> H + AKI (1.8%)	19* (10–33.5)
		N -> N (22.4%)	5* (3–9)
		N -> N + AKI (3.9%)	7.9* (4–14)

** = $p<1x10^{-5}$, zval range:-35.26 to 43.4, ranksum consistently > $1x10^{8}$, ns1 = not significant, p=0.33, zval=-0.98, ranksum= 4.9x108, ns2 = not significant, p=0.06, zval=1.85, ranksum= 9.5x108*

3.3.4 In-hospital mortality

In total, 6.4% (5,080) of the 79,949 patients died in hospital. In the binary analysis (Table 3.4), mortality was strongly associated with age, Ur:Cr ratio and development of AKI. There was an 87-fold mortality difference between patients who came into hospital with normal Ur:Cr, which remained normal until discharge, and those patients who arrived in hospital with a high Ur:Cr, which remained high, and who also developed AKI. AKI patients represented 36.7% (1,865) of all deaths, and 80% of these patients also had a high Ur:Cr at some point during their hospital stay. Compared to the theoretical optimal trajectory of a hospitalised patient, where the Ur:Cr ratio is reduced below 80 (regardless of admission value) and they do not develop AKI, the relative risk of dying for all other groups of patients ranged from 6.4 to 47.9 (all $p<1x10^{-16}$, Table 3.5).

Table 3.4: Mortality: Emergency admissions				
Total no of admissions (baseline)	**Admission Blood Test**	**-> Subsequent Blood Tests**		
Mortality % total number (died)	**Ur:Cr status at admission (percentage of baseline) Total No. (no. who died) = mortality%**	**Ur:Cr and AKI status (percentage of baseline)**	**Mortality % total no. (no. who died)**	
<65 y 32,183 (732) = 2.3 %	High = H (26%) 8,394 (353) = 4.2 %	H -> H (16.7%)	5,377 (207) =	3.8 % *
		H -> H + AKI (1.4%)	465 (116) =	24.9 % *
		H -> N (7.7%)	2,476 (15) =	0.6 % *
		H -> N + AKI (0.2%)	76 (15) =	19.7 % *
	Normal = N (74%) 23,789 (379) = 1.6 %	N -> H (12.3%)	3,946 (129) =	3.3 % *
		N -> H + AKI (1.4%)	438 (79) =	18 % *
		N -> N (55.3%)	17,787 (67) =	0.4 % *
		N -> N + AKI (5%)	1,618 (104) =	6.4 % *
≥65 y 47,766 (4,348) = 9.1 %	High = H (58.5%) 27,945 (3,138) = 11.2 %	H -> H (45.4%)	21,701 (2,073) =	9.6 %^{ns}
		H -> H + AKI (5.7%)	2,725 (950) =	34.9 %**
		H -> N (6.9%)	3,290 (51) =	1.6 %**
		H -> N + AKI (0.5%)	229 (64) =	27.9 %**
	Normal = N (41.5%) 19,821 (1,210) = 6.1 %	N -> H (14.5%)	6,926 (505) =	7.3 %**
		N -> H + AKI (2.2%)	1,027 (267) =	26 %**
		N -> N (20.7%)	9,888 (168) =	1.7 %**
		N -> N + AKI (4.1%)	1,980 (270) =	13.6 %**
** = p<0.001, baseline: <65 y, 2.3%; ** = p<1x10^{-6}, baseline: ≥65 y, 9.1%; ns = not significant, p=0.0583*				

The continuous nature of the relationship between mortality and both admission Ur:Cr and last Ur:Cr is highlighted by Figure 3.4, though a more complex relationship is seen when mortality is plotted against combinations of admission Ur:Cr ratio and its in-hospital change (Figure 3.5).

Decreases of 25% or increases of 50% in Ur:Cr during hospitalisation have minimal effect on low mortality rates (<3%) when Ur:Cr is low to begin with (Figure 3.5: A, B, C). Elevated Ur:Cr on admission, which remains uncorrected (Figure 3.5: D (180 to 180)) or worsens (Figure 3.5: E (180 to 315)) during hospitalisation, is associated with in-patient death rates ranging from 28% (95% Confidence Interval (CI) 22.8–33.1) to 89% (CI: 56.5–98) respectively. However, large (50%) reductions in Ur:Cr, in patients

in whom Ur:Cr is initially similarly elevated, are associated with a reduction in mortality to just 10% (CI: 8.2–13.9, Figure 3.5: F (180 to 90)) and 12.5% (CI :10–15.4, Figure 3.5: G (210 to 105)) respectively.

Table 3.5: Relative Risk of Death for unresolved or worsening Ur:Cr and AKI

Age Group	Sub-group Baseline	Sub-group	Relative Risk of Death
< 65y	N->N	N->N + AKI	17.1*
		N->H	8.7*
		N->H + AKI	47.9*
	H->N	H->N + AKI	32.6*
		H->H	6.4*
		H->H + AKI	41.2*
≥ 65y	N->N	N->N + AKI	8.0*
		N->H	4.3*
		N->H + AKI	15.3*
	H->N	H->N + AKI	18*
		H->H	6.2*
		H->H + AKI	22.5*

* = P<1x10^{-16} ; N=Ur:Cr <80; H=Ur:Cr ≥80; AKI=In-hospital development of Acute Kidney Injury

Figure 3.4: Urea:Creatinine (Admission and Last) vs Mortality

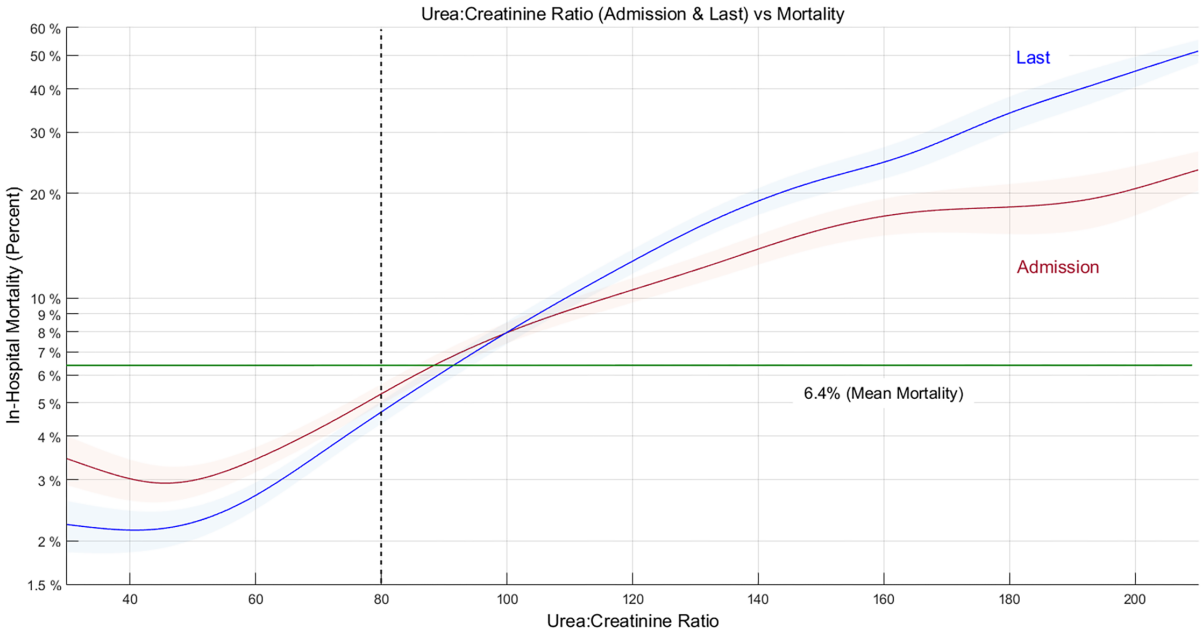
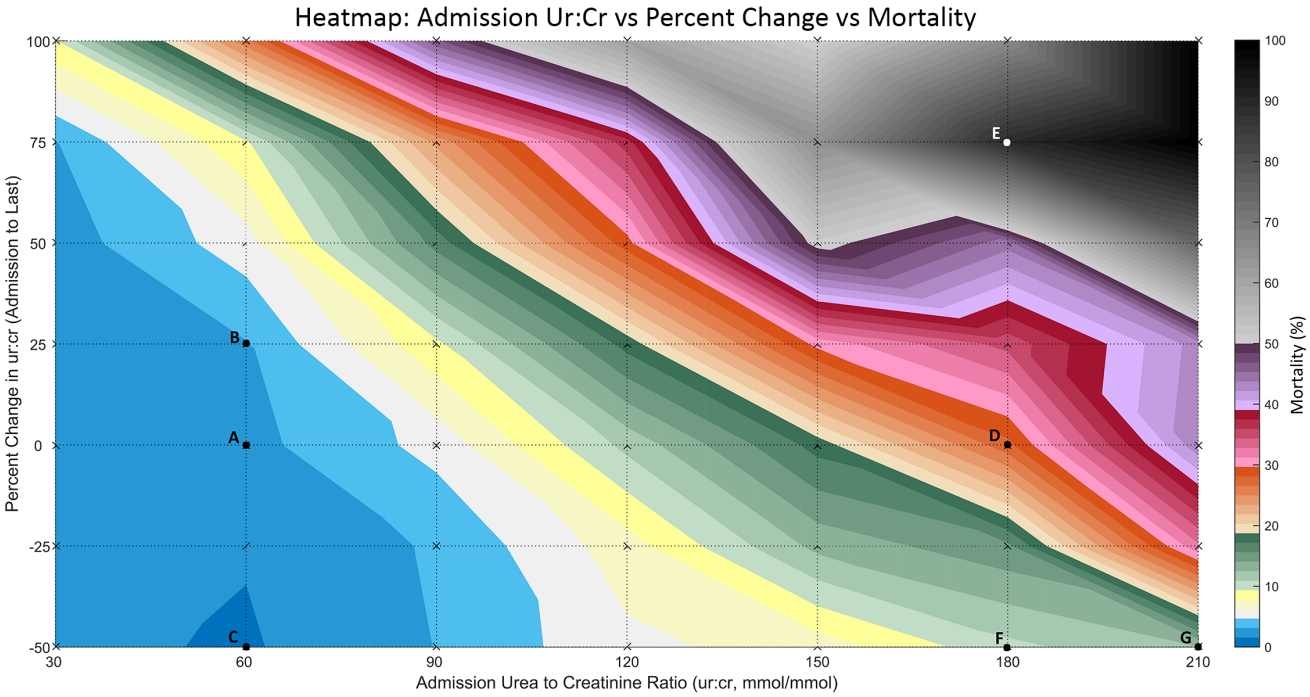
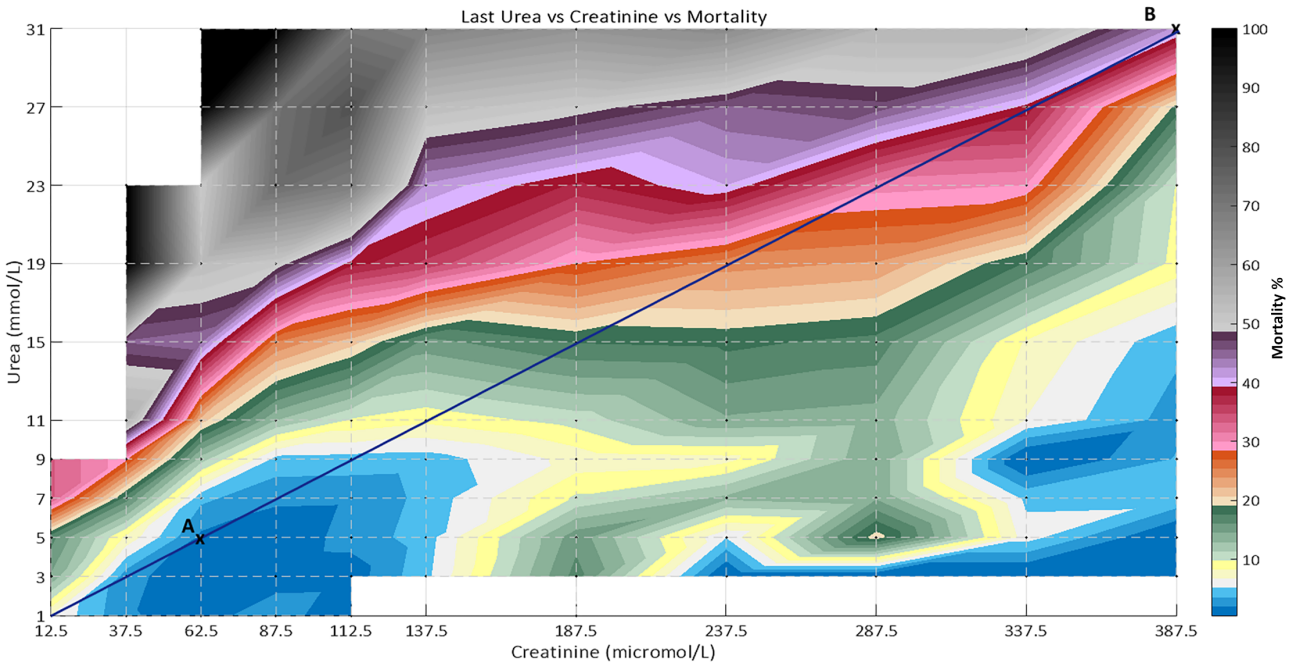


Figure 3.5



A: Admission Ur:Cr 60 -> Last Ur:Cr 60, no change; B: Admission Ur:Cr 60 -> Last Ur:Cr 75, 25% increase; C: Admission Ur:Cr 60 -> Last Ur:Cr 30, 50% decrease; D: Admission Ur:Cr 180 -> Last Ur:Cr 180, no change; E: Admission Ur:Cr 180 -> Last Ur:Cr 315, 75% increase; F: Admission Ur:Cr 180 -> Last Ur:Cr 90, 50% decrease; G: Admission Ur:Cr 210 -> Last Ur:Cr 105, 50% decrease.

Figure 3.6



Blue line = Ur:Cr ratio of 80. A=1.7% mortality, B=41.9% mortality

The relative contributions of urea and creatinine to their ratio, and their subsequent association with mortality, are also extremely complex, as demonstrated in a heatmap of the relationship of urea and creatinine (from the patient's last blood test before discharge or death) to mortality (Figure 3.6). With a Ur:Cr of 80, the mortality varies from 1.7% (CI: 1.69–1.70, Figure 3.6: A) to 41.9% (CI: 41.76–42.02, Figure 3.6: B). Patient age also non-linearly interacts with Ur:Cr to influence mortality (Figures 3.7 and 3.8).

Figure 3.7

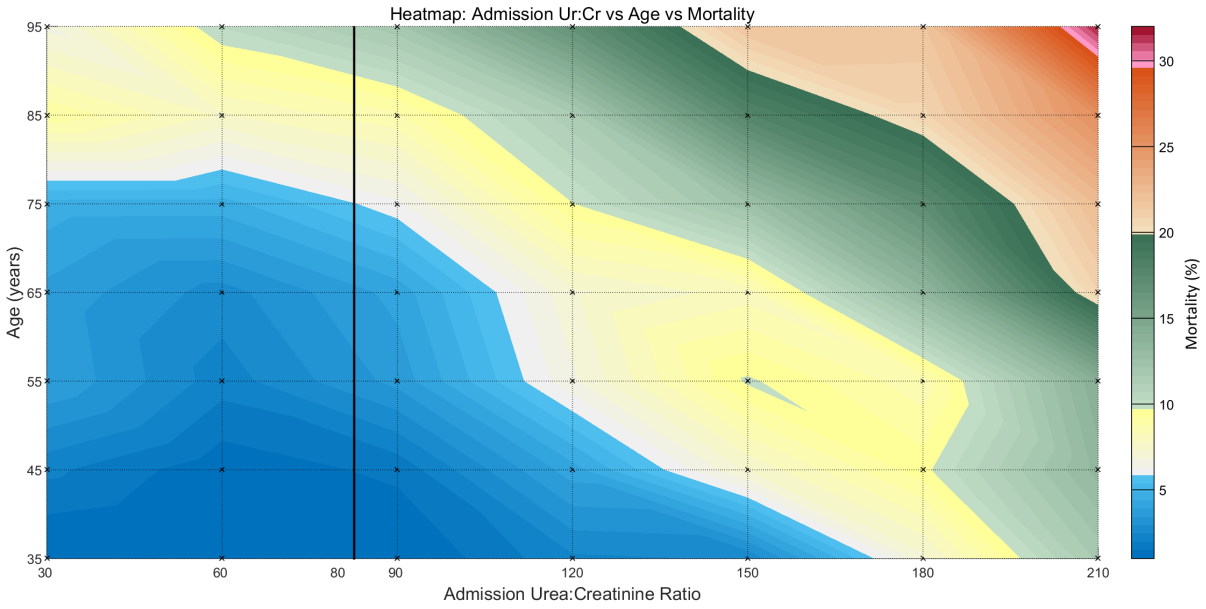
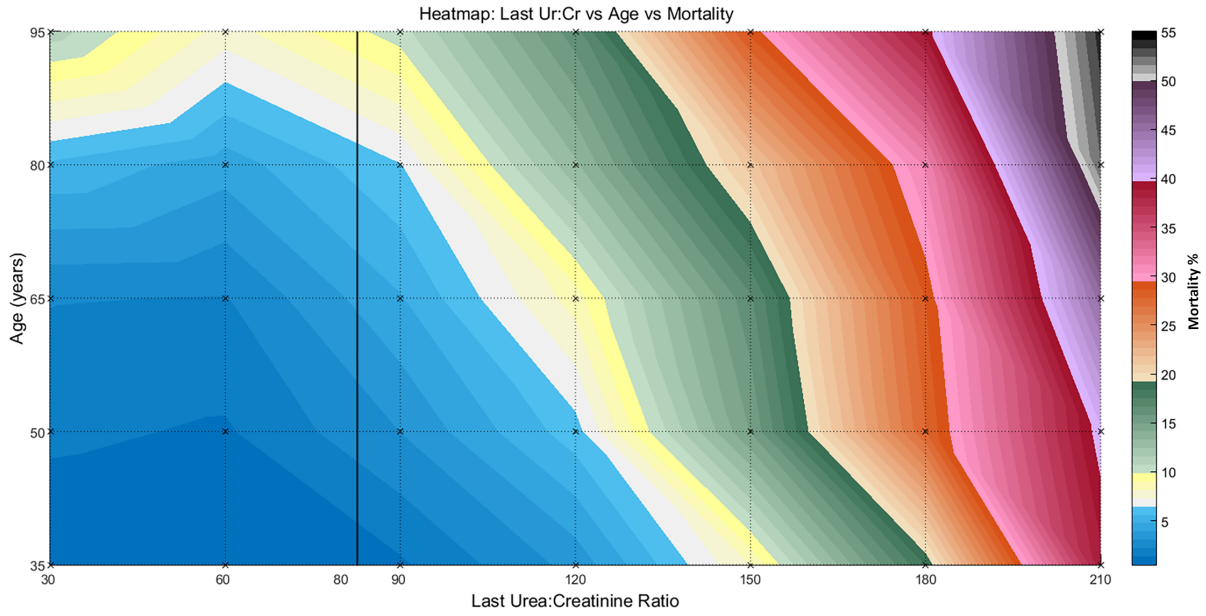


Figure 3.8



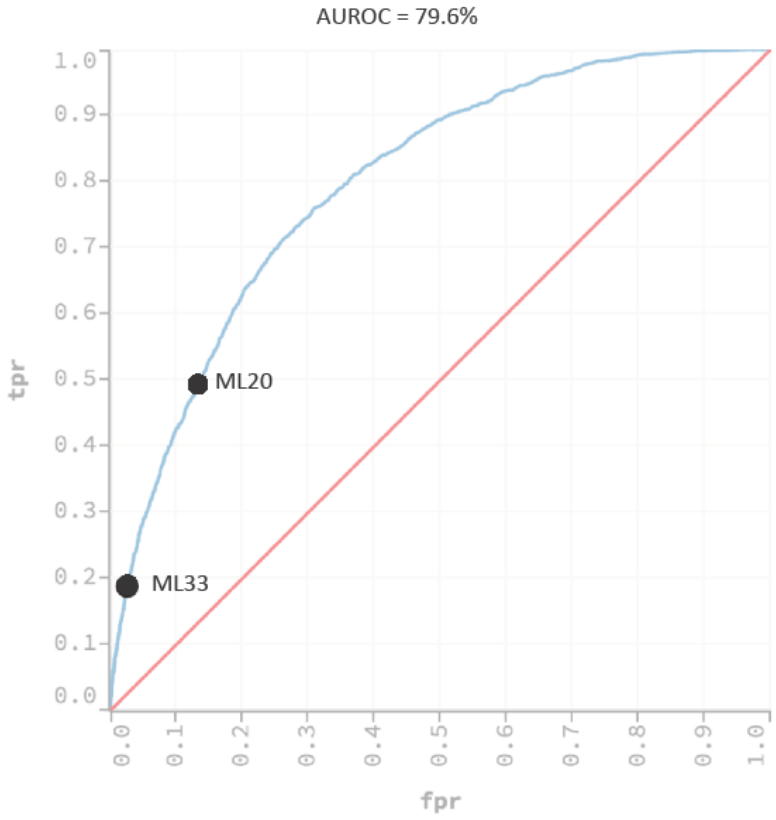
3.3.5 Predicting mortality: ML-Dehydration

There were no significant differences between the train and test datasets in terms of baseline characteristics (Table 3.6). The characteristics of the ML-Dehydration model were: max tree depth: 27; mean number of leaves: 551.7. ML-Dehydration achieved an overall area under the receiver operator curve (AUROC) of 79.6% (mean squared error: 0.055; logloss: 0.20; mean per-class error: 0.34) (Figure 3.9). This was similar to the cross-validation AUROC of 79.6%, confirming its validity. The training dataset's AUROC was 91.2%. At the positive predictive value threshold of ML20 (1:5 (20%)), the predictive performance of ML-Dehydration in identifying patients who died in hospital, at the point of their second blood test, had a sensitivity of 49.1% and a specificity of 86.7%. The ML33 (1:3 (33%)) threshold had a sensitivity of 16.6% and specificity of 97.7%.

Table 3.6: Baseline characteristics of the train and test datasets

Baseline characteristics	**Train Dataset n=55,964**	**Test Dataset n=23,985**
Median (IQR) age, years	71 (51–83)	71 (51–83)
No. (%) Male Sex	26,663 (47.6%)	11,505 (48%)
Median (IQR) length of stay, days	6.7 (3–13.5)	7 (3–13)
No. died in-hospital (%)	3,557 (6.4%)	1,523 (6.3%)

Figure 3.9: Area Under the Receiver Operator Curve (AUROC)



tpr: True Positive Rate (Sensitivity); fpr: False Positive Rate (1 – Specificity); the dots represent specific threshold values used to discriminate between patients likely to live or die

3.4 Discussion

I have undertaken my analysis with data for 79,949 patients across three NHS trust hospitals. These results represent the largest analysis to date that explores the association of Ur:Cr with in-hospital outcome for a large series of unselected emergency admissions to secondary care. In addition, I examined the value of assessing the dynamic change in Ur:Cr and the importance of its association with patient outcome.

I found that an admission Ur:Cr ratio of ≥80 was present in nearly half of all patients (consistent with previous smaller studies[58]), and Ur:Cr rose to ≥80 during hospital stay in over a quarter of those with a normal Ur:Cr on admission. A high Ur:Cr ratio status was strongly associated with increased LoS. Though I cannot attribute rises in Ur:Cr solely to dehydration, such data are consistent with the doubling of LoS in acute coronary syndrome patients suffering dehydration (7.8 vs 3 days), and the six-fold and tenfold rise in risk of AKI and cardiogenic shock respectively in such patients.[38]

AKI affected 10.7% of inpatients, matching recently published data[130], and this figure was similar to that suggested by a recent meta-analysis.[78] Likewise, my data confirmed the high prevalence of an elevated Ur:Cr amongst those who develop AKI, as reported by others[53], and its association with risk of dying with AKI. In contrast to previous work[53,130], I have explored the impact of initial ratio status and its change with mortality in AKI patients, demonstrating a wide range in risk of death.

Higher Ur:Cr on admission was non-linearly associated with in-hospital mortality, and change in Ur:Cr strongly correlated with both LoS and mortality via a complex relationship, further affected by age. The contributions of the individual components of the ratio also had a dramatic impact on the association of the ratio with mortality.

Such complex interactions suggest the need for non-linear models for the prediction of poor outcomes. The ML-Dehydration model, when used in a real-world scenario (sequential blood tests, not knowing whether it is to be the last), identifies 50% (sensitivity) of patients who will subsequently die, from only their second blood test onwards. The positive predictive value of 20% means that if the model classifies the patient as being likely to die, there is a 1:5 chance that this will subsequently occur.

Reductions in Ur:Cr, when the ratio at admission was ≥80, probably reflect effective intervention, though I am unable to describe the granularity of such interventions. It is likely that treatment of dehydration and sepsis (alone or in combination) might play a dominant role in generating this

reduction. Fluid therapy is an essential part of most sepsis bundles, improving relative or actual intravascular fluid depletion, and thus outcome.[131,132,133, 134,135]

Similarly, new or persistent rises in the Ur:Cr ratio increase relative risk of death (from 6.4 to 47.9), and may represent opportunities to improve patient management. Use of the Ur:Cr ratio and its trajectory in order to highlight patients at risk may provide guidance for interventions. Such interventions, proposed in the UK's National Institute for Health and Care Excellence guidelines on AKI[78] and on prescribing intravenous fluids[6], include automated alerts to clinicians, fluid administration, and appropriate training in the assessment of patients' fluid and electrolyte needs.

This specific investigation has particular strengths. It is the largest study, in terms of both number of patients and number of blood tests analysed, to have examined the following in hospitalised patients: 1) Ur:Cr, both its absolute value and its change; and 2) AKI, using quantitative biochemical definitions. It is also the first study that examines the relationship between Ur:Cr on admission *and its change during hospitalisation*, with outcome. The use of minimal exclusion criteria increases the generalisability of my findings. The use of data from three hospital groups, over approximately the same time period, thus reducing biases associated with single-site studies. Finally, the study is topical and relevant, as the impact of dehydration on outcome, and the search for measures to mitigate this, has become a concern for England's Care Quality Commission (CQC)[69], the Patient Association[70], and The British Parliamentary Ombudsman[71], with such concerns being echoed in an independent inquiry[72] and in the media.[73,74,75,76,77]

Nonetheless, this study does have limitations. The data was collected from hospitals which when compared to the national average had higher quality ratings (CQC and standardised mortality ratio) than the England average; this could introduce bias into the results. However, this bias, if it exists, would result in an underreporting of the problem of dehydration lending further support to the unreported scale of the problem, and thus the potential for improvement of patient care. It has not explored any excessive deaths amongst those who had a high Ur:Cr on admission, and who failed to survive to receive a second blood test. Nor has it examined the impact of those discharged prior to a second test. In identifying cases of AKI, baseline creatinine was defined as the earliest blood test performed within 24 hours of hospital admission; however, since these analyses were carried out, a National NHS England algorithm (NHSE-algorithm) definition of baseline creatinine has been mandated, which utilises creatinine blood results obtained up to one year prior to admission. This new NHSE-algorithm definition of AKI was not used. Therefore, a proportion of patients who presented to hospital with new AKI (based on community creatinine values) were not classified as such, nor were those who had a low out-of-hospital creatinine value. However, these limitations led to a conservative

estimate of AKI prevalence in my analyses. Data on urine output as part of the KDIGO classification were also unavailable. Further, I did not investigate interactions with co-morbid diagnoses, as the data linkage across all co-morbidities was beyond the scope of this initial investigation. Most of these limitations can be addressed in a subsequent study.

I have demonstrated that the use of two widely available blood tests (urea and creatinine), combined with two demographic variables (age and sex), when analysed in a binary or continuous manner, can be operationalised as a powerful predictor of outcomes. These variables interact with each other in a complex fashion in determining such outcomes. Reductions in the Ur:Cr ratio, when elevated, are associated with improved outcome, and rises may indicate patient management that can be improved with targeted interventions.

Dehydration, is a condition that is easy to communicate and in the majority of cases relatively simple to treat. Building care pathways that continuously track hydration status in patients, and have interventions to maintain hydration, would be simple and require relatively few resources. Such interventions could include encouragement to eat and drink more, along with nursing support to enable patients to do this, as well as robust implementation of existing guidelines on intravenous fluid administration.

Finally, from a practical viewpoint, my research demonstrated that manipulating and analysing large datasets with multiple non-numerical categorical values was extremely memory-intensive and cumbersome in MATLAB alone, despite Mathworks (the maker of MATLAB) having recently added database feature- 'Data Tables'. Therefore, for all further data management and analysis, I made the decision to completely switch from MATLAB to a dedicated database (Microsoft SQLServer), and to an alternative analytic suite, consisting of R and h2o.

Chapter 4: Acute kidney injury: dynamic single parameter to ML

4.1 Introduction

The trends of blood results, i.e. their rise or fall, have always been extremely valuable to clinicians. In Chapter 3, I demonstrated that in-hospital calculations based on the dynamic change in just two simple blood test results can help to risk-stratify patients. Here I expand this work by incorporating additional variables and grouping patients into known categories (chronic kidney disease stage, method of admission, ethnicity, etc.) to highlight both the variability, and complexity of patient outcome when these additional variables are used to aid in risk classification. I then consolidate all available a priori information (including pre-hospital), as well as the dynamic change in all the raw blood results, to create an interpretable ML model to determine risk of poor outcome (death or renal replacement therapy) in hospital. This approach is applied to my entire dataset, and specifically addresses the occurrence of (and outcome from) AKI, a condition defined by the dynamic change of just one variable: creatinine. I have also compared my ML approach with the existing NHS England AKI (NHSE-AKI) algorithm, to discriminate between a patient's outcome from AKI.

As detailed in Chapter 1 (Section 1.4.3), AKI is a prevalent syndrome (upto 21% of hospitalised patients worldwide) that has multiple varied aetiologies. It is associated with poor healthcare outcome and increased healthcare costs (in excess of £1 billion/year in England and Wales). Thus, initiatives have been put in place to automatically diagnose and escalate care for patients with AKI.

AKI just uses the dynamic change in one variable, creatinine, for its diagnosis. A defined rise in a patient's serum creatinine, from their baseline has been used to severity stage a patient's AKI. These AKI severity stages are linked to recommendations for escalation of clinical care (Figure 1.9).[83] However, while more severe AKI stages are considered to be associated with poorer outcomes (death and renal replacement therapy (Drrt)), and thus require prioritisation of care, the risk stratification by AKI stage when further variables/parameters are taken into account is not clear at an individual patient level. Indeed, it remains unknown the degree to which differences in individual demographic data, the context of a patient's admission (for example, day of the week, month of the year and emergency vs planned), their past medical history, and results of other investigations, influence outcome in any one AKI stage.

Furthermore, using the mandated NHSE-AKI algorithm, the same AKI stage can be reached via multiple trigger routes. It was unclear whether the different routes by which an AKI stage is reached influences the outcome associated with that stage.

Finally, most trigger routes for defining the AKI stage of the NHSE-AKI algorithm rely on the comparison of the current serum creatinine with a calculated baseline serum creatinine. The baseline serum creatinine is defined as either the minimum level in the last seven days, or the median in the last 8–365 days; whichever is lower. This results in a different baseline creatinine being defined for every subsequent creatinine result, and which is thus used for calculation of AKI status and stage. Therefore, there is a mathematical possibility that multiple creatinine tests over a short period could result in a rise in the baseline creatinine, such that an individual is no longer defined (via the NHSE-AKI algorithm) as having AKI: I have labelled this situation as 'false-negative-AKI'. I sought to investigate whether this occurred, and to what extent.

4.2 Objectives

The purpose of this chapter is to determine whether analytics that incorporate multiple correlated and uncorrelated variables, over multiple time-points, could be used to provide more precise individual risk prediction for patients who are deemed to have AKI. This could replace the current AKI severity stage, thus creating a new model that could be used as to escalate individual patient care. I specifically sought to identify:

1. The epidemiology of admissions who are diagnosed with AKI in hospital.
2. The relationship between the NHSE-AKI algorithm-defined AKI stage and in-hospital outcome (Death or Renal Replacement Therapy (Drrt)), and whether this relationship differs according to method of admission and existing co-morbidities.
3. Whether Drrt risk differs depending on the route by which the AKI stage is defined.
4. Whether the NHSE-AKI algorithm fails to identify patients who continue to have AKI.
5. Whether a ML approach can better stratify risk than the currently mandated NHSE-AKI algorithm.

4.3 Methods

4.3.1 Patients

The dataset described in Chapter 2 was used for these analyses. In brief, I collated continuous electronically stored data from all adults (18 years and over) admitted to fourteen UK NHS acute hospital trusts from early 2005 to late 2015 (dates varying between trusts). Specifically, for this analysis, I included patients with a complete administrative record, and at least one serum creatinine result during their admission. Excluded were patients transferred to another hospital for continued medical care, and those who self-discharged (discharge destination codes: 48, 49, 50, 51, 52, 53, 55, 56, 57, 84, 87, 88, 98 and 99).[136] Also excluded were those known to be dependent on renal replacement therapy prior to admission, identified by ICD10 code Z99.2 or by their elective admission for RRT alone (one-day planned admission and OPCS4 codes X40, X40.1, X40.2, X40.3, X40.8 and X40.9). All other patients who received renal replacement therapy in hospital were included (ICD10 codes Z490, Z491, Z492 and OPCS4 codes X40, X40.1, X40.2, X40.3, X40.8 and X40.9). These admissions were deemed to have progressed from AKI to complete renal failure, requiring renal replacement therapy.

Each hospital admission was analysed independently.

4.3.2 Data

All results for sodium (mmol/L), potassium (mmol/L), urea (mmol/L), creatinine (µmol/L), haemoglobin (g/L), platelet (10^9/L), mean cell volume(fL), red cell distribution width (%) and white cell count (10^9/L) were recorded throughout each admission. In addition, where such results existed, they were captured up to one year prior to admission. Routinely collected administrative data were also recorded: patient age, diagnoses (ICD10), treatment codes (OPCS-4) and dates, and details relating to hospital admission and discharge. I selected these variables because of their ubiquity (available for most hospitalised patients).

Admissions were grouped by their method of admission (planned vs emergency), and their chronic kidney disease (CKD) stage[137] at admission. The Chronic Kidney Disease Epidemiology Collaboration equation[138] was used to calculate the estimated glomerular filtration rate (eGFR) using the first baseline creatinine value on admission. Baseline creatinine was defined using the NHSE-AKI algorithm (Figure 4.1) as the lower of the following: minimum creatinine result in the last 7 days, or the median creatinine result in the last 8–365 days. The four CKD stages used were: (1) none/mild: no CKD/CKD1-2, eGFR > 60 ml/min/1.73 m^2; (2) moderate: CKD3, eGFR 30–60 ml/min/1.73 m^2; (3) severe/endstage: eGFR<30 ml/min/1.73 m^2; and (4) 'unknown' for when no baseline creatinine data were available.[137]

The epidemiology of the resultant dataset was described, detailing the number of admissions based on their demographics (sex, age and ethnicity) and admissions details (method of admission, day of the week, and month of year of admission).

For each creatinine result, the NHSE-AKI algorithm defined the baseline creatinine, the AKI stage and the trigger route for that AKI stage. The maximum AKI stage attained for each patient during each admission (AKI_{max}) was used to calculate prevalence. The ratio of measured creatinine:baseline creatinine [Cr:BCr] was also calculated, and its maximal value $[Cr:BCr]_{max}$ recorded.

I calculated Drrt rates for AKI_{max}, and for each algorithm route. However, AKI_{max} can only be determined post hoc (unless the first AKI stage is 3), and thus has limited utility as a 'live' predictor of AKI outcome. Therefore, Drrt rates for the *first* AKI event were also calculated- AKI_{first}. Finally, AKI-related Drrt rates were calculated for admissions grouped by their co-morbidities and their admission Ur:Cr ratio status ($\geq$80, historically used to suggest dehydration[12]).

Cases of false-negative-AKI were defined as those suffering from AKI in whom a sustained rise in creatinine (one which did not fall, or which continued to rise) then led the NHSE-AKI algorithm to define them as free from AKI.

In addition, I explored, for admissions who did not die or have renal replacement therapy in hospital, their final AKI stage (AKI_{last}) at discharge from hospital, as both a proportion of total AKI admissions and of total admissions. The number of creatinine results, and AKI alerts between the different admission groups, were also calculated.

4.3.3 Statistical and ML Analyses

I used the Wilson method to calculate the 95% confidence intervals for Drrt rate. Prevalence and Drrt rates were compared using Fisher's Exact Test. Pearson's correlation coefficients were also calculated. Statistical significance was defined as a p-value of less than 0.05.

ML, specifically the GB technique[139], was employed to create a tool for the early prediction of Drrt. Because early identification of individual risk might offer the greatest clinical utility, the only results used to generate the ML-AKI model were the summary (mean, standard deviation, frequency) blood results from the previous 365 days, the first set of admission blood results, and those that first triggered an AKI stage, with the administrative information available at those times. The a priori predictor variables I used to create the model are listed in Table 4.1. A random 70% data sample

defined the training-dataset. The model was built using 5-fold cross-validation. The hyper-parameters of the model were determined via a grid search. The final ML-AKI model (post hyper-parameter selection via grid search) was then tested on the remaining 30% of the data (test-dataset). Area under the receiver operator curve (AUROC) was calculated for the whole model. From the trained model, a threshold needed to be set to classify examples as either Drrt or non Drrt. I defined three thresholds (ML50, ML33, ML25), selected to produce positive predictive values (PPV) of 1:2, 1:3 and 1:4 respectively on the train data.

The relative importance of each predictor variable was calculated using a function developed by Friedman (2001)[140] and implemented in an h2o GB package. 'The measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees. The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response.'

The software used was Microsoft R Open 3.2.5 (Microsoft, USA: https://mran.microsoft.com/rro), Microsoft SQLServer 2014 Enterprise (Microsoft, USA: https://www.microsoft.com/en-gb/server-cloud/products/sql-server/overview.aspx), RStudio (RStudio, USA: https://www.rstudio.com), R (The R Foundation, Austria: https://www.r-project.org), h2o (h2o.ai, USA: www.h2o.ai), and .Net and C# (Microsoft, USA) scripts.

Table 4.1: ML-AKI Model Variables

Variable Name		**Type of Variable**
Age		Numerical
Sex		Categorical
Ethnic Category		
Patient Category		
Admission Method Code		
Source of Admission Code		
Day of Admission		
Month of Admission		
Blood Results		
Admission	eGFR group	Categorical
	eGFR	Numerical
	Baseline Creatinine	

	Creatinine	
	Urea	
	Sodium	
	Potassium	
	Urea:Creatinine Ratio	
	Haemoglobin	
	Platelets	
	White Cell Count	
	Mean Cell Volume	
	Red Cell Distribution Width	
Blood results at the First AKI Stage (SI units)		
First AKI Stage	Baseline Creatinine	Numerical
	Creatinine	
	Urea	
	Sodium	
	Potassium	
	Urea:Creatinine Ratio	
	Haemoglobin	
	Platelets	
	White Cell Count	
	Mean Cell Volume	
	Red Cell Distribution Width	
Ratio of (Ur:Cr first AKI stage) : (Ur:Cr admission)		Numerical
Creatinine:Baseline Creatinine Ratio at first AKI stage		
Creatinine 48 hr rise at first AKI stage		
AKI stage at first AKI stage		Categorical
Summary Pre-Admission Blood Results (last year (365 days), SI units)		
Number	Creatinine	Numerical
	Urea	
	Sodium	
	Potassium	
	Haemoglobin	
	Platelets	
	White Cell Count	
	Mean Cell Volume	
	Red Cell Distribution Width	
Mean	Creatinine	Numerical
	Urea	
	Sodium	
	Potassium	
	Haemoglobin	
	Platelets	
	White Cell Count	

	Mean Cell Volume	
	Red Cell Distribution Width	
	Creatinine	
Standard deviation	Urea	Numerical
	Sodium	
	Potassium	
	Haemoglobin	
	Platelets	
	White Cell Count	
	Mean Cell Volume	
	Red Cell Distribution Width	
Number of AKI Stage 1 Alerts		Numerical
Number of AKI Stage 2 Alerts		
Number of AKI Stage 3 Alerts		
Co-morbidities		
Condition	**ICD10 code**	Categorical
Essential (primary) hypertension	I10	
Type 2 diabetes mellitus	E11	
Atrial fibrillation or flutter	I48	
Chronic ischaemic heart disease	I25	
Chronic kidney disease	N18	
Heart Failure	I50	
Disorders of lipoprotein metabolism and other lipidaemias	E78	
Personal history of diseases of the circulatory system	Z867	
Personal history of psychoactive substance abuse	Z864	
Other chronic obstructive pulmonary disease	J44	
Personal history of long-term (current) use of anticoagulants	Z921	
Personal history of allergy to penicillin	Z880	
Other hypothyroidism	E03	
Asthma	J45	
Smoking	F17	
Angina Pectoris	I20	
Depressive episode	F32	
Other symptoms and signs involving the nervous and musculoskeletal systems	R29	
Presence of aortocoronary bypass graft	Z951	
Presence of electronic cardiac devices	Z950	

Presence of orthopaedic joint implants	Z966	
Mental and behavioural disorders due to use of alcohol	F10	
Secondary malignant neoplasm of other and unspecified sites	C79	
Presence of coronary angioplasty implant and graft	Z955	
Obesity	E66	
Myeloid leukaemia	C92	
Personal history of allergy to other drugs, medicaments and biological substances	Z888	
Personal history of malignant neoplasm of digestive organs	Z850	
Epilepsy	G40	
Multiple myeloma and malignant plasma cell neoplasms	C90	
Personal history of diseases of the nervous system and sense organs	Z866	
Secondary and unspecified malignant neoplasm of lymph nodes	C77	

4.4 Results

Overall 1,972,130 hospital admissions (942,061 unique patients; median 1 (inter-quartile range: 1 to 2) admissions per patient) met the inclusion criteria (Consort Diagram: Figure 4.1).

4.4.1 Baseline characteristics

Of the 1,972,130 admissions, 39.6% (780,970) were male. Emergency admissions accounted for 57.8% (1,139,220) of all admissions, with planned admissions accounting for 42.2% (832,910). Their baseline characteristics are described in Table 4.2. A full breakdown by ethnicity, date of admission and method of admission is provided in Tables 4.3, 4.4 and 4.5 respectively. The largest ethnic group was British (49.6%) and the smallest White and Black African (0.2%). The distribution of the top 25 co-morbidities of all the admissions is also shown in the baseline table (Table 4.2). The prevalence of the top three conditions were hypertension at 22.3% (439,521), type 2 diabetes mellitus at 10.1% (200,011) and chronic ischaemic heart disease at 9.1% (179,026). The distribution and relationship between by age, CKD group and method of admission is shown in Figure 4.2.

Figure 4.1 Consort Diagram

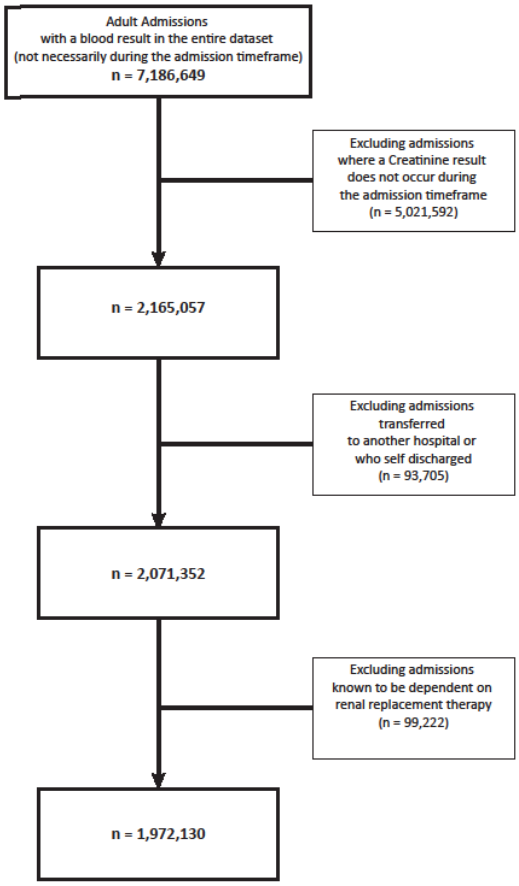


Figure 4.2: Histogram of Age and Chronic Kidney Disease status
CKD status (No CKD/CKD 1–2– 1,020,742 ; CKD 3– 242,239; CKD 4–5– 63,244 ; Unknown– 645,905

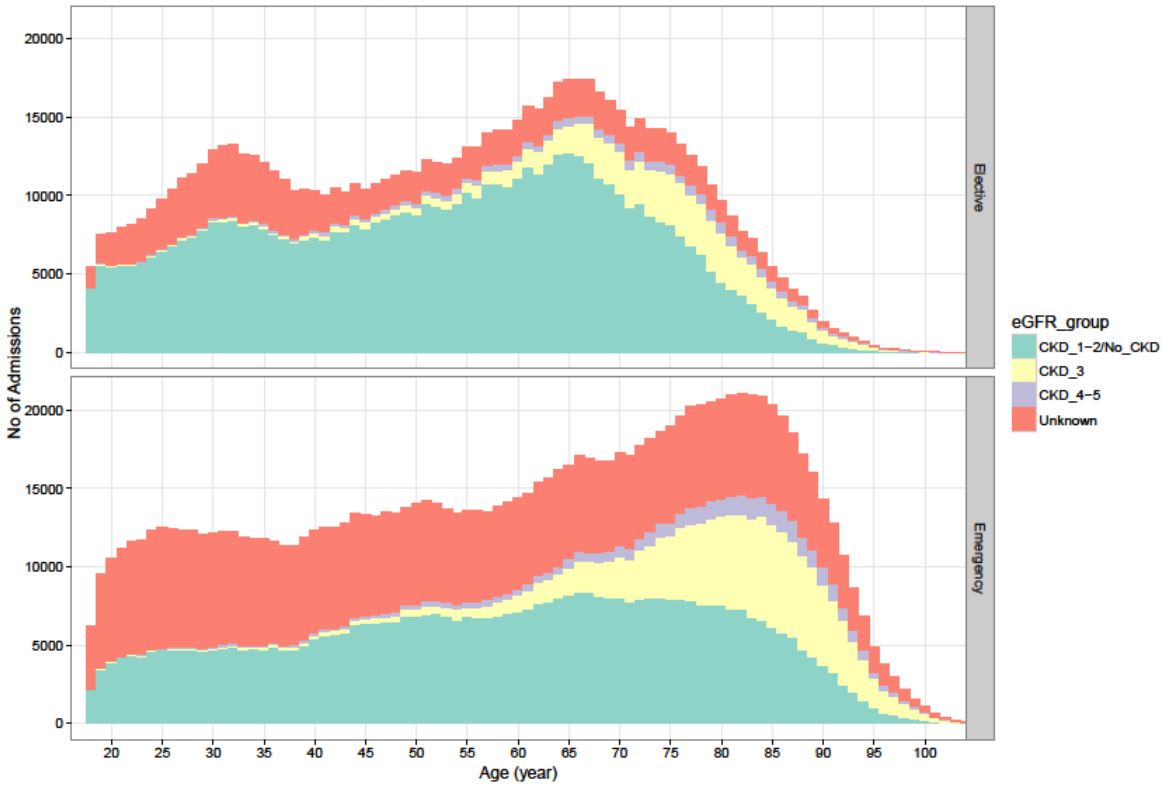


Table 4.2: Baseline Characteristics						
		Emergency		**Planned**		**p value**
Number of admissions		1,139,220		832,910		
Male		425,528 (47.1%)		355,442 (42.7%)		
Age (median (IQR)		67 (47–80) years		56 (37–69) years		$<1 \times 10^{-2}$
Length of stay (median (IQR))		4 (1–9) days		0 (0–4) days		$<1 \times 10^{-2}$
Drrt (%)		53,206 (5.9%)		9,031 (1.1%)		$<1 \times 10^{-2}$
CKD Stage		**Frequency (percentage of Emergency, Planned)**				
None to Mild (CKD 0–2)		470,517 (52.1%)		556,547 (66.8%)		$<1 \times 10^{-2}$
Moderate (CKD 3)		157,655 (17.5%)		85,453 (10.3%)		$<1 \times 10^{-2}$
Severe (CKD 4–5)		39,941 (4.4%)		23,419 (2.8%)		$<1 \times 10^{-2}$
Unknown (CKD ?)		234,474 (26%)		167,491 (20.1%)		$<1 \times 10^{-2}$
Co-Morbidities – ICD10 code		**Frequency (percentage of Emergency, Planned)**				
Essential (primary) hypertension	I10	303,446	26.6%	136,075	16.3%	$<1 \times 10^{-2}$
Type 2 diabetes mellitus	E11	144,337	12.7%	55,674	6.7%	$<1 \times 10^{-2}$
Chronic ischaemic heart disease	I25	130,211	11.4%	48,815	5.9%	$<1 \times 10^{-2}$
Atrial fibrillation or flutter	I48	122,345	10.7%	31,513	3.8%	$<1 \times 10^{-2}$
Disorders of lipoprotein metabolism and other lipidaemias	E78	116,971	10.3%	50,492	6.1%	$<1 \times 10^{-2}$
Personal history of psychoactive substance abuse	Z864	86,855	7.6%	34,458	4.1%	$<1 \times 10^{-2}$
Other chronic obstructive pulmonary disease	J44	81,506	7.2%	16,871	2.0%	$<1 \times 10^{-2}$
Personal history of diseases of the circulatory system	Z867	80,334	7.1%	32,419	3.9%	$<1 \times 10^{-2}$
Heart failure	I50	80,118	7.0%	10,684	1.3%	$<1 \times 10^{-2}$
Asthma	J45	75,699	6.6%	35,592	4.3%	$<1 \times 10^{-2}$
Smoking	F17	68,267	6.0%	23,266	2.8%	$<1 \times 10^{-2}$
Personal history of allergy to penicillin	Z880	64,676	5.7%	31,708	3.8%	$<1 \times 10^{-2}$
Personal history of long-term (current) use of anticoagulants	Z921	57,213	5.0%	24,221	2.9%	$<1 \times 10^{-2}$
Chronic kidney disease	N18	56,625	5.0%	21,040	2.5%	$<1 \times 10^{-2}$
Depressive episode	F32	54,119	4.8%	14,018	1.7%	$<1 \times 10^{-2}$
Other hypothyroidism	E03	51,859	4.6%	21,345	2.6%	$<1 \times 10^{-2}$
Angina Pectoris	I20	46,328	4.1%	13,616	1.6%	$<1 \times 10^{-2}$

Other symptoms and signs involving the nervous and musculoskeletal systems	R29	41,548	3.6%	2,351	0.3%	$<1 \times 10^{-2}$
Mental and behavioural disorders due to use of alcohol	F10	37,550	3.3%	3,381	0.4%	$<1 \times 10^{-2}$
Presence of orthopaedic joint implants	Z966	33,363	2.9%	16,687	2.0%	$<1 \times 10^{-2}$
Presence of aortocoronary bypass graft	Z951	30,427	2.7%	10,747	1.3%	$<1 \times 10^{-2}$
Presence of electronic cardiac devices	Z950	28,586	2.5%	7,503	0.9%	$<1 \times 10^{-2}$
Epilepsy	G40	28,253	2.5%	8,672	1.0%	$<1 \times 10^{-2}$
Presence of coronary angioplasty implant and graft	Z955	25,646	2.3%	10,508	1.3%	$<1 \times 10^{-2}$
Secondary malignant neoplasm of other and unspecified sites	C79	17,895	1.6%	23,590	2.8%	$<1 \times 10^{-2}$
Personal history of allergy to other drugs, medicaments and biological substances	Z888	17,533	1.5%	14,844	1.8%	$<1 \times 10^{-2}$
Personal history of diseases of the nervous system and sense organs	Z866	17,347	1.5%	6,028	0.7%	$<1 \times 10^{-2}$
Obesity	E66	16,717	1.5%	17,294	2.1%	$<1 \times 10^{-2}$
Personal history of malignant neoplasm of digestive organs	Z850	13,839	1.2%	9,192	1.1%	$<1 \times 10^{-2}$
Secondary and unspecified malignant neoplasm of lymph nodes	C77	6,470	0.6%	20,853	2.5%	$<1 \times 10^{-2}$
Multiple myeloma and malignant plasma cell neoplasms	C90	5,381	0.5%	34,780	4.2%	$<1 \times 10^{-2}$
Myeloid leukaemia	C92	4,116	0.4%	24,218	2.9%	$<1 \times 10^{-2}$

A total of 6,343,530 serum creatinine tests were performed during hospital admissions, with the median creatinine per admission being 1 (IQR: 1 to 3). For emergency admissions, there was a total of 4,315,120 during-admission creatinine results, with a median per admission of 2 (IQR: 1 to 4). For planned admissions, the total for during-admission creatinine results was 2,028,410, with a median per admission of 1 (IQR: 1 to 2). However, a total of 22,601,915 serum creatinine results were analysed, comprising during-admission blood results (6,343,530) as well as those obtained up to one year before admission (16,258,385). Because each admission was treated independently, patients admitted more than once during the span of a year would have the results from those separate admissions analysed separately. Some continuous variables were correlated (r: -0.26 to 0.83, $p < 1 \times 10^{-5}$; Table 4.3).

Table 4.3: Correlation Coefficients

First AKI	Cr 48h	Cr:BCr	BCr	Cr	Ur	Na	K	Hb	Plts	WCC	MCV	RDW
Age	-0.04	-0.02	0.01	0.01	0.27	0.02	0.08	0.03	0.07	0.07	0.09	0.04
Creatinine (Cr)48h Rise		0.44	0.37	0.61	0.34	-0.03	0.21	0.00	-0.03	0.01	0.00	-0.02
Cr:BCr			-0.11	0.33	0.28	-0.05	0.15	0.03	0.03	0.04	-0.01	0.01
Baseline Cr (BCr)				0.83	0.48	0.00	0.20	-0.12	-0.09	-0.04	0.00	-0.01
Creatinine (Cr)					0.63	-0.02	0.29	-0.10	-0.07	-0.02	0.01	0.00
Urea (Ur)						0.02	0.34	-0.05	-0.04	0.06	0.02	0.08
Sodium (Na)							-0.17	0.04	-0.07	-0.01	0.16	0.06
Potassium (K)								-0.05	0.13	0.04	0.01	0.03
Hemoglobin (Hb)									0.09	0.04	0.07	-0.26
Platelets (Plts)										0.14	-0.12	-0.01
White cell count (WCC)											0.00	0.06
Mean Cell Volume (MCV)												-0.15

4.4.2 Epidemiology / Variation in Prevalence and Progression of AKI Admissions

AKI occurred in 170,596 (8.6%) admissions, representing 122,696 (13%) unique patients. Of these, 170,596 admissions with AKI, 65,772 (38.6%) were diagnosed with AKI on their first creatinine test in hospital, indicating community-acquired AKI. Tables 4.4, 4.5 and 4.6 show the AKI prevalence by ethnic group, and by month, day and method of admission respectively. The highest proportion of AKI admissions occurred in Bangladeshi (10.2%), Caribbean (9.8%), Indian (9.7%) and Irish (9.7%) ethnicities, with the lowest in those with a mixed background (6.5% to 7.2%). The highest proportion

of AKI admissions occurred over winter (January (9.2%) and December (9.2%)) and the weekend (Saturday (10.3%) and Sunday (10.5%)). Analysing the data by method of admission, AKI was almost twice as common in emergency as in planned admissions (10.7% (122,346) vs 5.8% (48,240), $p<1 \times 10^{-10}$). However, even within these groups (Planned and Emergency) there was significant variation. Within Planned admissions, maternity related admissions had the lowest AKI prevalence (maternity (3%) vs rest of Planned (6.4%)) and those transferred from another hospital the highest (17%).

Table 4.4: Ethnicity

Ethnicity	All Admissions N (% of total) Total = 1,972,130	AKI only Admissions N (% Total AKI admissions) (Total AKI s = 170,596)	Sub-category prevalence (Percentage of AKI admissions by Total admissions)
British (A)	977,717 (49.6%)	85,463 (50.1%)	8.7%
Not stated (Z)	481,395 (24.4%)	42,173 (24.7%)	8.8%
Any other white background (C)	144,337 (7.3%)	10,744 (6.3%)	7.4%
Any other ethnic group (S)	63,986 (3.2%)	4,930 (2.9%)	7.7%
African (N)	53,095 (2.7%)	4,245 (2.5%)	8.0%
Indian (H)	44,810 (2.3%)	4,357 (2.6%)	9.7%
Caribbean (M)	41,297 (2.1%)	4,042 (2.4%)	9.8%
Any other Asian background (L)	40,313 (2.0%)	3,628 (2.1%)	9.0%
Irish (B)	33,739 (1.7%)	3,283 (1.9%)	9.7%
Bangladeshi (K)	25,430 (1.3%)	2,600 (1.5%)	10.2%
Any other black background (P)	20,318 (1.0%)	1,563 (0.9%)	7.7%
Pakistani (J)	18,414 (0.9%)	1,528 (0.9%)	8.3%
Chinese (R)	8,585 (0.4%)	757 (0.4%)	8.8%
Any other mixed background (G)	8,165 (0.4%)	531 (0.3%)	6.5%
White and black Caribbean (D)	4,404 (0.2%)	315 (0.2%)	7.2%
White and Asian (F)	3,082 (0.2%)	214 (0.1%)	6.9%
White and black African (E)	3,043 (0.2%)	213 (0.1%)	7.0%

Table 4.5: Date of Admission			
	All Admissions N (% of total) Total = 1,972,130	**AKI only Admissions N (% of Total AKI admissions) (Total AKI admissions = 170,596)**	**Sub-category prevalence (Percentage of AKI admissions by Total admissions)**
Admission Month			
January	172,004 (8.7%)	15,868 (9.3%)	9.2%
February	157,964 (8.0%)	14,285 (8.4%)	9.0%
March	166,252 (8.4%)	14,609 (8.6%)	8.8%
April	162,938 (8.3%)	13,889 (8.1%)	8.5%
May	169,118 (8.6%)	14,136 (8.3%)	8.4%
June	157,386 (8.0%)	13,221 (7.8%)	8.4%
July	166,086 (8.4%)	13,861 (8.1%)	8.3%
August	158,295 (8.0%)	13,336 (7.8%)	8.4%
September	161,344 (8.2%)	13,500 (7.9%)	8.4%
October	171,459 (8.7%)	14,618 (8.6%)	8.5%
November	166,452 (8.4%)	14,325 (8.4%)	8.6%
December	162,832 (8.3%)	14,938 (8.8%)	9.2%
Admission Day of Week			
Monday	327,787 (16.6%)	27,769 (16.3%)	8.5%
Tuesday	349,597 (17.7%)	28,863 (16.9%)	8.3%
Wednesday	332,633 (16.9%)	26,951 (15.8%)	8.1%
Thursday	330710 (16.8%)	27,357 (16%)	8.3%
Friday	301005 (15.3%)	25,384 (14.9%)	8.4%
Saturday	162,819 (8.3%)	16,697 (9.8%)	10.3%
Sunday	167,579 (8.5%)	17,565 (10.3%)	10.5%

Table 4.6: Method of Admission			
Method of admission	**All Admissions N (% of total) Total = 1,972,130**	**AKI only Admissions N (% of Total AKI admissions) Total AKI admissions= 170,596**	**Sub-category prevalence (Percentage of AKI admissions by Total admissions)**
Emergency – Accident and emergency of the health care provider (21)	855,020 (43.4%)	90,929 (53.3%)	10.6%
Planned – Elective planned (13)	305,875 (15.5%)	13,233 (7.8%)	4.3%
Planned – Elective waiting list (11)	257,598 (13.1%)	18,012 (10.6%)	7.0%
Emergency – General practitioner (22)	156,829 (8.0%)	13,362 (7.8%)	8.5%
Planned – Elective booked (12)	118,528 (6.0%)	6,888 (4.0%)	5.8%
Planned – Maternity admitted ante-partum (31)	106,904 (5.4%)	3,228 (1.9%)	3.0%
Emergency – Transfer of any admitted patient from another hospital provider in an emergency (28)	80,034 (4.1%)	11,611 (6.8%)	14.5%
Planned – Transfer of any admitted patient from another hospital provider other than in an emergency (81)	39,033 (2.0%)	6,655 (3.9%)	17.0%
Emergency – Consultant clinic, of this or another health care provider (24)	36,202 (1.8%)	5023 (2.9%)	13.9%
Emergency – Bed bureau (23)	8,608 (0.4%)	998 (0.6%)	11.6%
Planned – Admitted post-partum (32)	4,510 (0.2%)	188 (0.1%)	4.2%
Emergency – Other emergency admission (2D)	1,749 (0.1%)	260 (0.2%)	14.9%
Emergency – Admission via mental health crisis resolution team (25)	414 (0.0%)	29 (0.0%)	7.0%
Planned – Not known / validation error (99)	308 (0.0%)	25 (0.0%)	8.1%
Emergency – Other means (2B)	259 (0.0%)	103 (0.1%)	39.8%
Planned – The birth of a baby in this health care provider (82)	125 (0.0%)	8 (0.0%)	6.4%
Emergency – Accident and emergency department of another provider where the patient had not been admitted (2A)	105 (0.0%)	31 (0.0%)	29.5%
Planned – Transfer of any admitted patient from another hospital provider other than in an emergency (83)	29 (0.0%)	3 (0.0%)	10.3%

The total number of in-hospital creatinine tests performed on those with (or who developed) AKI was 1,927,488, representing 30.4% of all in-hospital creatinine tests. Of these 1,927,488 results, 461,546 (23.9%) triggered an AKI alert as per the NHSE-AKI algorithm. Breaking this down by method of admission, 24% (100 * (337,101 / 1,407,425)) of creatinine tests done in Emergency admissions with AKI resulted in an AKI alert, which was similar to that of Planned admissions with AKI at 23.9% (100 * (124,445 / 520,063)). The median number of AKI alerts per admission with AKI was 1 (IQR: 1 to 3). This remained the same even when admissions were further divided into Emergency and Planned.

AKI prevalence in admissions with hypertension, type 2 diabetes mellitus, heart failure and chronic obstructive pulmonary disease (COPD) were 13.3%, 16.8%, 24.5% and 15.4% respectively. Overall, an AKI_{max} of stage 1, 2 and 3 occurred in 5.6%, 1.5% and 1.5% respectively. There was also variation in the characteristics of admissions with AKI, with respect to their CKD status/AKI triggering route (range: 0.01%–17.3%, Figure 4.3), and co-morbidities (range: 1.4%–41%, Table 4.7). For example, 41% of AKI admissions were dehydrated on admission, 13% had heart failure and 8.9% had COPD.

Figure 4.3: Prevalence of Max AKI by trigger route and Chronic Kidney Disease (CKD) status. Percent of total AKI (1,2,3; n = 170,596)

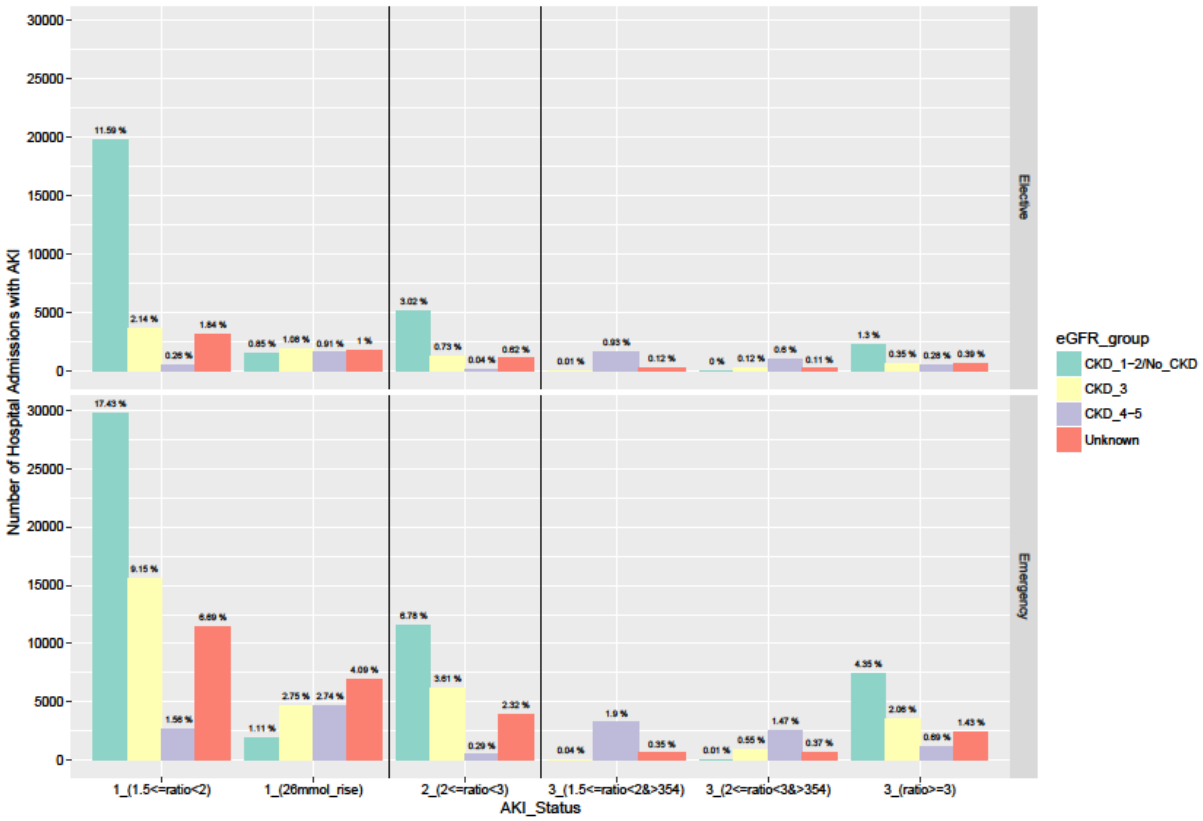


Table 4.7: Co-morbidities: Prevalence/Drrt rates in AKI admissions

Co-morbidity	ICD10 code	Prevalence % (n)	Drrt rate (%)		p < 0.05
			Condition Present	Condition Not Recorded	
Essential (primary) hypertension	I10	34.2% (58,278)	21.9% (21.5–22.2)	20.6% (20.3–20.8)	Y
Type 2 diabetes mellitus	E11	19.7% (33,572)	23.7% (23.3–24.2)	20.3% (20.1–20.6)	Y
Atrial fibrillation or flutter	I48	17.2% (29,421)	32.7% (32.1–33.2)	18.6% (18.4–18.8)	Y
Chronic ischaemic heart disease	I25	17.0% (28,961)	26.2% (25.7–26.7)	19.9% (19.7–20.2)	Y
Chronic kidney disease	N18	13.5% (23,110)	31.8% (31.2–32.4)	19.3% (19.1–19.5)	Y
Heart Failure	I50	13.0% (22,217)	36.9% (36.2–37.5)	18.6% (18.4–18.8)	Y
Disorders of lipoprotein metabolism and other lipidaemias	E78	12.1% (20,665)	19.2% (18.7–19.7)	21.3% (21.0–21.5)	Y
Personal history of diseases of the circulatory system	Z867	9.1% (15,524)	23.7% (23.0–24.4)	20.7% (20.5–20.9)	Y
Personal history of psychoactive substance abuse	Z864	9.0% (15,340)	24.7% (24.1–25.4)	20.6% (20.4–20.8)	Y
Other chronic obstructive pulmonary disease	J44	8.9% (15,135)	30.8% (30.1–31.5)	20.1% (19.9–20.3)	Y
Personal history of long-term (current) use of anticoagulants	Z921	7.0% (11,905)	22.8% (22.1-23.6)	20.9% (20.7-21.1)	Y
Personal history of allergy to penicillin	Z880	5.6% (9,608)	18.1% (17.4–18.9)	21.2% (21.0–21.4)	Y
Other hypothyroidism	E03	5.6% (9,530)	23.1% (22.2–23.9)	20.9% (20.7–21.1)	Y
Asthma	J45	5.6% (9,494)	18.3% (17.5–19.0)	21.2% (21.0–21.4	Y
Smoking	F17	4.8% (8,257)	20.4% (19.6–21.3)	21.0% (20.8–21.2)	N
Angina Pectoris	I20	4.3% (7,321)	24.0% (23.0–25.0)	20.9% (20.7–21.1)	Y
Depressive episode	F32	4.2% (7,142)	19.6% (18.7–20.6)	21.1% (20.9–21.3)	Y
Other symptoms and signs involving the nervous and musculoskeletal systems	R29	4.0% (6,806)	26.8% (25.7–27.9)	20.8% (20.6–21.0)	Y
Presence of aortocoronary bypass graft	Z951	3.7% (6,336)	24.0% (23.0–25.1)	20.9% (20.7–21.1)	Y

Presence of electronic cardiac devices	Z950	3.6% (6,131)	28.8% (27.7–29.9)	20.7% (20.5–20.9)	Y
Presence of orthopaedic joint implants	Z966	3.5% (6,019)	21.7% (20.7–22.8)	21.0% (20.8–21.2)	N
Mental and behavioural disorders due to use of alcohol	F10	3.0% (5,137)	23.0% (21.8–24.1)	20.9% (20.7–21.1)	Y
Secondary malignant neoplasm of other and unspecified sites	C79	2.8% (4,734)	36.0% (34.6–37.4)	20.6% (20.4–20.8)	Y
Presence of coronary angioplasty implant and graft	Z955	2.6% (4,391)	20.7% (19.5–21.9)	21.0% (20.8–21.2)	N
Obesity	E66	2.4% (4,129)	21.6% (20.4–22.9)	21.0% (20.8–21.2)	N
Myeloid leukaemia	C92	2.3% (3,853)	11.5% (10.5–12.5)	21.2% (21.0–21.4)	Y
Personal history of allergy to other drugs, medicaments and biological substances	Z888	2.2% (3,819)	11.4% (10.4–12.5)	21.2% (21.0–21.4)	Y
Epilepsy	G40	2.1% (3,639)	20.8% (19.5–22.2)	21.0% (20.8–21.2)	N
Personal history of malignant neoplasm of digestive organs	Z850	2.0% (3,389)	19.6% (18.3–20.9)	21.0% (20.8–21.2)	Y
Multiple myeloma and malignant plasma cell neoplasms	C90	1.9% (3,159)	16.5% (15.2–17.8)	21.1% (20.9–21.3)	Y
Personal history of diseases of the nervous system and sense organs	Z866	1.7% (2,948)	23.3% (21.8–24.9)	21.0% (20.8–21.2)	Y
Secondary and unspecified malignant neoplasm of lymph nodes	C77	1.4% (2,354)	25.2% (23.5–27.0)	20.9% (20.8–21.1)	Y
			Urea : Creatinine Ratio		
			≥ 80	**< 80**	
Admission (First) Urea : Creatinine Ratio		41% (69,941)	26.5% (26.2–26.8)	17.4% (17.2–17.6)	Y

The relationship of first to subsequent maximum AKI stage is shown in Table 4.8: AKI stage advanced from that first detected in 13.6% (23,226) of cases.

Table 4.8: First AKI to Max AKI stage			
First AKI		**Subsequent Max AKI**	
AKI Stage	**Frequency**	**AKI Stage**	**Frequency (percentage of First AKI Stage)**
1	131,324	1	111,163 (84.6%)
		2	11,211 (8.5%)
		3	8,950 (6.8%)
2	21,537	2	18,472 (85.8%)
		3	3,065 (14.2%)
3	17,725	3	17,725 (100%)

4.4.3 AKI_{last} and false-negative-AKI

Of 134,754 AKI admissions who were successfully discharged from hospital and did not have renal replacement therapy (no Drrt), 43,369 (32.2%) still had an AKI stage on their last creatinine test before discharge. The AKI stage of these non-Drrt AKI discharges were: AKI_{last} stage 1: 35,378; AKI_{last} stage 2: 6,101; and AKI_{last} stage 3: 1,890.

Separately, 51.9% (88,609) of those suffering from AKI had a subsequent in-hospital creatinine result that indicated they were no longer classified as having AKI. Of these, over a fifth (21.3%, 18,902) had a false-negative-AKI. The Drrt rate in these was higher than in those cases in which a fall in serum creatinine had appropriately led to an AKI diagnosis ceasing to be made (29.4% (5,564) vs 15.8% (10,986), $p<0.01$).

4.4.4 Variation of Drrt in AKI admissions

A total of 64,296 (3.3% of admissions) either died (56,076; 2.8%) or had renal replacement therapy (10,053; 0.5%) in hospital. Of the 64,296, 55.7% (35,832: died=30,047; renal replacement therapy=7,476) had suffered AKI. Overall, Drrt occurred in 21% (35,832) of those who developed AKI, compared with 1.6% (28,464: died=26,029; RRT=2,577) ($p<0.001$) in those who did not. Such poor outcome in AKI admissions was more prevalent in emergency than planned admissions (25.1% (30,726) vs 10.6% (5,106) respectively, $p<1 \times 10^{-5}$). Drrt risk was related to higher AKI_{max} and AKI_{first} stages, with Drrt rates ranging from 13.6% to 42.4% for AKI_{max} and 18.7% to 32.8% for AKI_{first} (Table 4.9).

The incidence of Drrt also varied with the algorithm route by which any given AKI stage was reached. Thus, Drrt occurred in 12.7% (95% CI: 12.5–12.9) of those whose stage 1 AKI_{max} was triggered by a 1.5-fold increase in creatinine over baseline, but 16.6% (CI: 16.2–17.1) when triggered by a rise of >26 mmol/L over 48 hours. For AKI 3, Drrt rates varied from 38.6% to 45.9% depending on the triggering route (Figure 4.4).

Grouping by AKI_{max} stage (and trigger route), admission route and CKD status results in 56 groups, further increasing the heterogeneity in Drrt rates (Figure 4.5). The absolute creatinine result as well as the Cr:BCr ratio also influences the Drrt rate. The complex pattern of this interaction is shown in the heatmap (Figure 4.7) which plots Drrt rates against $[Cr:BCr]_{max}$: here, Drrt rates for AKI_{max} 3 range from ~0.5% to ~70% ($p<0.01$). Thus, some stage 1 AKI_{max} cases may have a 100-fold higher risk of Drrt (50%: Figure 4.7, X) than some stage 3 AKI_{max} cases (0.5%: Figure 4.6, Y).

Prevalence of Drrt associated with AKI_{max} also varied with the presence of co-morbidities, being 11.5% in those with myeloid leukaemia and 36.9% in those with heart failure (Table 4.7).

Table 4.9: Drrt rates/positive predictive values

	Drrt	Total	Drrt rate (95% Confidence Interval)		Sensitivity	Specificity	Positive Predictive Value
MAX AKI Stage							
1	15,114	111,163	13.6 %	(13.4–13.8)	42.2	65.1	**1 in 7.4**
2	8,101	29,683	27.3 %	(26.8–27.8)	22.6	80.3	**1 in 3.7**
3	12,617	29,740	42.4 %	(41.9–43)	35.2	83.5	**1 in 2.4**
Combining 2&3	20,718	59,423	34.9 %	(34.4–35.4)	57.8	86.4	**1 in 2.9**
First AKI Stage							
1	24,528	131,324	18.7 %	(18.5–18.9)	68.5	71.2	**1 in 5.4**
2	5,493	21,537	25.5 %	(25–26.1)	15.3	79.6	**1 in 3.9**
3	5,811	17,725	32.8 %	(32.1–33.5)	16.2	80.4	**1 in 3.1**
Combining 2&3	11,304	39,262	28.8 %	(28.2–29.5)	31.5	81.3	**1 in 3.5**
ML-AKI Model							
ML25	na	na	25 %	na	99.6	21.3	**1 in 4**
ML33	na	na	33.3 %	na	95.3	49.4	**1 in 3**
ML50	na	na	50 %	na	74.3	80.3	**1 in 2**

Figure 4.4 AKI Died or Renal Replacement Therapy (Drrt) Incidence

(by trigger route of Max AKI Stage; n = 170,596 (D−RRT) = 35,832)

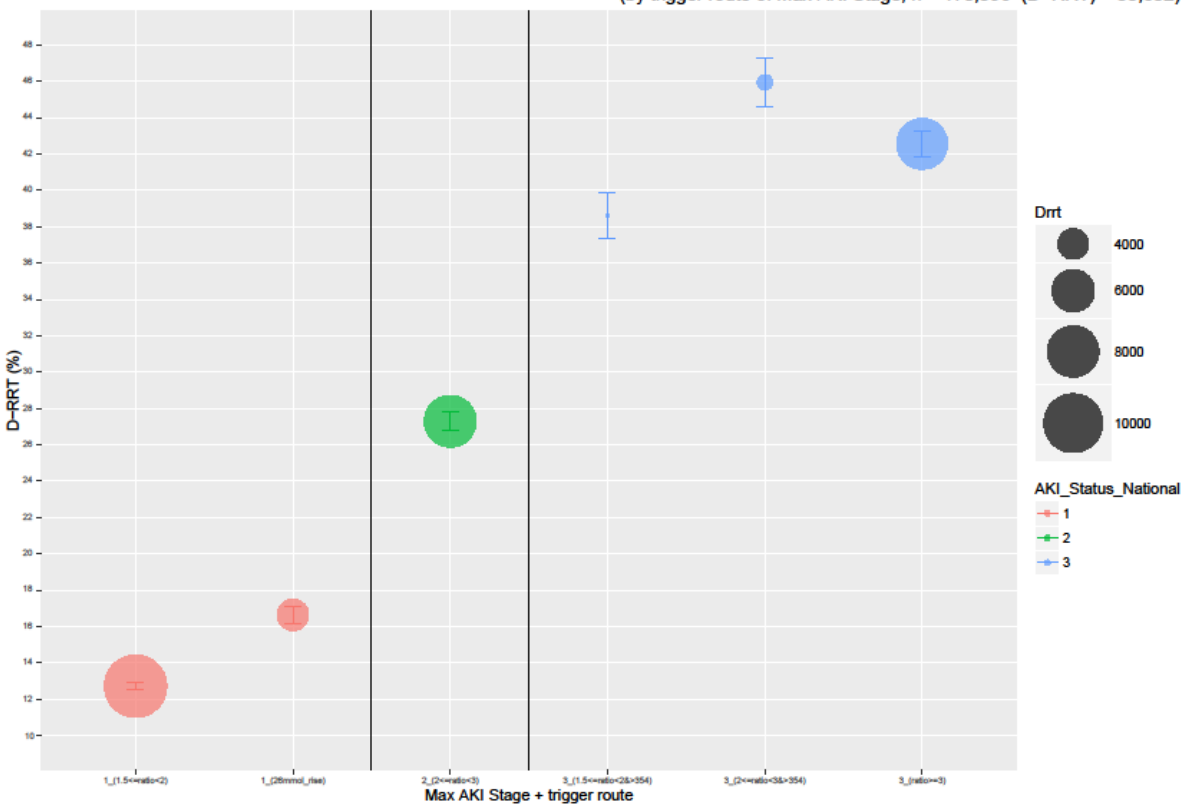


Figure 4.5: AKI Died or Renal Replacement Therapy (Drrt) Incidence for each CKD group

(by trigger route of Max AKI Stage; n = 170,596 (D−RRT) = 35,832)

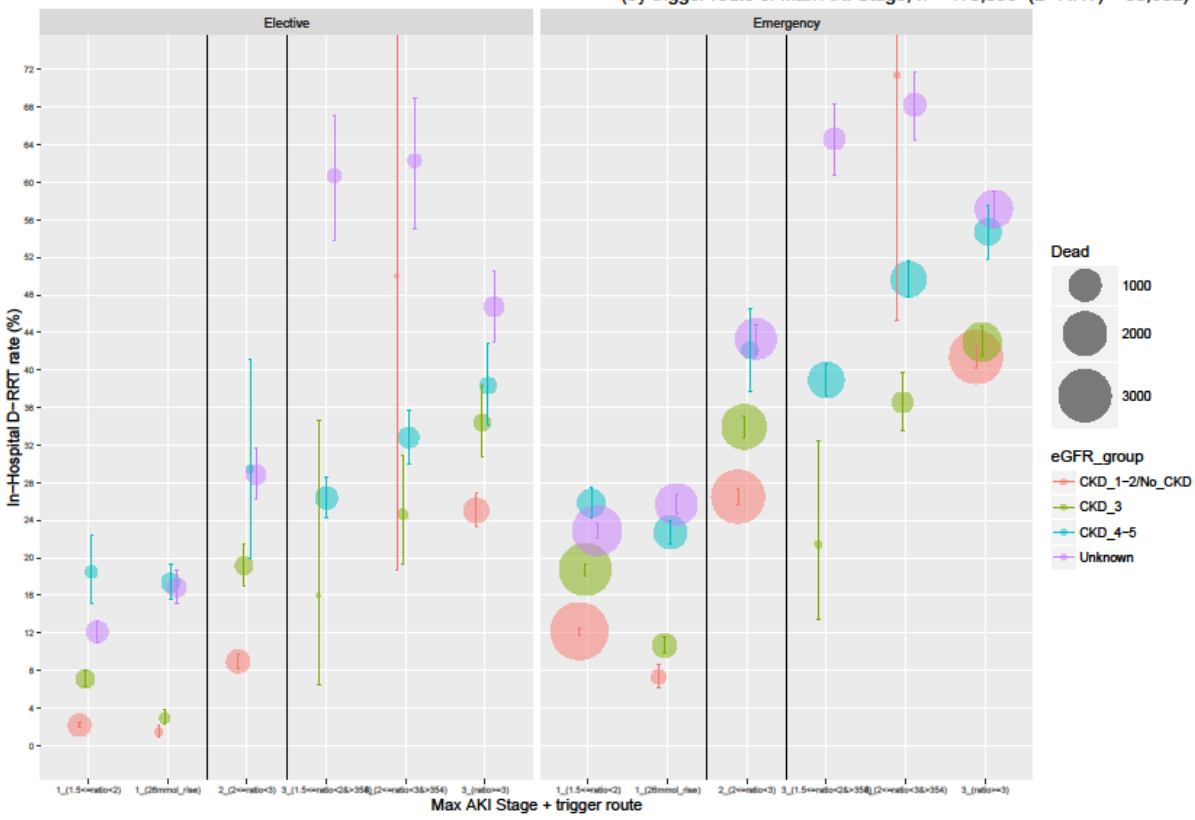
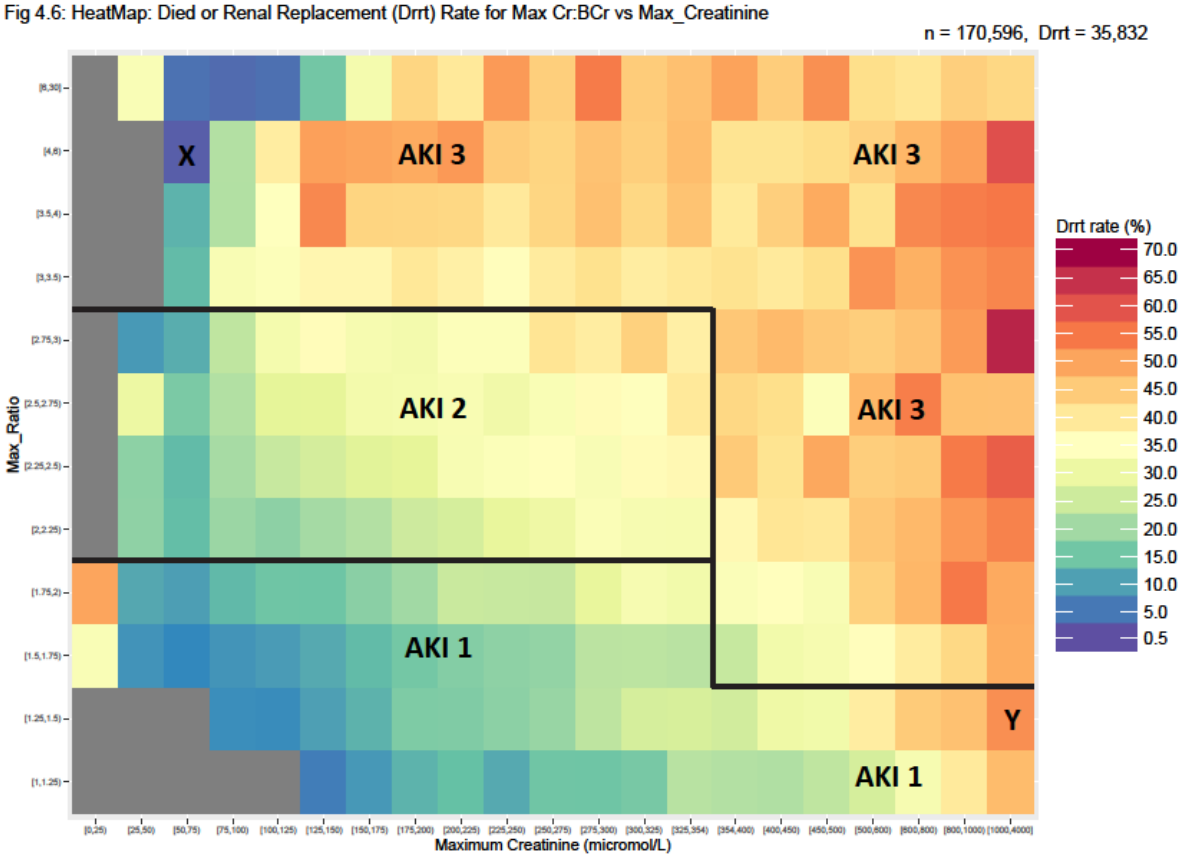


Fig 4.6: HeatMap: Died or Renal Replacement (Drrt) Rate for Max Cr:BCr vs Max_Creatinine



4.4.5 ML Model: ML-AKI

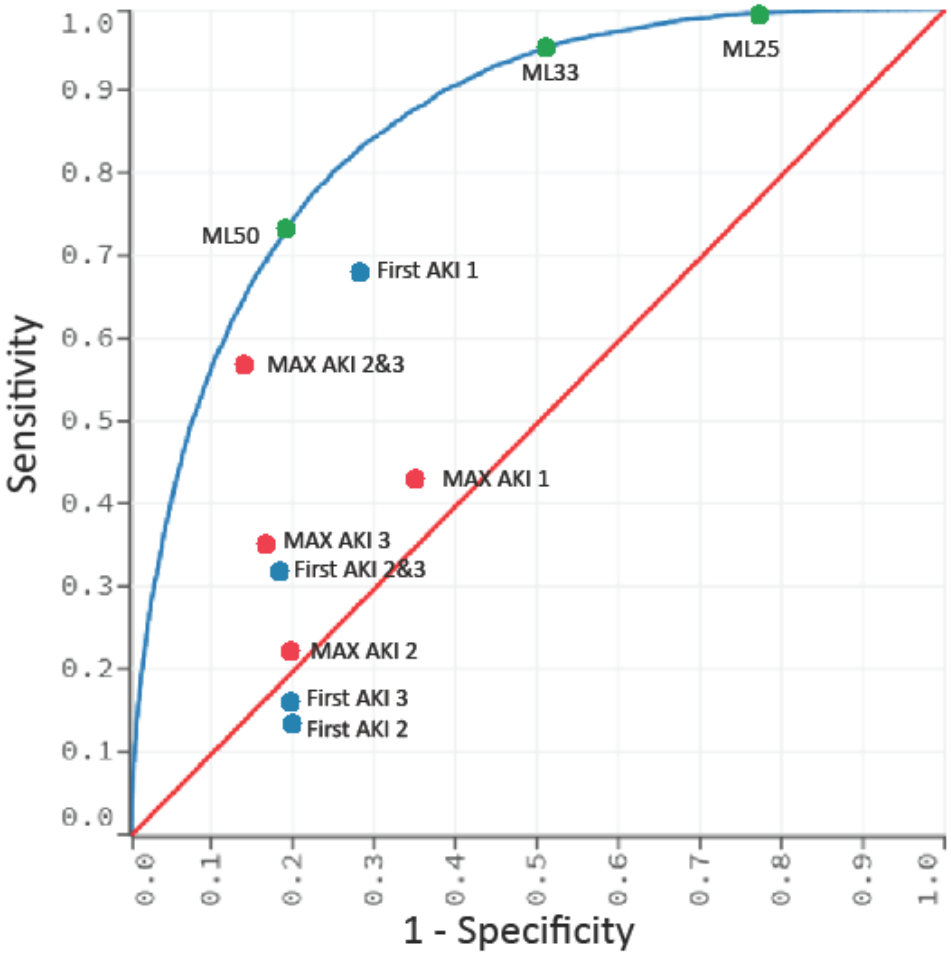
Increased AKI severity stage is a poor indicator of individual risk of death or renal replacement therapy in hospital (Drrt), due to the substantial heterogeneity within patient populations, and the complex interaction of these heterogeneous variables. Modelling all of these, the ML-AKI model achieved an AUROC for Drrt of 85.9% (with Drrt was coded as 0/1: log loss: 0.36, mean squared error: 0.11, and mean per class error: 0.24) on the test dataset. The training dataset and the cross-validation AUROCs were 96% and 85.2% respectively. The sensitivities, specificities and PPVs of the ML-model and NHSE-AKI algorithm are shown in Figure 4.7 and Table 4.9. The final hyper-parameters of the model were; number of trees: 3000, max depth: 7, maximum leaves: 90, and learn rate: 0.02.

Overall, predictive performance of the ML-AKI model exceeded that of the NHSE-AKI algorithm. The rate of Drrt for ML33 (33.3%) and ML50 (50%) was higher than for any AKI_{first} stage (18.7 to 32.8%). Even when merging AKI stages 2 and 3 together, their performance fell short of both ML33 and ML50 for Drrt (AKI_{first} 2&3 = 28.8%) and sensitivity (AKI_{first} 2&3 = 31.5%, ML33 = 95.3%, ML50 = 74.3%).

The application of the ML-AKI model would better target escalation of care when compared with the existing stages of the AKI algorithm. If, at first AKI presentation, the ML50 threshold were used to

escalate care to specialist staff, the referral rate would be 312 per 1,000 admissions with AKI, of which 156 of those seen would suffer Drrt in the future. Were all AKI_{first} stage 2 and 3 admissions referred, 230 of every 1,000 AKI admissions would be seen, of which 66 would develop Drrt. Thus, 26.3% (82) fewer admissions would be seen than if escalated by ML50, but at the expense of missing 90 future Drrt cases (66 vs 156: 57.7% fewer than those seen with the ML50 trigger). ML50's performance is also superior when compared to AKI_{max} 2&3 stage (were it possible to recognise this a priori), with 13.1% (312 vs 359 per 1,000) lower referral rates and 24.8% (156 vs 125) more Drrt admissions seen.

Figure 4.7: ML-AKI Model: Area Under the Receiver Operator Curve 85.9%



The green dots (ML20, ML33, ML50) represent specific threshold values used to discriminate between patients likely to have Drrt or not; The red and blue dots are representations of the sensitivity and specificity of the AKI Stages of the existing NHSE-AKI algorithm.

For ML33, the referral rate would be 600 per 1,000 AKI admissions, of which 200 (95.6% of all Drrt) would suffer from Drrt. This would result in 366 (159%) more admissions being seen than for AKI_{first} stages 2 and 3, but 40% less than all AKI admissions (AKI stages 1, 2 and 3). However, the Drrt cases identified would be 134 (203%) more or 10 (4.7%) less than AKI_{first} stages 2 and 3 or all AKI (stages 1, 2

and 3) respectively. For ML25, the referral rate would be 836 per 1,000 AKI admissions, of which 209 (99.3% of all Drrt) would suffer from Drrt.

Finally, Table 4.10 lists the relative importance of the top thirty variables of the ML-AKI model in descending order, highlighting the influence of demographic and administrative data in an admission's AKI outcome.

Table 4.10: ML-AKI Variable Importance

Variable	Scaled Importance	Percentage importance
Urea at first AKI stage	1.000	21.65%
Admission method code	0.409	8.85%
Ethnic category	0.284	6.15%
White cell count at first AKI stage	0.211	4.57%
Age	0.197	4.26%
Sodium at first AKI stage	0.188	4.07%
Creatinine 48 hr rise at first AKI stage	0.155	3.35%
Source of admission code	0.154	3.34%
Platelets at first AKI stage	0.135	2.92%
AKI stage at first AKI stage	0.124	2.68%
Red cell distribution width at first AKI stage	0.123	2.67%
Month of admission	0.119	2.57%
Urea:creatinine ratio at first AKI stage	0.102	2.20%
I50 (ICD10 code)	0.088	1.90%
Admission estimated glomerular filtration rate	0.084	1.81%
Creatinine at first AKI stage	0.083	1.80%
Mean sodium	0.072	1.56%
Potassium at first AKI stage	0.065	1.41%
Haemoglobin first AKI stage	0.061	1.32%
Mean creatinine	0.060	1.30%
I48 (ICD10 code)	0.059	1.27%
Admission baseline creatinine	0.057	1.23%
Admission urea:creatinine ratio	0.047	1.02%
C79 (ICD10 code)	0.045	0.97%
Day of admission	0.033	0.72%
Admission sodium	0.033	0.71%
Admission red cell distribution width	0.030	0.66%
Admission haemoglobin	0.029	0.63%

Admission potassium	0.027	0.58%
Admission white cell count	0.027	0.57%

4.5 Discussion

I have performed the one of largest analysis yet undertaken relating to AKI incidence, outcomes and prediction of mortality and renal replacement therapy, having studied 1.9 million hospital admissions in 14 NHS acute trusts, using data for over 22.6 million creatinine results. I found AKI to be common, occurring in over 1 in 11 (8.6%) of hospital admissions, in line with data from China (7%) [141], but lower than both the pooled northern European (14.7%) [84] and worldwide rates (21%). [84] Incidence amongst emergency admissions (10.7%) was also lower than in recently reported studies (25.4%[142] and 12.3%[143]). The proportion of community-acquired AKI (AKI diagnosed on the first in-hospital creatinine result, i.e. on admission) in my study was 38.6%, with no firm comparison available. AKI patients in my study did, however, have a similar prevalence of co-morbidities to that reported in past studies (e.g. hypertension and diabetes mellitus, 34.2% vs 33.8%[144] and 19.7% vs 23%[144] respectively), and the proportion developing AKI who presented with a Ur:Cr ratio >80 was also similar (41% vs 39%[145]). The prevalence of heart failure in my study was, however, a little lower than in previous research (13% vs 18.3%[144]).

I confirmed that AKI is associated with poor outcome. Drrt occurred in a little over 1 in 5 of those with AKI (21%; death, 17.6%), but less than 1 in 62 (1.6%; death, 1.6%) of those without. Such findings are not dissimilar from those reported worldwide (AKI mortality = 23.3%).[84] Finally, I confirmed that a higher AKI stage was associated with poorer outcome: both Drrt (Table 4.8) and mortality (AKI 1: 12.3%; 2: 26.6%; 3: 28.4%) rose with rising stages.

Over a decade of research culminating in the KDIGO guidelines, and the potential for improving care of patients developing AKI led to NHS England to developing and subsequently mandating the NHSE-AKI algorithm. This is without a doubt a positive incremental improvement in the management of patients with renal injury. However, I have demonstrated there is room for a further significant step change improvement, as there are weaknesses in the KDIGO-based NHSE-AKI algorithm. Outcome is dependent not just on the AKI stage, but also on the route within the algorithm by which it is determined, resulting in significant variability of Drrt risk. More importantly, Drrt is affected by multiple factors, including patient demographics, disease states, and method and dates of admission. None of these are incorporated into either the NHSE-AKI algorithm or the original KDIGO guidelines from which the former is derived. Overall, the current AKI staging results in poorly stratified individual risk.

Consistent with these observations, I found that the performance of my ML-AKI model greatly exceeded that of the NHSE-AKI algorithm in predicting Drrt. Indeed, my ML-AKI model achieved an AUROC of 85.9%. Using the ML50 threshold (1 in 2 cases predicted to die, or need RRT, did so), I was able to identify 74.3% of all AKI admissions who suffered such outcomes (sensitivity). If this threshold were used to direct escalation of care, as opposed to the existing KDIGO AKI staging, only 31.2% of admissions on first presentation with AKI would trigger the efferent limb (rapid response team or dedicated AKI nephrologist), while capturing almost three out of every four patients likely to suffer from Drrt. The positive predictive value of 1 in 2, means that for every two patients escalated, one would suffer from Drrt subsequently in-hospital. The ML50 Drrt rate (50%) is higher than that of any of the existing KDIGO AKI Stages at AKI_{first} 1: 18.7%, AKI_{first} 2: 25.5%, and AKI_{first} 3: 32.8% (i.e. higher PPV and specificity), while at the same time identifying 1.1, 4.9, and 4.6 as many patients likely to suffer from Drrt (higher sensitivity), compared to the KDIGO AKI Stages, respectively.

My study does have some limitations. A small number of admissions received renal replacement therapy without AKI having been diagnosed by the algorithm (2,577). Of these, 2,019 actually had existing CKD-4/5, and may thus already have been dependent on renal replacement therapy. This suggests a possible failure of appropriate in-hospital diagnostic coding. Alternatively, acute on chronic kidney injury associated with a <1.5-fold increase in creatinine from the baseline (and thus no AKI diagnosis) may yet have led to renal replacement therapy. However, as these patients were not classified as having AKI, they were excluded from the AKI analysis, regardless of whether the renal replacement therapy was planned or due to undiagnosed AKI (which cannot be discriminated between from my dataset). It is also worthy of note that patients transferred to another hospital were excluded, which would not have been known a priori. Nor were all 'detrimental outcomes' (length of stay, hospital readmission or development of CKD) modelled. Future studies might also expand my definition of 'false-negative-AKI' to include cases where creatinine did fall (leading to the lack of an AKI alert), but incompletely and without clinically significant resolution; such studies might also incorporate complete resolution of AKI (return to baseline values). Finally, this is a retrospective study.

A major strength of my study is the use of the AKI_{first} stage. My ML-AKI model also only utilised data up to this point. I chose this for being both a practical measurement (as AKI_{max} is only known post hoc) and probably the most effective, reflecting when early intervention might stand the greatest chance of improving the subsequent clinical course. Another strength is that the data were collected before the NHSE-AKI algorithm was implemented, and thus the Drrt rates were not confounded by NHSE-AKI algorithm-related detection/interventions.

The 'importance' of non-creatinine-related variables in the ML-AKI model (a 48 hr change in creatinine is seventh in relative importance) does highlight that it is misleading to focus on just the creatinine result to determine AKI stage, as a proxy risk of poor outcome. Other blood results – urea, white cell count and sodium – all make higher relative contributions to the ML-AKI model, as do the variables of ethnicity, month and method of admission (Table 4.10). Further work is needed to understand these associative links and to determine causality. Nonetheless, these a priori data are used by clinicians in their diagnosis/treatment of a patient, and should therefore be incorporated into any risk prediction model for AKI-diagnosed patients.

Finally, it is of great concern that 32.2% of all AKI admissions who do not die or have renal replacement therapy in hospital are discharged from hospital when still demonstrating an AKI stage. Future studies should further investigate and follow up these patients to deem the appropriateness of this course of action (discharge from hospital with an AKI stage).

In summary, I have shown that:

1. AKI stage is poor indicator of individual risk of Drrt.
2. Method of admission, demographics, other blood results and co-morbidities significantly modify Drrt risk.
3. ML applied to common blood tests and administrative data can offer individual risk prediction with a precision well beyond that delivered by existing/conventional risk stratification.
4. Variables other than creatinine have more influence on the risk of Drrt in AKI admissions than creatinine itself.

My ML-AKI model (implemented in Java code) can be integrated with existing hospital systems, and would immediately target clinical resources to patients at the highest risk of Drrt, on their first presentation with AKI. This would enable early diagnosis and intervention, which has been shown to increase clinical effectiveness, and which may also reduce costs associated with health care.

Chapter 5: ML EWSCs

5.1 Introduction

Patients may die in hospital from a variety of conditions. They may not be dehydrated, nor may they develop AKI. They also arrive in hospital via multiple methods (emergency, elective, maternity, etc.); thus, an ideal EWSC would be applicable to all hospitalised patients, regardless of their method of admission or disease state. Preferably, such an EWSC would be able to risk-stratify patients as early in their hospitalisation as possible, i.e. on admission. In addition, the capability to provide an updated prediction of in-hospital mortality once initial treatment has taken place would also help target resources to those patients who are either at continued risk or have deteriorated since admission. As discussed in Chapter 1 (Section 1.4.4), there have been attempts at building such laboratory-based EWSCs in the past. However, none of these previous attempts can be applied to all hospitalised patients, due to limitations which include:

a) Poor generalisability: their effectiveness has not been proven in diverse hospital patients, as they are either specific to a type of patient (medical, surgical, emergency, elective, etc.), or they have been created and tested using patients from only one or two hospitals.
b) Simplistic analysis: they take account of neither the dynamic change (trends), nor the inter-correlations between blood results.
c) Lack of patient specific information: they do not account for a patient's co-morbidities.
d) Lack of context sensitive information: they account for neither the date of a patient's admission (e.g. weekday vs weekend, summer vs winter) nor the time a blood test was performed (e.g. 3pm vs 3am).
e) Poor performance: their ability to discriminate between patients likely to live or die is low.

To overcome the limitations of these previous approaches in this thesis, it was necessary to 1) collect a large dataset from multiple diverse hospitals, and 2) to apply advanced analytical techniques that both accounted for multicollinearity and were able to produce highly accurate models.

5.2 Objectives

To build advanced EWSCs (hereafter referred to as a 'model' that 1) use a wide range of existing available information, 2) incorporate the dynamic change in the input variables, and 3) can be utilised on all patients admitted to hospital.

In this chapter, I have focussed on creating two such ML models. Both can be applied to all hospitalised patients. The first model, 'ML_Admission', to be applied at the point of a patient's admission, predicts the risk of in-hospital mortality using the very first set of blood test results. The second model, 'ML-Two-Tests', provides a prediction of a patient's in-hospital mortality at the moment when a second set of blood test results is available; i.e. at a later time point during a patient's stay in hospital, during which it is expected that some treatment has been initiated.

The input variables of the models consist of:

- A subset of the most common blood tests performed on hospitalised patients: Urea and Electrolytes (sodium, potassium, urea, and creatinine), Full Blood Count (haemoglobin, white cell count, platelets, red cell distribution width, and mean cell volume), and Albumin;
- Known a priori information about the patient:
    - method and source of admission
    - co-morbidities
    - demographics (age, sex, and ethnicity)

5.3 Methods

5.3.1 Patients

The dataset described in Chapter 2 was used as the basis for this analysis. In brief, I collated continuous electronically stored data for all adults (18 years and over) admitted to fourteen UK NHS acute hospital trusts from early 2005 to late 2015. Specifically, for this analysis, I included patients with a complete administrative record, and at least one sodium and haemoglobin blood test result, taken during their admission. Excluded were patients transferred to another hospital for continued medical care and those who self-discharged from hospital, for whom the outcome of their hospitalisation was unknown. Each hospital admission was analysed independently.

5.3.2 Data

The following results were collated from each admission: the first and second sodium (mmol/L), potassium (mmol/L), urea (mmol/L), creatinine (µmol/L), haemoglobin (g/L), platelet (10^9/L), mean cell volume (fL), red cell distribution width (%), white cell count (10^9/L) results, and albumin (g/L). In addition, where such data existed, they were captured up to one year prior to admission. Routinely collected administrative data were also recorded: patient age, diagnoses (ICD-10), treatment codes (OPCS4) and dates, and details relating to hospital admission and discharge. I selected these variables because of their ubiquity (available for most hospitalised patients).

I performed feature engineering on this raw data, to extract:

- For blood results tested during an admission:
  - The hour of the day each blood test result became available (24-hour clock).
  - The absolute difference between the second and first blood results, for each test per admission.
  - The relative difference (second blood test / first blood test), for each test per admission.
- For blood results in the year prior to a patient's admission:
  - The total number of each test performed, for each test per admission.
  - The mean of each blood test, for each test per admission.
  - The standard deviation of each blood test, for each test per admission.
- For administrative data:
  - The day of the week of admission (Monday to Sunday).
  - The month of admission (January to December).

For ML-Admission, this entire dataset was used. For ML-Two-Tests, only admissions with both an additional in-hospital sodium and haemoglobin result were included.

Prevalence and in-hospital mortality were calculated for: each ethnic category, admission method, day of week of admission, month of admission, and each ICD-10 code. I also calculated the mean and interquartile range, and the number of missing values for each blood test.

5.3.3 ML

I used Wilson's method to calculate the 95% confidence intervals for mortality rate. Prevalence and mortality rates were compared using Fisher's Exact Test.

I used the GB ML technique to build the models to predict in-hospital mortality. The two models built were:

1. ML_Admission: to be used on all patients who have at least **one** sodium and **one** haemoglobin blood result in-hospital, at the point their first set of blood results is available.
2. ML_Two_Tests: to be used on all patients who have at least **two** sodium and **two** haemoglobin in-hospital blood results, at the point their second set of blood results is available.

The full list of variables used to create these two models is shown in Tables 5.1 and 5.2 respectively.

Table 5.1: ML-Admission: Variables		
Variable Name		**Type of Variable**
Age		Numerical
Sex		Categorical
Ethnic category		
Admission method code		
Source of admission code		
Day of admission		
Month of admission		
Blood Results		
Admission Sodium		Numerical
Admission Potassium		
Admission Urea		
Admission Creatinine		
Admission Urea:Creatinine Ratio		
Admission Haemoglobin		
Admission Platelets		
Admission White Cell Count		
Admission Mean Cell Volume		
Admission Red Cell Distribution Width		
Admission Albumin		
Hour of Validation of the Admission Blood Result		
Hour of validation of admission	Sodium Potassium Urea Creatinine Haemoglobin Platelets White Cell Count Mean Cell Volume Red Cell Distribution Width Albumin	Numerical (0 – 24)
Summary Pre-Admission Blood Results (last year (365 days), SI units)		
Number of	Sodium Potassium Urea Creatinine Haemoglobin	Numerical
Mean		Numerical

Standard deviation of	Platelets White Cell Count Mean Cell Volume Red Cell Distribution Width Albumin	Numerical
Co-morbidities		
Condition	**ICD10 code**	
Essential (primary) hypertension	I10	Categorical
Angina Pectoris	I20	
Chronic ischaemic heart disease	I25	
Atrial fibrillation or flutter	I48	
Heart Failure	I50	
Secondary and unspecified malignant neoplasm of lymph nodes	C77	
Secondary malignant neoplasm of other and unspecified sites	C79	
Multiple myeloma and malignant plasma cell neoplasms	C90	
Myeloid leukaemia	C92	
Other hypothyroidism	E03	
Type 2 diabetes mellitus	E11	
Obesity	E66	
Disorders of lipoprotein metabolism and other lipidaemias	E78	
Mental and behavioural disorders due to use of alcohol	F10	
Smoking	F17	
Depressive episode	F32	
Epilepsy	G40	
Other chronic obstructive pulmonary disease	J44	
Asthma	J45	
Chronic kidney disease	N18	
Other symptoms and signs involving the nervous and musculoskeletal systems	R29	
Personal history of malignant neoplasm of digestive organs	Z850	
Personal history of psychoactive substance abuse	Z864	
Personal history of diseases of the nervous system and sense organs	Z866	
Personal history of diseases of the circulatory system	Z867	
Personal history of allergy to penicillin	Z880	
Personal history of allergy to other drugs, medicaments and	Z888	

biological substances		
Presence of electronic cardiac devices	Z950	
Presence of aortocoronary bypass graft	Z951	
Presence of coronary angioplasty implant and graft	Z955	

For each model I split the appropriate dataset into a seventy per cent 'Train Dataset' and a thirty per cent 'Test Dataset'. Five-fold cross-validation was employed to prevent overfitting and to help with final model selection. I built multiple models for each use case (ML-Admission and ML-Two-Tests) by modifying the hyper-parameters of the GB model. The hyper-parameters in this case were 1) learning rate, 2) depth of trees, 3) number of trees and 4) nodesize. A grid search of these hyper-parameters was used to determine the best model, based on its mean cross-validation performance (i.e. the model with the smallest mean cross-validation log loss). The range of the hyper-parameters for the grid search were: learning rate: 0.1 to 0.01; depth of trees: 5 to 9; number of trees: 800 to 5,000; and nodesize: 500 to 12,000. The selected model was then used to predict outcome on the independent Test Dataset to determine the AUROC. Two threshold values were selected that resulted in positive predictive values (PPV) of 1:5 and 1:3, for identifying whether a patient had died in hospital or not. A PPV of 1:5 means that if the model predicts a patient is likely to die, then one out of every five predicted patients will do so (20%). Similarly, a PPV of 1:3 means that if the model predicts a patient is likely to die, then one out of every three predicted patients will do so (33.33%). For each PPV threshold, the sensitivity and specificity were also calculated.

The relative importance of each predictor variable was calculated as in Chapter 4 (Section 4.3.3). The same software as in Chapter 4 (Section 4.3.3) was also used.

Table 5.2: ML-Two-Tests: Variables

Variable Name	**Type of Variable**
All Variables from Table 5.1	Mixed
Blood Results	
Second Sodium	Numerical
Second Potassium	
Second Urea	
Second Creatinine	
Second Urea:Creatinine Ratio	
Second Haemoglobin	

Second Platelets		
Second White Cell Count		
Second Mean Cell Volume		
Second Red Cell Distribution Width		
Second Albumin		
Hour of Validation of Blood Result		
Hour of validation of second blood result:	Sodium Potassium Urea Creatinine Haemoglobin Platelets White Cell Count Mean Cell Volume Red Cell Distribution Width Albumin	Numerical (0–24)
Difference between Second and Admission Blood Result		
Absolute difference between Second and Admission: (second blood result / admission blood result)	Sodium Potassium Urea Creatinine Ur:Cr Haemoglobin Platelets White Cell Count Mean Cell Volume Red Cell Distribution Width Albumin	Numerical
Relative difference between Second and Admission: (second blood result / admission blood result)		Numerical
Number of hours' difference between Second and Admission blood result:		Numerical

5.4 Results

5.4.1 Admission_Cohort

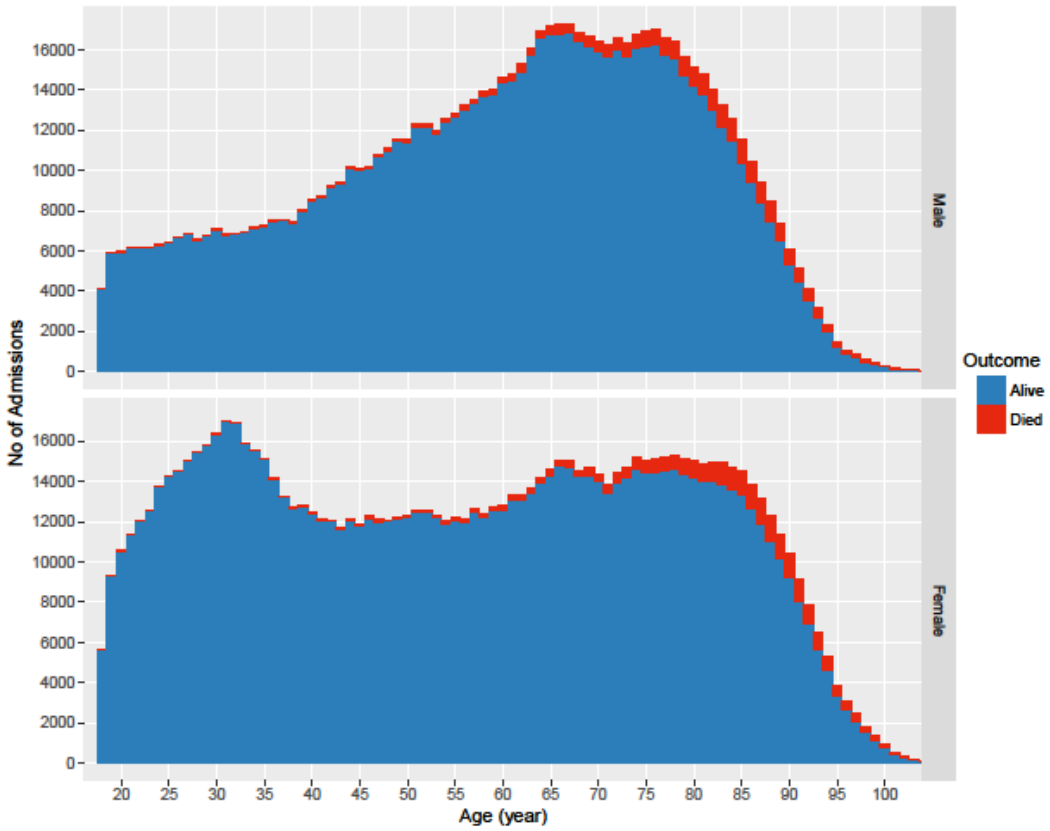
5.4.1.1 Baseline Characteristics

There were 1,874,325 admissions who met the inclusion criteria. Of these, 3.14 % (58,843) died in hospital. The median length of stay (LoS) was 2 days (IQR): 0–6 days). Female admissions were 54.86% of the total. The median age was 60 years (IQR 41–75). The histogram of the ages of admissions, along with the number of patients who died, is shown in Figure 5.1.

Grouped emergency, elective and maternity admissions accounted for 57.6%, 34.6% and 5.7% of the total respectively (Table 2.3). The full distribution of the admissions by ethnic category, method, month, and day of week are shown in Tables 5.3, 5.4, 5.5 and 5.6 respectively. The ethnic category 'British' had the highest proportion of admissions at 42.8%, followed by the 'Not Stated' category at

30.4%. The prevalence of the various ICD-10 codes is shown in Table 5.7. The median (IQR) values of all the blood tests, including the number of admissions with missing values, are shown in Table 5.8. Of particular note are the proportions of admissions with no results for platelets, RDW or albumin, at 28.6%, 29.4% and 21.08% respectively. The presence of both a sodium and a haemoglobin result were part of the inclusion criteria, hence the presence of these in all admissions.

Figure 5.1: Histogram of Age



The in-hospital mortality rates for each group of admissions are also shown in Tables 5.3 to 5.6. The highest in-hospital mortality occurred in the 'Not Stated' and 'British' ethnic categories, at 4.1% and 3.4% respectively. Both of these were significantly higher than for all other ethnicities. Mortality by method of admission (Emergency: 4.9%; Elective: 0.4%; and Maternity: 0.07%; $p<0.01$) substantiates the view that patients admitted as an emergency are significantly more likely to die than those admitted via either the elective or maternity methods. The 'Other' method of admission (the bulk of which was made up of non-emergency inter-hospital transfers) has a small prevalence (2.1%), but the highest mortality rate of all, at 9%. The months of admission with the highest mortality were January (3.5%) and December (3.9%). Weekend admissions had higher mortality when compared with weekdays (4.3% vs 2.9%, $p<0.01$). However, this weekend vs weekday difference in mortality rates decreased from 1.4% to 0.31% when only emergency admissions were examined (5.13% vs 4.82%).

Table 5.3: Admission_Cohort: Ethnic category prevalence, and in-hospital mortality rates			
Ethnic Category	**Number of Admissions**	**Prevalence % of total Total = 1,874,325**	**In-Hospital Mortality (95% CI)**
British (A)	802,570	42.8 %	3.37 % (3.33–3.41)
Irish (B)	31,973	1.71 %	2.90 % (2.72–3.09)
Any other white background (C)	139,363	7.44 %	1.84 % (1.77–1.91)
White and black Caribbean (D)	4,329	0.23 %	0.97 % (0.72–1.31)
White and black African (E)	2,753	0.15 %	1.09 % (0.76–1.55)
White and Asian (F)	2,588	0.14 %	1.31 % (0.94–1.83)
Any other mixed background (G)	7,944	0.42 %	1.27 % (1.05–1.54)
Indian (H)	41,824	2.23 %	1.96 % (1.83–2.1)
Pakistani (J)	17,658	0.94 %	1.39 % (1.23–1.58)
Bangladeshi (K)	27,899	1.49 %	1.47 % (1.34–1.62)
Any other Asian background (L)	39,093	2.09 %	1.57 % (1.45–1.7)
Caribbean (M)	43,785	2.34 %	1.77 % (1.65–1.9)
African (N)	55,177	2.94%	0.85 % (0.78–0.93)
Any other black background (P)	21,603	1.15%	1.11 % (0.98–1.26)
Chinese (R)	8,864	0.47%	1.83 % (1.57–2.13)
Any other ethnic group (S)	58,089	3.10%	1.97 % (1.86–2.09)
Not Stated (Z)	568,813	30.35%	4.08 % (4.03–4.13)

Table 5.4: Admission_Cohort: Method of admission prevalence, and in-hospital mortality rates

Method of Admission		Number of Admissions	Prevalence: % of total Total = 1,874,325	In-Hospital Mortality (95% CI)
Elective	Elective Waiting list (11)	219,310	11.70 %	0.54 % (0.51–0.57)
	Elective Booked (12)	97,146	5.18 %	0.79 % (0.74–0.85)
	Elective Planned (13)	333,124	17.77 %	0.20 % (0.19–0.22)
Emergency	Accident and emergency or dental casualty department of the health care provider (21)	762,307	40.67 %	5.03 % (4.98–5.08)
	General practitioner (22)	163,152	8.70 %	5.15 % (5.05–5.26)
	Bed bureau (23)	8,814	0.47 %	3.8 % (3.42–4.22)
	Consultant clinic, of this or another health care provider (24)	33,385	1.78 %	2.83 % (2.66–3.02)
	Admission via mental health crisis resolution team (25)	474	0.03 %	5.27 % (3.6–7.67)
	Transfer of an admitted patient from another hospital provider in an emergency (28)	107,525	5.74 %	4.19 % (4.07–4.31)
	Accident and emergency department of another provider where the patient had not been admitted (2A)	149	0.01 %	10.07 % (6.2–15.95)
	Other means (2B)	385	0.02 %	12.47 % (9.53–16.14)
	Other emergency admission (2D)	2,498	0.13 %	0.76 % (0.49–1.18)
Maternity	Admitted ante-partum (31)	101,982	5.44 %	0.01 % (0.01–0.02)
	Admitted post-partum (32)	4,317	0.23 %	0.05 % (0.01–0.17)
Other	Transfer of any admitted patient from another hospital provider other than in an emergency (81)	39,262	2.09 %	9.07 % (8.79–9.36)
	The birth of a baby in this health care provider (82)	140	0.01 %	1.43 % (0.39–5.06)
	Baby born outside the health care provider except when born at home as intended (83)	30	0.00 %	6.67 % (1.85–21.32)
	Not known / validation error (99)	325	0.02 %	4.92 % (3.05–7.85)

Table 5.5: Admission_Cohort: Month of admission prevalence, and in-hospital mortality rates			
Month	**Number of Admissions**	**Prevalence: % of total Total = 1,874,325**	**In-Hospital Mortality (95% CI)**
January	163,785	8.74 %	3.54 % (3.46–3.63)
February	150,886	8.05 %	3.35 % (3.26–3.44)
March	157,378	8.4 %	3.29 % (3.2–3.38)
April	154,892	8.26 %	3.14 % (3.05–3.23)
May	160,491	8.56 %	2.85 % (2.77–2.93)
June	149,226	7.96 %	2.78 % (2.69–2.86)
July	155,801	8.31 %	2.84 % (2.75–2.92)
August	150,825	8.05 %	2.91 % (2.83–3)
September	153,985	8.22 %	2.81 % (2.73–2.89)
October	162,718	8.68 %	3.1 % (3.02–3.18)
November	158,649	8.46 %	3.18 % (3.1–3.27)
December	155,689	8.31 %	3.86 % (3.76–3.95)

Table 5.6: Admission_Cohort: Day of admission prevalence, and in-hospital mortality rates				
Day of Admission		**Number of Admissions**	**Prevalence: % of total Total = 1,874,325**	**In-Hospital Mortality** *(95% CI)*
Weekday	Monday	307,401	16.4 %	2.97 % (2.91–3.03)
	Tuesday	326,864	17.44 %	2.83 % (2.77–2.88)
	Wednesday	323,169	17.24 %	2.74 % (2.69–2.8)
	Thursday	320,625	17.11 %	2.78 % (2.72–2.84)
	Friday	281,399	15.01 %	3.24 % (3.17–3.3)
Weekend	Saturday	155,174	8.28 %	4.43 % (4.33–4.53)
	Sunday	159,693	8.52 %	4.2 % (4.11–4.3)

Table 5.7: Admission_Cohort: Prevalence of ICD10 codes

Co-morbidity	ICD10 code	No. of admissions	Prevalence: % of total (1,874,325)
Essential (primary) hypertension	I10	431,609	23.03 %
Angina Pectoris	I20	60,181	3.21 %
Chronic ischaemic heart disease	I25	174,610	9.32 %
Atrial fibrillation or flutter	I48	151,808	8.1 %
Heart Failure	I50	79,226	4.23 %
Personal history of diseases of the circulatory system	Z867	111,968	5.97 %
Secondary malignant neoplasm of lymph nodes	C77	25,351	1.35 %
Secondary malignant neoplasm of other and unspecified sites	C79	37,953	2.02 %
Multiple myeloma and malignant plasma cell neoplasms	C90	37,608	2.01 %
Myeloid leukaemia	C92	25,511	1.36 %
Other hypothyroidism	E03	70,281	3.75 %
Type 2 diabetes mellitus	E11	199,260	10.63 %
Obesity	E66	32,739	1.75 %
Disorders of lipoprotein metabolism and other lipidaemias	E78	155,255	8.28 %
Mental and behavioural disorders due to use of alcohol	F10	43,076	2.30 %
Smoking	F17	94,122	5.02 %
Depressive episode	F32	66,164	3.53 %
Epilepsy	G40	35,559	1.90 %
Other chronic obstructive pulmonary disease	J44	99,895	5.33 %
Asthma	J45	113,837	6.07 %
Chronic kidney disease	N18	117,982	6.29 %
Other symptoms and signs involving the nervous and musculoskeletal systems	R29	38,476	2.05 %
Personal history of malignant neoplasm of digestive organs	Z850	21,552	1.15 %
Personal history of psychoactive substance abuse	Z864	119,672	6.38 %
Personal history of diseases of the nervous system and sense organs	Z866	24,251	1.29 %
Personal history of allergy to penicillin	Z880	90,515	4.83 %
Personal history of allergy to other drugs, and biological substances	Z888	32,564	1.74 %
Presence of electronic cardiac devices	Z950	34,612	1.85 %
Presence of aortocoronary bypass graft	Z951	38,428	2.05 %
Presence of coronary angioplasty implant and graft	Z955	31,323	1.67 %

Table 5.8: Admission_Cohort: Median (Interquartile Range) of Blood Tests		
Blood Test	**Median (Interquartile Range)**	**Percentage missing of total (total = 1,872,345)**
Sodium	139 (136–141)	0 %
Potassium	4.2 (3.9–4.5)	3.26 %
Urea	5.3 (3.9–7.6)	3.64 %
Creatinine	77 (62–100)	8.88 %
Ur:Cr	66.2 (51.2–84.7)	9.88 %
Haemoglobin	12.4 (10.9–13.8)	0 %
Platelets	236 (185–295)	28.62 %
White Cell Count	8.2 (6.1–11)	0.2 %
Mean Cell Volume	89.2 (85.1–93.3)	0.13 %
Red Cell Distribution Width	13.9 (13.1–15.4)	29.44 %
Albumin	40 (35–44)	21.08 %

5.4.1.2 ML-Admission Model

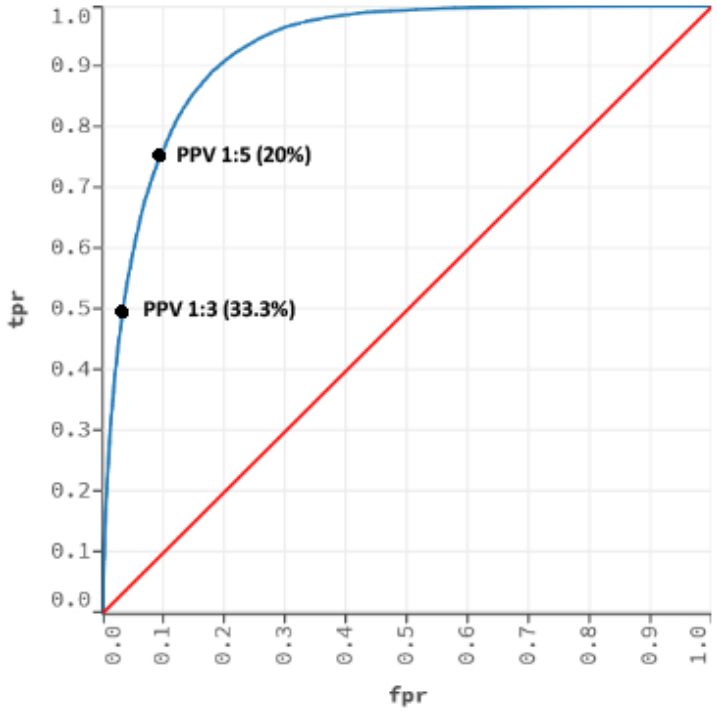
The Train Dataset and Test Dataset were randomised effectively, with both having the same proportion of patients who died, at 3.14%. The hyper-parameters that produced the lowest mean logloss and highest mean AUROC via cross-validation were: number of trees: 1,600; tree depth: 7; and learning rate: 0.04.

The ML-Admission model achieved an AUROC of 93% (logloss: 0.088, Figure 5.2) on the Test Dataset, which was similar to the mean cross-validation AUROC of 92.9% (logloss: 0.088). At the PPV 1:5 (20%) threshold, the model achieved a sensitivity of 76.1%, and a specificity of 90.2% (the highest ever achieved). This means that, if a rapid response team were directed to all admissions who triggered the model at this threshold, only 11.9% of admissions would be referred, but these referrals would include 76.1% of all patients likely to die. In other words, for every 1,000 admissions to hospital, there would be approximately 31 (31.4) admissions who would die in hospital. The referral rate of the model in this scenario would be 119 per 1,000 admissions, within which would be 24 of the 31 patients likely to die.

Similarly, for the PPV 1:3 (33%) threshold, the model's sensitivity was 49.3%, with a specificity of 96.8%. For every 1,000 admissions to hospital, the referral rate would be approximately 46 per 1,000 (4.6%) admissions, within which would be 15 of the 31 patients likely to die.

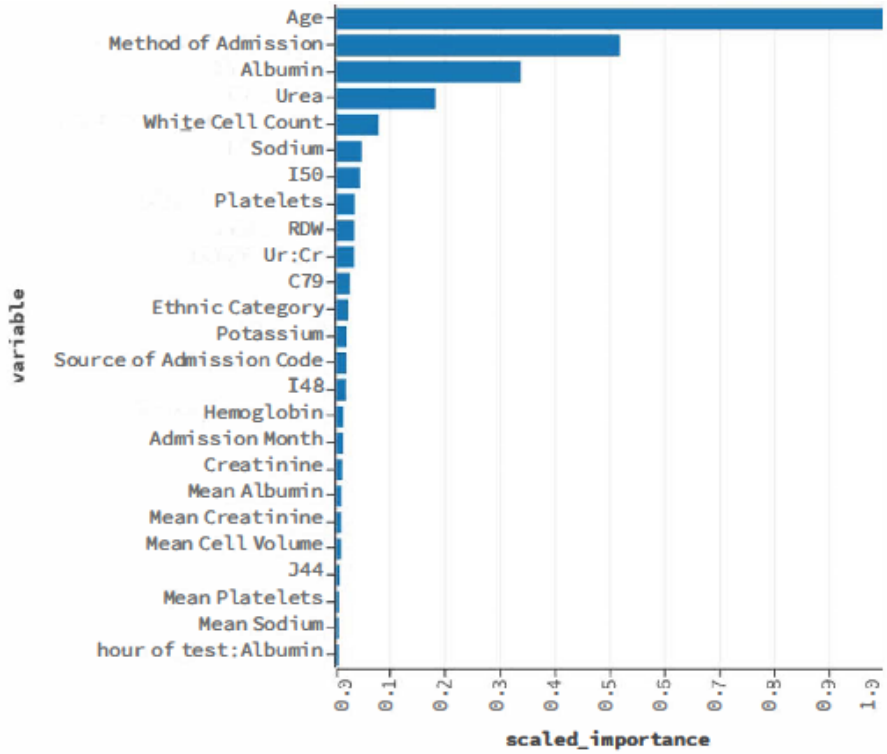
The scaled variable importance of the twenty-five most 'important' predictor variables is displayed in Figure 5.3. The four most important variables were Age, Method of Admission, Albumin and Urea.

Figure 5.2: ML- Admission: Receiver Operator Curve (AUROC 93%)



tpr (true positive rate): sensitivity; fpr (false positive rate): 1 – specificity; PPV 1:5 : positive predictive value of 1 in 5 threshold; PPV 1:3 : positive predictive value of 1 in 3 threshold

Figure 5.3: ML-Admission: Scaled variable importance for the top 25 variables

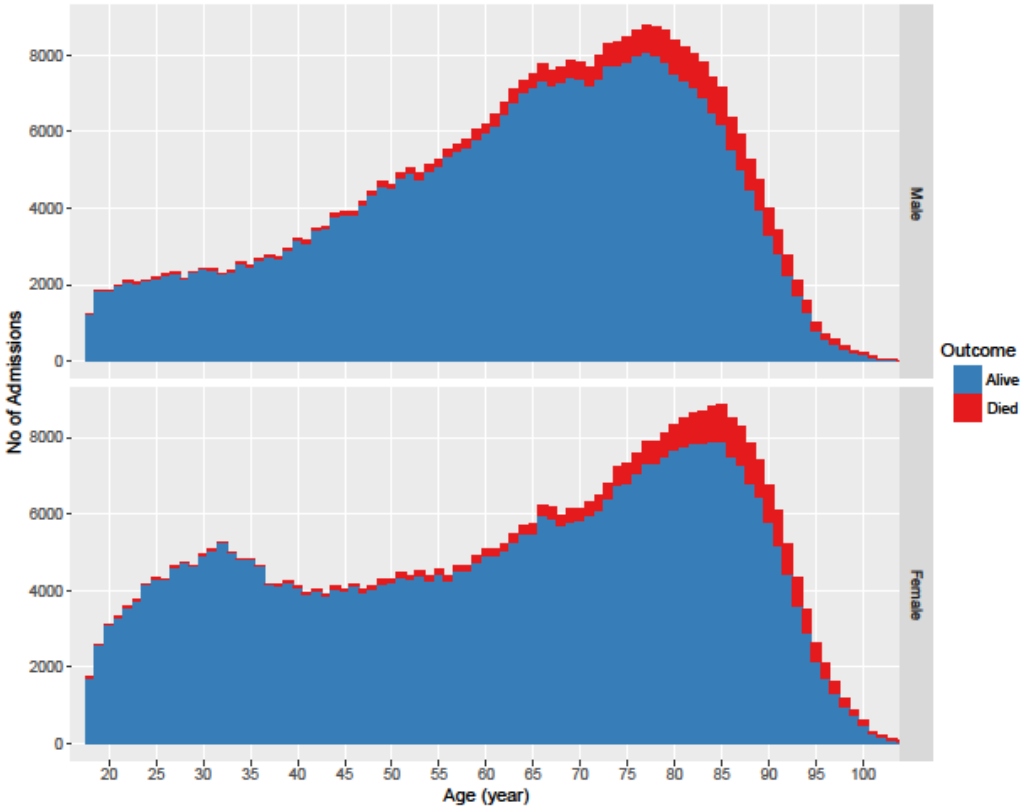


5.4.2 Two_Tests_Cohort

5.4.2.1 Baseline characteristics

Of all admissions to hospital (1,874,325), 43.28% remained there and had a second set of blood tests. Thus 811,268 admissions met the inclusion criteria for the ML_Two_Tests_Cohort. Of these, 6.1% (49,818) died in hospital. The median LoS was 6 days (IQR: 3–13 days). Female admissions were 52.64% of the total. The median age was 66 years (IQR 47–80). The histogram of the ages of admissions is shown in Figure 5.3.

Figure 5.4: Histogram of Age



Emergency, elective, and maternity admissions accounted for 71.9%, 20.6% and 4% of the total respectively, the remainder being composed of the 'Other' category. The full distribution of the admissions by ethnic category, method, month, and day of the week are shown in Tables 5.9 to 5.12 respectively. The ethnic category 'British' had the highest proportion of admissions at 43.7%, followed by the 'Not Stated' category at 31.7%. The prevalence of the various ICD-10 codes is shown in Table 5.13. Hypertension (I10) and type 2 diabetes mellitus had the highest prevalence, at 29.8% and 14.7% respectively. The median values of all the second blood tests, including the number with missing values, are shown in Table 5.14; of note are the proportion of admissions with no results for platelets, RDW and albumin, at 31.1%, 31.1% and 34.8% respectively. The presence of both a sodium and a haemoglobin result were part of the inclusion criteria, hence the presence of these in all admissions.

The in-hospital mortality rates for each group of admissions is shown in Tables 5.9 to 5.12. Within ethnic categories, the highest mortality rates (significantly higher than in all other ethnicities) occurred in the 'Not Stated' and 'British' categories, at 7.6% and 6.5% respectively. Mortality by method of admission (Emergency: 7.6%; Elective: 1.5%; and Maternity: 0.04%; $p<0.01$) substantiates the view that patients admitted as an emergency are more likely to die than those admitted via either the elective or the maternity method. The 'Other' method of admission (the bulk of which consisted of non-emergency inter-hospital transfers) had a small prevalence (1.6%), but the highest mortality rate of all, at 10.3%. The month of admission with the highest mortality was December, at 7.5%. Weekend admissions had marginally higher mortality rates when compared with weekdays, at 6.81% vs 5.97% ($p<0.05$), but the difference was only 0.84%.

Table 5.9: Two_Tests_Cohort: Ethnic category prevalence, and in-hospital mortality rates

Ethnic Category	Number of Admissions	Prevalence % of total Total = 811,268	In-Hospital Mortality (95% CI)
British (A)	354,127	43.65 %	6.45 % (6.37–6.53)
Irish (B)	14,776	1.82 %	5.48 % (5.13–5.86)
Any other white background (C)	55,958	6.9 %	3.85 % (3.7–4.02)
White and black Caribbean (D)	1,618	0.2 %	2.29 % (1.66–3.14)
White and black African (E)	951	0.12 %	2.94 % (2.04–4.22)
White and Asian (F)	1,001	0.12 %	2.9 % (2.02–4.13)
Any other mixed background (G)	3,114	0.38 %	2.79 % (2.27–3.43)
Indian (H)	16,297	2.01 %	4.4 % (4.1-4.73)
Pakistani (J)	6,910	0.85 %	3.13 % (2.74–3.56)
Bangladeshi (K)	11,429	1.41 %	3.08 % (2.78–3.41)
Any other Asian background (L)	15,182	1.87 %	3.58 % (3.29–3.88)
Caribbean (M)	16,680	2.06 %	4 % (3.72–4.31)
African (N)	21,084	2.6 %	1.98 % (1.80–2.18)

Any other black background (P)	8,404	1.04 %	2.52 % (2.21–2.88)
Chinese (R)	3,333	0.41 %	3.96 % (3.35–4.68)
Any other ethnic group (S)	23,396	2.88 %	4.17 % (3.92–4.43)
Not Stated (Z)	257,008	31.68 %	7.63 % (7.52–7.73)

Table 5.10: Two_Tests_Cohort: Method of admission prevalence, and in-hospital mortality rates

Method of Admission		Number of Admissions	Prevalence: % of total Total = 811,268	In-Hospital Mortality *(95% CI)*
Elective	Elective Waiting list (11)	96,854	11.94%	1.16 % (1.09–1.23)
	Elective Booked (12)	34,476	4.25%	2.03 % (1.89–2.19)
	Elective Planned (13)	35,586	4.39%	1.68 % (1.55–1.82)
Emergency	Accident and emergency or dental casualty department of the health care Provider (21)	423,340	52.18%	7.60 % (7.52–7.68)
	General practitioner (22)	77,195	9.52%	9.22 % (9.01–9.42)
	Bed bureau (23)	5,016	0.62%	6.02 % (5.4–6.71)
	Consultant clinic, of this or another health care provider (24)	20,562	2.53%	4.27 % (4.01–4.56)
	Admission via mental health crisis resolution team (25)	239	0.03%	10.04 % (6.84–14.51)
	Transfer of an admitted patient from another hospital provider in an emergency (28)	55,374	6.83%	6.9 % (6.69–7.11)
	Accident and emergency department of another provider where the patient had not been admitted (2A)	124	0.02%	11.29 % (6.85–18.06)
	Other means (2B)	352	0.04%	13.07 % (9.94–16.99)
	Other emergency admission (2D)	896	0.11%	2.01 % (1.27–3.15)
Maternity	Admitted ante-partum (31)	30,753	3.79%	0.04 % (0.02–0.06)

	Admitted post-partum (32)	1,470	0.18%	0.14 % (0.04-0.49)
Other	Transfer of any admitted patient from other hospital provider other than in an emergency (81)	28,756	3.54%	10.36 % (10.01–10.71)
	The birth of a baby in this health care provider (82)	55	0.01%	3.64 % (1–12)
	Baby born outside the health care provider except when born at home as intended (83)	18	0.00%	11.11 % (3.1–32.8)
	Not known / validation error (99)	202	0.02%	6.93 % (4.17–11.3)

Table 5.11: Two_Tests_Cohort: Month of admission prevalence, and in-hospital mortality rates

Month	Number of Admissions	Prevalence: % of total Total = 811,268	In-Hospital Mortality (95% CI)
January	72,058	8.88%	6.78 % (6.6–6.97)
February	65,963	8.13%	6.54 % (6.35–6.73)
March	68,533	8.45%	6.44 % (6.26–6.62)
April	66,598	8.21%	6.17 % (5.99-6.35)
May	68,435	8.44%	5.64 % (5.47–5.82)
June	64,384	7.94%	5.4 % (5.23–5.58)
July	66,981	8.26%	5.6 % (5.43–5.77)
August	64,552	7.96%	5.8 % (5.62–5.98)
September	67,004	8.26%	5.5 % (5.33–5.67)
October	70,569	8.70%	6.1 % (5.92–6.28)
November	68,395	8.43%	6.29 % (6.11–6.47)
December	67,796	8.36%	7.35 % (7.16–7.55)

Table 5.12: Two_Tests_Cohort: Day of admission prevalence, and in-hospital mortality rates				
Day of Admission		**Number of Admissions**	**Prevalence: % of total Total = 811,268**	**In-Hospital Mortality *(95% CI)***
Weekday	Monday	131,627	16.22%	5.91 % (5.79–6.04)
	Tuesday	137,863	16.99%	5.72 % (5.6–5.84)
	Wednesday	130,647	16.10%	5.75 % (5.62–5.87)
	Thursday	127,933	15.77%	5.96 % (5.83–6.09)
	Friday	117,523	14.49%	6.58 % (6.44–6.73)
Weekend	Saturday	79,278	9.77%	7.18 % (7–7.36)
	Sunday	86,397	10.65%	6.47 % (6.31–6.64)

Table 5.13: Two_Tests_Cohort: Prevalence of ICD10 codes			
Co-morbidity	**ICD10 code**	**No. of admissions**	**Prevalence: % of total (811,268)**
Essential (primary) hypertension	I10	241,620	29.78%
Angina Pectoris	I20	34,132	4.21%
Chronic ischaemic heart disease	I25	109,753	13.53%
Atrial fibrillation or flutter	I48	102,601	12.65%
Heart Failure	I50	61,157	7.54%
Personal history of diseases of the circulatory system	Z867	64,779	7.98%
Secondary malignant neoplasm of lymph nodes	C77	8,263	1.02%
Secondary malignant neoplasm of other sites	C79	16,522	2.04%
Multiple myeloma and malignant plasma cell neoplasms	C90	7,309	0.90%
Myeloid leukaemia	C92	6,847	0.84%
Other hypothyroidism	E03	39,889	4.92%
Type 2 diabetes mellitus	E11	118,876	14.65%
Obesity	E66	18,622	2.30%
Disorders of lipoprotein metabolism and other lipidaemias	E78	87,028	10.73%
Mental and behavioural disorders due to use of alcohol	F10	24,108	2.97%

Smoking	F17	47,218	5.82%
Depressive episode	F32	34,828	4.29%
Epilepsy	G40	18,547	2.29%
Other chronic obstructive pulmonary disease	J44	65,617	8.09%
Asthma	J45	54,683	6.74%
Chronic kidney disease	N18	58,916	7.26%
Other symptoms and signs involving the nervous and musculoskeletal systems	R29	26,540	3.27%
Personal history of malignant neoplasm of digestive organs	Z850	12,958	1.60%
Personal history of psychoactive substance abuse	Z864	70,416	8.68%
Personal history of diseases of the nervous system and sense organs	Z866	13,896	1.71%
Personal history of allergy to penicillin	Z880	46,756	5.76%
Personal history of allergy to other drugs, and biological substances	Z888	17,042	2.10%
Presence of electronic cardiac devices	Z950	21,788	2.69%
Presence of aortocoronary bypass graft	Z951	23,444	2.89%
Presence of coronary angioplasty implant and graft	Z955	17,357	2.14%

Table 5.14: Two_Tests_Cohort: Median (Interquartile Range) of Blood Tests

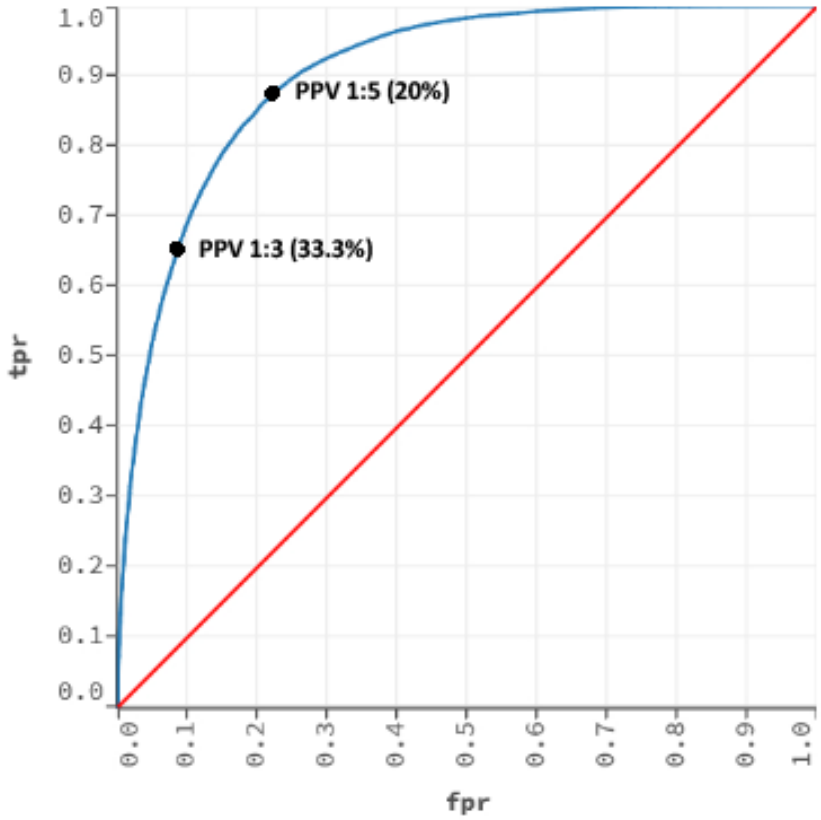
Blood Test	Median (Interquartile Range)	Percentage missing of total (total = 811,268)
Sodium	138 (136–141)	0 %
Potassium	4.1 (3.8–4.5)	3.9 %
Urea	5.4 (3.8–8.4)	4.9 %
Creatinine	78 (62–107)	9.9 %
Ur:Cr	65.8 (50–86.3)	10.9 %
Haemoglobin	11.6 (10.2–13.1)	0 %
Platelets	227 (173–292)	31.1 %
White Cell Count	8.5 (6.3–11.4)	1.1 %
Mean Cell Volume	89.2 (85–93.4)	1 %
Red Cell Distribution Width	14.2 (13.2–15.6)	31.1 %
Albumin	35 (30–40)	34.8 %

5.4.2.2 ML-Two-Tests Model

The Train Dataset, and Test Dataset were randomised effectively, and both had the same mortality rate of 6.1%. The hyper-parameters (determined by a grid search) that produced the lowest mean logloss and highest mean AUROC via cross-validation were: number of trees: 5,000; tree depth: 8; and learning rate: 0.01.

The ML-Two-Tests model achieved an AUROC of 90.6% (log loss: 0.152, Figure 5.5) on the Test Dataset, which was similar to the mean cross-validation AUROC of 90.3% (logloss: 0.154). At the PPV 1:5 (20%) threshold, the model achieved a sensitivity of 88.2%, and a specificity of 77% (the highest ever achieved). This means that, if a rapid response team were directed to all admissions who triggered the model at this threshold (at an in-hospital mortality rate of 6.1%), only 27% of admissions would be referred; however, 88.2% of all patients likely to die would be seen. In other words, for every 1,000 admissions to hospital for whom there are two sets of blood tests, approximately 61 admissions would be likely to die. The referral rate of the model in this scenario would be 270 per 1,000 admissions, within which would be 54 of the 61 patients likely to die.

Figure 5.5: ML-Two-Tests: Receiver Operator Curve (AUROC 90.6%)

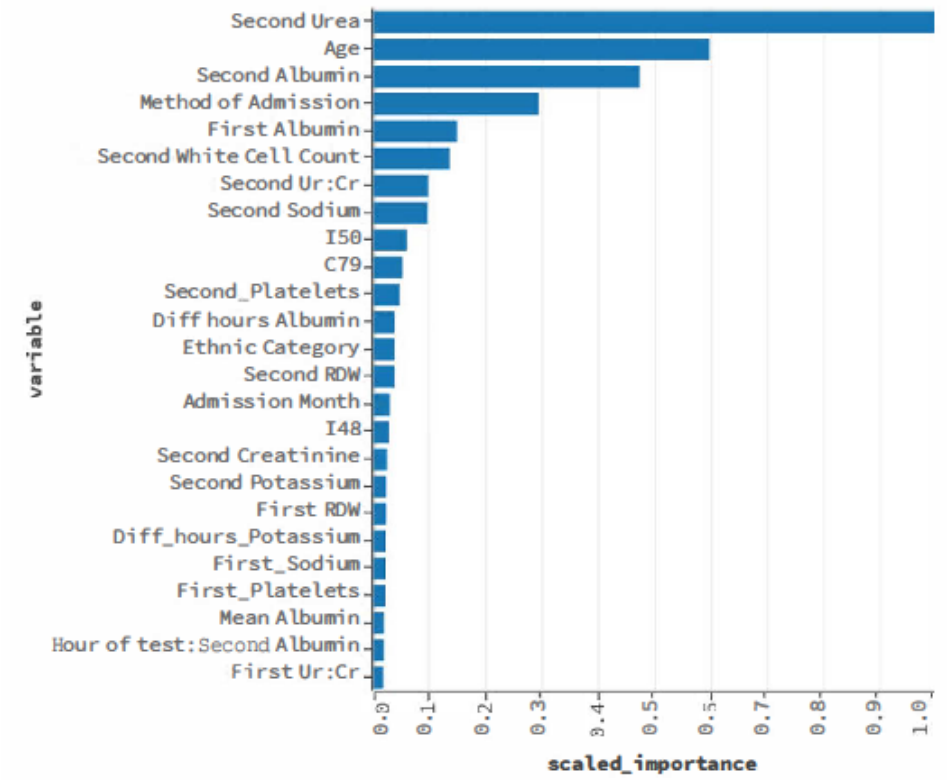


tpr (true positive rate): sensitivity; fpr (false positive rate): 1 – specificity; PPV 1:5 : positive predictive value of 1 in 5 threshold; PPV 1:3 : positive predictive value of 1 in 3 threshold

Similarly, for the PPV 1:3 (33%) threshold, the model's sensitivity was 65%, with a specificity of 91.5%. For every 1,000 admissions to hospital, the referral rate would be approximately 120 per 1,000 (12%) admissions, within which would be 40 of the 61 patients likely to die.

The scaled variable importance of the twenty-five most important predictor variables is shown in Figure 5.6. The four most important variables were Second Urea, Age, Second Albumin and Method of Admission.

Figure 5.6: ML-Two-Tests: Scaled variable importance of the top 25 variables



5.5 Discussion

This analysis uses the largest cohort of admissions (1,874,325) yet investigated for mortality in relation to blood test results. Examining the distribution of patient admissions highlights a few key patterns, which vary depending on whether or not the patient stayed in hospital to have a second set of blood tests (i.e. between the two cohorts). For the Admission cohort, the mode of age for male admissions was 67 years. However, for female admissions this was 32 years; on further investigation, this was found to be due to the prevalence of maternity admissions in this lower age group. The effect of maternity admissions on the burden of in-hospital care decreased in the Two_Tests_Cohort, where, although a slight 'bump' was seen at age 32 years, the mode for female admissions was now 85 years. The mode age for male admissions in the Two_Tests_Cohort increased by ten years to 77 years. Thus,

older patients underwent more investigations in hospital. The burden of admissions also changed between the cohorts when the patients' method of admission was examined. The prevalence of emergency admissions rose from 57.6% to 71.9% ($p<0.01$) from the Admission to the Two_Tests_Cohort, with concurrent falls in elective admissions and maternity admissions. This indicates that elective admissions, when compared with emergency admissions, are more likely to be discharged after their first set of blood results. Patients who have a second set of blood tests are sicker than those who just have one set. This is demonstrated by the increase in the prevalence of co-morbidities in the Two_Tests_Cohort, in comparison with the Admission cohort: the proportion of patients with hypertension, chronic ischaemic heart disease, heart failure, type 2 diabetes and COPD all increased, by 6.7%, 4.1%, 3.3%, 4% and 2.7% respectively (p values <0.01). In addition, the median LoS increased by four days, and the mortality by 3%, from the Admission to the Two_Tests_Cohort.

In the Admission_Cohort, regarding admissions over the weekend vs weekdays, the mortality rate difference in these groups fell to non-clinically significant levels (1.4% to 0.3%) when only emergency admissions were examined, as opposed to all admissions. This indicates that the much referred-to 'weekend effect' may be purely due to a cohort difference: i.e. there are far fewer elective admissions over the weekend, and elective admissions have an extremely low mortality rate.

The in-hospital mortality results for both cohorts do, however, reinforce the notion that without any statistical corrections (for age, disease groups or method of admission), mortality is higher for patients admitted over the winter (December and January) vs the summer (May–September). Patients admitted as emergencies had higher rates of dying in hospital than either elective or maternity admissions (Admission: 4.9% vs 0.4% & 0.07%; Two_Tests: 7.6% vs 1.5% & 0.04%). Surprisingly, however, non-emergency transfers from another hospital had the highest in-hospital mortality of all methods of admission, which was even greater than for emergency transfers (Admission: 9.1% vs 4.2%; Two_Tests: 10.4% vs 6.9% (p values <0.01)). A possible explanation is that these patients may have been pre-arranged transfers between critical care units, or the movement of especially complex patients to or from specialist care centres, i.e. the sickest patients in hospital. Further investigation would need to be carried out to explore this further.

The ML-Admission and ML-Two-Tests models are by far the most accurate models to date for predicting a patient's risk of dying in hospital. Not only do the models when used together enable an initial prediction to be made at the point of admission, but they also allow for a modification of this prediction, as the patient undergoes further investigations in hospital in the context of a changing hospital patient population. The AUROCs of 93% and 90.1% surpass all previous EWSCs, thus providing better discrimination between patients likely to live or die. Two previous EWSCs by Assadalohi[31] and

Jarvis[96] achieved AUROCS of only 76.1% and 78.9% respectively, for my Admission cohort. What makes my two EWSCs especially remarkable is the fact that the models achieved this performance despite significant proportions of missing blood test results: creatinine, platelets, RDW, and albumin results were missing in 8.9%, 28.6%, 29.4% and 21.1% of admissions respectively for the Admission_Cohort, and in 9.9%, 31.3%, 31.1% and 34.8% of admissions respectively in the Two_Tests_Cohort.

The change in importance of different variables (Figures 5.3 and 5.6) in both these models is complex, but they do both reinforce known risk predictors as well as indicate new avenues for further investigation. The latest blood results (second) become more important for predicting risk than the admission blood results as patients have multiple tests done. This makes sense, as the latest blood result is a more accurate reflection of an individual patient's current physiology. The fall in importance Method of Admission from ML-Admission to ML-Two-Tests could however be because of the change in the underlying data, there are a lot less patients in the Two_Tests_Cohort admitted as elective/ maternity vs emergency. As emergency patients have a higher mortality, the probability of mortality gleaned from Method of Admission decreases slightly (as the probability of patient being an emergency increases), this is also probably the reason that Urea at First Aki stage in ML-AKI (Table 4.10) has a higher importance than Method of Admission. Albumin however, despite being absent in >21.1% (Admission_Cohort) and 34.5% (Two_Tests_Cohort) cases still have a high influence on the model result, reinforcing the complexity of the contribution of a variable to the model. It is important to note that the scaled variable importance is purely based on mathematical analysis of the model structure itself, and does not mean a higher or lower value results in increased risk, just that the defined variable influences the final result of the model more than (or less than others).

Deciding on the optimum threshold at which a patient should be escalated to either 1) immediate review by a consultant (as opposed to the current guidelines' recommendation of review within 12 hours), or 2) a rapid response team, is still a matter of discussion. From a patient's perspective, it would be understandable to expect be told if he/she has an in-hospital risk of death of at least 20%, and also for enhanced clinical care to be initiated, to reduce such risk. Thus, this would support the lower threshold of PPV 1:5 (20%). On admission, this threshold would reliably identify 75% of all patients likely to die within their current hospital stay, but would also result in 12% of all admissions having their care escalated. A single dedicated RRT may struggle to cope with assessing and managing the care of such a high total number of hospital admissions. Therefore, at the point of admission, it may be prudent to redirect all referrals at this level for an immediate consultant review, rather than requiring a review within 12 hours, as per current guidelines. However, by the time a patient has a second set of blood tests, the situation has changed, as follows: 1) there are 56% fewer patients, and 2) as some time has elapsed between sets of blood tests, it could be expected that some treatment or

intervention to improve the patient's outcome has been initiated. At this point, a continued proven 20% risk of dying in hospital should be unacceptable; unfortunately, this encompasses 27% of the remaining patients in hospital. Therefore, although at the outset a lower threshold (PPV 1:5) seems to be the appropriate trigger level for escalation of patient care, the existing resources in hospitals may not be able to deliver this effectively. Hence, the PPV 1:3 (33.3%) could be an alternative, where the referral/escalation rates would be 4.6 % at admission and 12% at the Two_Tests stage. The PPV 1:3 (33.3%) threshold could thus provide a balance between: 1) severity of a patient's condition, 2) RRT workload, and 3) the proportion of patients likely to die being identified (sensitivity). The disadvantage of using a higher threshold is that it identifies a smaller proportion of the total number of patients likely to die (sensitivity): Two_Tests PPV 1:3: 65% vs PPV 1:5: 88.2%; Admission PPV 1:3: 49.3% vs PPV 1:5: 76.1%.

It is important to note, though, that these models were from data obtained in an environment where patients are already receiving treatment in hospital. This means that if a patient is predicted to live, this does not necessarily equate to that patient being well; it could also mean that this patient is unwell but being effectively treated by their existing clinical teams. Similarly, a prediction that a patient is likely to die does not always mean that this patient's death is completely preventable. Other limitations of my models include the fact that patients on a palliative pathway are not reliably identified; therefore, some of the in-hospital deaths may be expected. Another issue with this approach is that the cause of a patient's imminent mortality is not described; thus the rapid response team may have to perform additional investigations on the patient to determine the optimum intervention to prevent the patient's death. These limitations are common to all EWSCs, whether they are laboratory-based, vital sign or co-morbidity based, and are not unique to my models. Despite these issues, EWSCs are mandated across the globe.

In summary, I have created two EWSCs, that when deployed together can be used for all hospitalised patients, to both provide risk of death at the point of admission, and also modify this risk while in hospital. The positive predictive values of identifying patients likely to die exceed those of any other EWSC yet developed, while at the same time capturing the majority of patients likely to die (i.e. a high sensitivity). The efficacy of affecting patient care is still open for debate; however, the deployment of EWSCs are a requirement of multiple national guidelines, and are thus standard practice. The increased accuracy of my two models (ML-Admission and ML-Two-Tests) would better direct resources to those groups of patients most likely to die than any current system. The technological mechanism needed to implement such a blood test EWSC for real-time patient data is, however, not straightforward: the solving of this technological problem will be the focus of my next chapter.

Finally, it is important to note, that there are a multitude of additional data sources that could be easily be incorporated into a future EWSC. The most simple being activity data from wearables (Fitbit, apple watch, etc.) or patient's smartphones. Mobility after an operation is a key metric to judge the success (and recovery) of a patient after certain operations, and now tracking this has become simple. Monitoring could also continue once a patient is discharged home, with no additional complex hardware, allowing for a home based EWSC on the most vulnerable patients. In fact, there are companies providing such services in the community already such as Sentrian. Sentrian allows for the early identification of patients at risk of deterioration at home, and enables the direction of the appropriate resources to that patient. The opportunities to track and enhance patient care are thus just beginning.

Chapter 6: Patient Rescue: technology platform

6.1 Introduction

So far in this thesis, I have focussed on creating models that can identify hospitalised patients who are deteriorating or likely to deteriorate, using existing electronic data sets. However, none of these models were implemented in a real world hospital environment. A certified technology platform that would enable the implementation of such models as I have created in real time, i.e. milliseconds after data is generated, in a clinical environment, does not, as far as I know exist. Therefore, it is necessary to create such a platform if real time tracking of a patient's AE risk is to be implemented.

Such a system could be created by either 1) significant modifications to an existing electronic health records system, or 2) by building a new system focussed on real-time analytics. As access to and modification of a proprietary commercial electronic system was not feasible (the commercial products do not allow for such modification), this chapter describes the creation of an independent real-time analytics platform for use in health care. This technology proof of concept I have created, titled 'Patient Rescue', is based on principles and technologies that are now common practice in existing financial and event-based analytic systems.

In order to fund the significant development of 'Patient Rescue', I successfully applied for an extremely competitive NHS England 'Regional Innovation Fund Award', and in January 2014 was awarded £183,000 to create the prototype technology platform. The focus of this award was to create a real-time analytics system and demonstrate its capability by implementing the NHS England Acute Kidney Injury Algorithm, which at the time could not be implemented by the incumbent LIMS providers.

I conceived, led, managed, and wrote part of the software code for the Patient Rescue technology platform. I was supported by a core team consisting of a part-time software developer (Simon Brown), a part-time NHS information analyst (Aniruddha Dwarakanath), and a part-time project manager (Prashant Lele). The programming and implementation was a team effort, and the system described in this chapter had contributions from the core team.

In addition to the core team above, representatives from partner hospitals advised on issues relating to information governance, and specifically their hospital's AKI service. Members of the various hospitals' IT teams also created (by replicating existing information channels) and maintained the 'data pipes' that inputted data into my technology platform.

The key challenges to creating a real-time analytics platform for healthcare, and a summary of the key components of my platform, are described below.

6.2 The Challenges

The implementation of such models as I have developed, especially in real-time, would be a challenge in any hospital. There are three primary reasons for this:

1. Access
2. Analytics
3. Clinical Communication

6.2.1 Access

In order to implement a model, all the various data variables of the model must be accessible. The variables for my models are inputted into and reside in different systems within an NHS hospital. Blood result data is part of the laboratory information management system (LIMS). Co-morbidity data is stored within an electronic health record (if such exists) or in a business intelligence database (used to create the NHS Trust's CDS.APC submission to HSCIC (NHS England)). Admission and discharge from hospital would traditionally be part of a patient administration system (PAS). Thus, although the various data elements exist in electronic form, accessing them requires extracting them from multiple systems. The two approaches to access such data are 1) 'Push': to 'push' the data out of the system as and when it is created; and 2) 'Pull': to allow an external system to 'query' the system the data resides in, and if relevant data is present, for the host system to then send this data, i.e. 'pull' the data out. The problem with the 'pull' approach is that most hospital IT systems are not designed for intensive querying, so that if many queries are made on the system, the entire systems slows down or crashes. These host information systems are based on old database architectures and optimised for the performance of very specific queries, which were defined when the information system was first integrated with the rest of the hospital's systems. The 'Push' approach, on the other hand, is more viable, as the workload on the existing hospital system is quantifiable; and more importantly, most hospital systems already communicate their core information to other systems within the hospital. Acquiring the data required to implement the model could, therefore, theoretically be extracted from one of these existing 'Push' data channels. Although possible, this is a substantial task: the previously described (Chapter 2, Section 2.3) challenges of pre-processing the data are all present when the data are accessed in real-time using this method, and need to be overcome. My solution to this challenge comprised the following steps: 1) to consume the data from these 'Push' data channels; 2) to store all

the data in an unstructured raw format in a 'data lake'; and 3) to pre-process only those data elements required for the implementation of the current models.

The standard communication protocol for health care systems are HL7 messages. 'HL7' refers to Health Level Seven International[146], an American National Standards Institute accredited standards-developing organisation dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. An HL7 message is a hierarchically structured message with multiple segments in a defined sequence.

The term 'data lake' was supposedly coined by James Dixon, the Chief Technology Officer of Pentaho. In a blog entry he said: 'If you think of a datamart as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples.'[147] [148]

6.2.2 Analytics

I have referred to analytics as an all-encompassing term that includes pre-processing of the acquired data, as well as the implementation of the pre-defined models. However, this is an oversimplification. Some of the tasks that form part of pre-processing are described in Chapter 2, Section 2.3; all those steps now need to be implemented for data that is constantly being ingested into the system. Thus, the system needs to carry out these tasks automatically, and must also have robust quality control and monitoring mechanisms built in, to confirm that the automatic tasks are indeed performing their tasks accurately. Another issue with the application of the model concerns the time when such a model should be applied to the data. The results of the various data variables may arrive at different times (for example, a sodium test result may arrive 30 minutes after the sample is received in the laboratory, while the haemoglobin test result may take 60 minutes), or they may arrive for patients no longer in hospital (outpatient/GPs' tests). Therefore, rules to cater for all of these scenarios must be created. A specific user case for Acute Kidney Injury would be a patient newly admitted to hospital via accident and emergency, where he/she had already undergone blood tests. In this scenario, an AKI algorithm should be run on the data of this new admission, despite the possibility of no new blood tests being requested on admission (as blood tests had been carried out recently). In addition to the rules described above, the analytics platform must also be designed to such a standard that it will eventually be approved by the appropriate Medical Health Regulatory Agency (MHRA) standards for a medical device. Such a system would be a 'Class 2' (2A or 2B depending on implementation) medical device,

which would require robust quality control and logs of all calculations performed, among many other requirements.

6.2.3 Clinical Communication

Finally, once a model has successfully been applied to a patient's data in a timely manner (real-time), relevant information relating to the outcome of the model's result must be communicated to the relevant clinical team in the hospital. However, merely informing the clinical team member of a positive or negative outcome of the model may not necessarily aid patient care. Therefore, early in the process I made the decision that the system would need to provide 'actionable insight': clinical decision support, by incorporating not only the result of the model, but also other information on the patient. As the demonstration of 'Patient Rescue' was focused on AKI, I created a report that incorporated information that a nephrologist would request if referred a patient diagnosed with AKI. The actual delivery of the report to the clinician was beyond the scope of the project; but to demonstrate communication, a simple secure email system was built. This system utilised the secure and encrypted NHS.net email facility, and allowed reports for each AKI patient to be sent to dedicated NHS.net email addresses within 1 second of the system identifying a patient as having new AKI.

Communicating the condition of a deteriorating patient is, however, a much bigger problem, which needs further investigation and was beyond the scope of my study. The challenges involved include identifying the appropriate team and the appropriate method of communication, and verification of communication. All these parameters may change on a daily basis for the multiple hospital teams. The teams may themselves change, when a patient is transferred to another physical location or to a different medical/surgical specialty. The time of day will also influence the choice of which teams are to be contacted.

6.3 The Solution: Patient Rescue

6.3.1 System Overview

The Patient Rescue platform ingests HL7 messages from multiple channels, then performs real-time analysis of the information in these messages with respect to previous information received about a patient. Based on the outcome of this analysis, the platform can generate alerts and/or reports.

Data received via HL7 channels is persisted in a database and combined with any bulk imported data (for example, historic patient information). Along with real-time analysis of messages, algorithms can be run on the persisted data in order to perform one-off analyses.

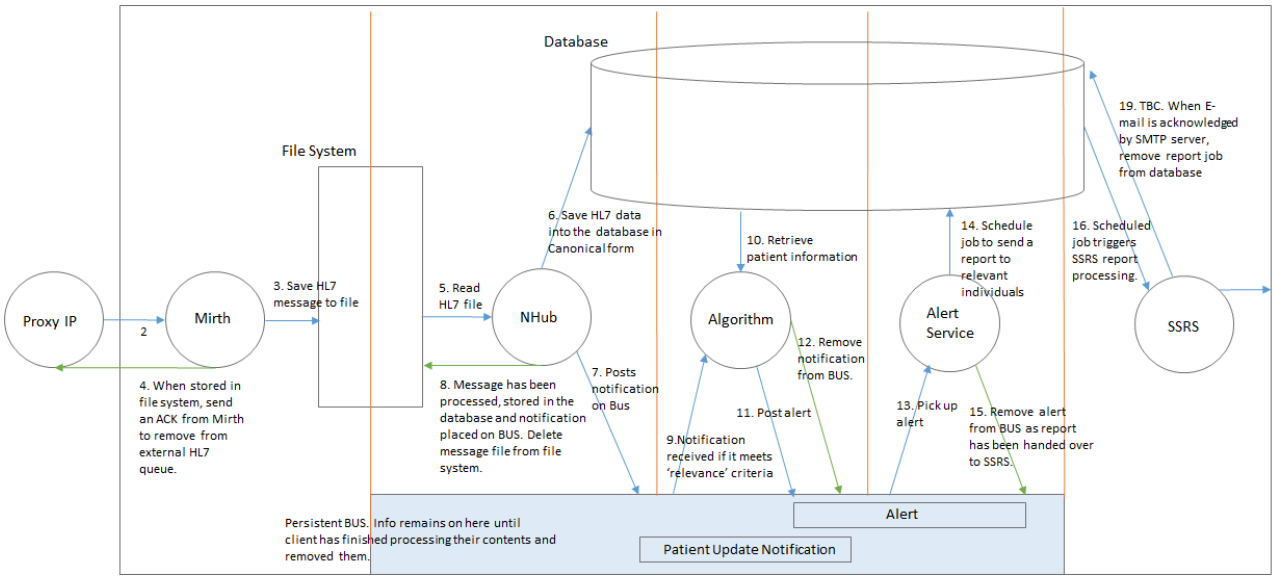
The platform has the following features:

1. Data from different sources are mapped to a canonical form before being analysed and persisted. This overcomes the issue of variations in the terminology and meaning of information sent (which are allowed within the HL7 framework).
2. The components of the platform are loosely coupled. This means that different implementation technologies, algorithms and actions as a result of analysis can be plugged in to the system. This design leads to increased flexibility and allows rapid deployment of new algorithms and actions triggered by these algorithms.
3. The platform is highly scalable, in that:
    a. New sources of data can be added quickly.
    b. The database schema is optimised for large quantities of data, and the database software makes use of in-memory tables, allowing exceptionally fast queries to be performed on large datasets.
    c. Components can be distributed across multiple servers.
    d. Components that are 'plugged into' the system (e.g. algorithms and alerting modules) can subscribe to 'topics' of interest. Given the large volumes of data flowing through the system, this means they only receive the information they define as necessary to perform their function, which reduces their processing demands.

6.3.2 System Schematic

A schematic of the live element of the Patient Rescue platform is shown in Figure 6.1

Figure 6.1 Patient Rescue System Schematic



6.3.3 Component Description

6.3.3.1 HL7 Gateway

The platform uses the gateway component of the open source Mirth Connect[149] to consume HL7 messages. This ensures resilient delivery of messages to the platform from multiple sources, and manages the protocol 'conversation' between the platform and the message source. Its admin console allows new sources (channels) to be added quickly, and configures the delivery and persistence (to file) of these messages. Each individual HL7 message from a given channel is saved as an individual file, and also concatenated into a single file for that channel and message type. Both individual messages and concatenated files are saved as plain text files, allowing visual inspection and replay of messages (thus facilitating resilience and testing).

6.3.3.2 NHUB

This is a bespoke .NET component, which performs the following tasks:

- Consuming HL7 messages saved to disk by the Mirth Connect.
- Mapping the information in the messages into the canonical form.
- Persistence of the information in the messages to the database.
- Placing the information in the messages on the Service Bus, to enable notification of interested consumers (e.g. algorithms).

Multiple instances of NHub are run on the server, typically one per data source (hospital HL7 feed). Each data source has different ways of representing medical data, so each instance of NHub uses a different configuration file and mapping data in order to translate the data into a canonical form. While each instance is single-threaded, as many instances can be run as are needed to deliver the throughput required.

For resilience, messages are placed on the Service Bus only once they have been persisted into the database. Following acknowledgement of successful posting on the Service Bus, the individual HL7 text file is deleted, thus preventing the duplicate processing of messages.

6.3.3.3 Service Bus

An open-source messaging framework, Rabbit MQ[150], with .NET extensions was selected for the Service Bus, and configured/coded for the platform. The Service Bus allows processes to post information (for example, a set of blood test results), with metadata describing that information (for example, the types of blood tests), which is termed the 'topic'. Other processes can subscribe to listen to topics. When a topic relevant to the subscriber is published on the Service Bus, the subscriber

receives an update and can act upon the information. Information is held on the Service Bus until all the subscribers have acknowledged receipt of the posted information, thus facilitating resilience.

In the case of the Patient Rescue platform, the NHub process publishes information in canonical form to the Service Bus, and relevant algorithms are subscribed to data updates containing information they require to perform their function.

6.3.3.4 CDS.APC Gateway

This component allows for the monthly ingestion of the CDS.APC data that the NHS hospitals submit to NHS England. This component was built using Microsoft SQL Server Integration Services. As with the HL7 gateway, this component enables the conversion of the raw hospital-submitted CDS.APC data to be quality checked, converted to canonical form, and persisted into the database. The data from the CDS.PCS reports is regarded as the highest-quality data; thus, during persistence into the database, the entire database is updated and quality checked.

6.3.3.5 Algorithms

Algorithms for the platform were written in C# and Matlab. These algorithms subscribe to topics posted on the Service Bus. When relevant data (for example, creatinine blood result) is posted on the Bus, they are notified and executed. Typically, the algorithms require contextual information (patient history and demographics) to determine their output. They extract this data from the database, using Structured Query Language (SQL). For the purposes of this proof of concept, the only algorithm written and tested was the NHS England AKI algorithm. After execution, the algorithms:

- Persist information to the database for audit purposes (at what time they were executed, for which patient, etc.).
- Publish messages to the Service Bus containing information that Alert processes can act upon (in this case, AKI level and patient identifiers).

6.3.3.6 Alert Service

Alert Services subscribe to the Service Bus, listening for topics of interest (AKI alerts). On receipt of an alert, they enrich it with the additional information they require by querying the database. Following this, they schedule an AKI report to be generated and sent to relevant parties. It is important to note that Alert Services can easily be developed to perform other functions, for example sending an SMS message to relevant parties.

6.3.3.7 SQL Server Reporting Services (SSRS)

SQL Server Reporting Services (SSRS) is a Microsoft product which generates reports in various formats from data stored in SQLServer. It also allows these reports to be sent via email. The Alert Services trigger SSRS to create and send AKI alerts to practitioners via email. These emails contain complex PDF (portable document format) reports, including patient history (in tabular and chart form), demographics and identifiers.

The Patient Rescue reporting system provides two levels of insight:

1) **Patient level:** real-time detection, alerting, decision support, management tracking and automated referral. The system:
    a) Tracks all blood and radiology result data relating to every hospitalised patient.
    b) Applies the NHS England AKI Patient Safety Alert Algorithm to all newly admitted patients and all creatinine results.
    c) For patients identified to have AKI (stages 1 to 3), a custom report is created. An example of one of my AKI reports on a fictional patient is shown in Figure 6.2. This report, the decision support tool, contains patient-specific information:
        i) Administrative Information: patient demographics, location in hospital, and days since admission.
        ii) Background Risk Factors: Past medical history of the patient that is relevant in the context of AKI, e.g. previous AKI, chronic kidney disease, neoplasms, heart disease.
        iii) Acute complications of AKI: markers of sepsis, hyperkalaemia, hyperphosphatamia.
        iv) Result Visualisation: Blood results of creatinine (Cr), urea (Ur), Ur:Cr, c reactive protein and white cell count are graphically visualised to enable the clinician to understand the trends and context of the current alert; i.e. whether it is a repeat, whether the patient has been progressively worsening or improving, etc.
        v) Radiology Reports: Reports on any renal-relevant radiological investigations performed on the patient in the last month are displayed, potentially enabling the cause of the current AKI to be determined.
        vi) Further Advice: Additional analyses are performed on the entire patient dataset to determine if there are any other abnormalities, or whether further investigations should be requested to exclude specific pathologies.
        vii) Automating the referral pathway: The report is sent (using dedicated NHS.net email addresses) to the Critical Care Outreach Team and for predefined severity levels of AKI also to the Renal Consultant in charge of AKI. Tracking of report delivery, reading of reports and interventions are all built into the system.

1) **Trust-Wide AKI Dashboards**: Patient Rescue also populates a continuously updating dashboard of all patients with AKI in the hospital. This allows the AKI caseload to be tracked and prioritised

across a large, multi-site organisation such as the Royal Free NHS Trust. Patients can be selected from the dashboard, and the full AKI report (as described above) may be viewed. It will also continuously populate a live AKI performance report. This presents data on the number of AKI episodes per timeframe, incidences of hospital- vs. community-acquired AKI, AKI mortality, AKI recovery, AKI length of stay, and the proportion of AKI patients who have timely follow-up blood tests and renal imaging.

The Patient Rescue platform was deployed in three NHS Hospitals. It was hosted in compliance with all NHS information governance principles and requirements (IGSOC), and was independently tested to ensure security and resilience. No research was carried out on the data processed by the platform, and the only algorithm deployed was the NHSE-AKI algorithm. No ML models were implemented on the live data flows, as these had not been peer reviewed, published nor validated as a standard for patient care. Patient Rescue was officially only a Data Processor for legal, regulatory and governance purposes. Although we reached required standards of quality, as the platform was not MHRA-certified, it was never used to treat any patients. The Patient Rescue platform was disconnected from the NHS hospitals once it had demonstrated that such a real-time analytics system could be built and deployed on real-time data flows in the NHS, and that such a system could monitor patients in real-time for their risk of deterioration. The core team and myself continue to work together, and are pursuing avenues to enable deployment of such a platform at scale.

Since the start of my PhD, there have been a number of initiatives/solutions, in the UK and internationally, that are tackling the problem of real time integration of disparate healthcare repositories and enabling live analytics. The most prominent in the UK being Datawell, a Connected Health Cities project of the Northern Health Science Alliance. Datawell builds on a platform developed by LumiraDx, and focusses on connecting multiple providers of healthcare (GP's, hospitals, private providers, etc.) and enabling patient centred care and analytics. In the USA, multiple hospital groups have built their own data and analytic solutions, and are carrying out machine learning analysis, two particular examples being the Mayo Clinic and the Children's Hospital of Orange County. The Watson Health division within IBM is also building on its cognitive computing solutions to gain insight from existing data, both within the hospital but also from public research repositories (e.g. pubmed). Overall, this is a rapidly advancing field, and I believe in just a few years, it will be commonplace for IT solutions to be augmenting the intelligence of doctors (and other healthcare professionals) and enabling better quality of patient care.

Figure 6.2 Example Patient Rescue: AKI report

Renal Rescue - Acute Kidney Injury Report

NHS

Name	**Mr. Nephron Kidneyson**				Sex	M	Age 65
Hospital	St. Elsewhere's	Hospital Number	999999	NHSNumber 123456789		DOB	01/01/1950
Location	Acute Kidney Unit						

Stage of AKI: 3 AKI alert time: 19 January 2015 16:00:00

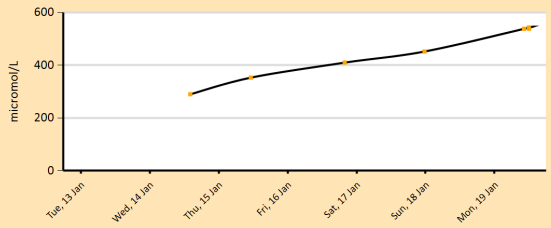
AKI Complications

Urea is high: 29.5 mmol/L- assess for uraemia
C reactive protein is high: 33 mg/L- consider sepsis
Phosphate is high: 2.89 mmol/L

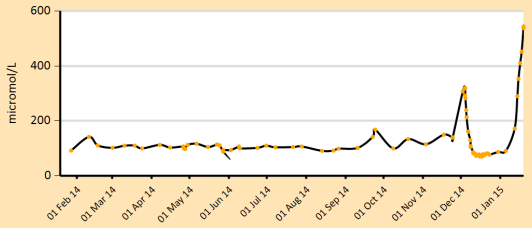
Background risk factors

Neoplasms

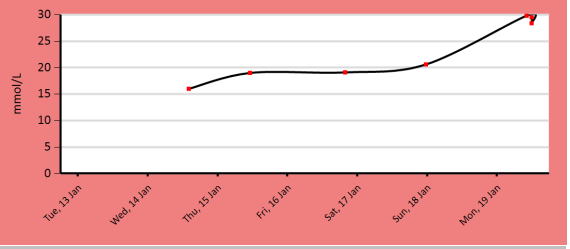
Creatinine - 7 day history



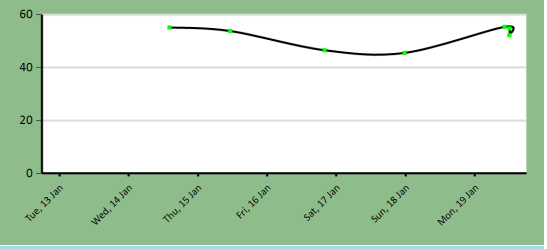
Creatinine - 1 yr history



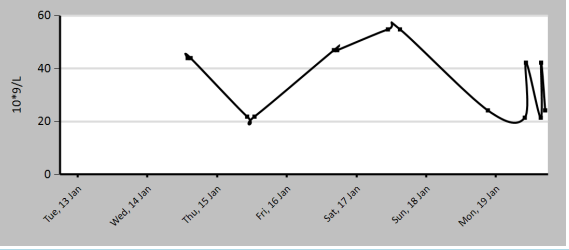
Urea - 7 day history



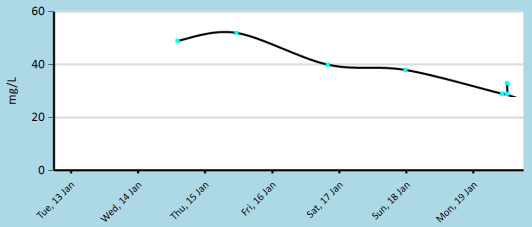
Urea:Creatinine - 7 day history



White Cell Count - 7 day history



C Reactive Protein - 7 day history



Further advice:

Platelets low: Consider microangiopathy- perform blood film, reticulocytes and haptoglobins
Lactate dehydrogenase: no results available in the last 48 hours
Immunoglobulins are abnormal: consider myeloma

Radiology report date: 17 January 2015 10:00

US RENAL TRACT: Both kidneys are enlarged, with increased reflectivity in keeping with a nephritic process. There is a complex cyst at the left upper pole .

No other renal lesion identified.

Chapter 7: Summary and Discussion

Hospitalised patients are at risk of deteriorating and suffering an AE while in hospital. These AEs include unexpected admissions to the intensive care unit, surgery, cardiac arrest, and death; and almost half of these AEs are thought to be preventable.[4,5,8,9,11] Identifying these deteriorating patients, or those at risk of deteriorating, is a current priority for our health care system.

It is well known (Chapter 1, Section 1.2) that prior to an AE, distinct patterns of vital signs occur. Therefore, current initiatives have focussed on identifying patients with these vital sign patterns, and escalating their care in accordance with the severity of their vital sign abnormalities (as indicated by their EWSC) to a RRT. This combination of a pattern detection mechanism (the afferent pathway) and the RRT (the efferent pathway) is known as a 'rapid response system'. However, the effectiveness of these vital-sign-based models is problematic. Patient vital sign measurements are carried out infrequently, and even when they are measured, they are often inaccurate. In addition, the vital-sign-based EWSCs themselves have poor sensitivity and specificity, resulting in a poor positive predictive value in discriminating between patients likely to suffer an AE and those who are not. Thus, the current systems do not adequately capture those at risk of an AE in hospital.

In this thesis, I have sought to address these issues by assessing if variables other than vital signs can better identify patients most at risk of an AE during a hospital admission. I have explored variables which are routinely and universally collected on hospitalised patients, and stored electronically, thus enabling the easy implementation of any system that is developed. Specifically, I have focussed on blood test results as well as administrative data such as demographics, method of admission, dates of admission, ICD10 codes of diagnoses and co-morbidities, and OPCS (procedure/intervention) codes. A few existing studies have explored the use of other variables such as these. The limitations of these existing non-vital-sign-based EWSCs include:

a) Poor generalisability: their effectiveness has not been proven in diverse hospital patients, as they are either specific to a type of patient (medical, surgical, emergency, elective, etc.), or they have been created and tested using patients from only one or two hospitals.
b) Simplistic analysis: they take account of neither the dynamic change (trends), nor the inter-correlations between blood results.
c) Lack of patient specific information: they do not account for a patient's co-morbidities.
d) Lack of context sensitive information: they account for neither the date of a patient's admission (weekday vs weekend, summer vs winter) nor the time a blood test was performed (3pm vs 3am).
e) Poor performance: their ability to discriminate between patients likely to live or die is low.

To overcome the limitations of these previous approaches in this thesis, it was necessary to 1) collect a large dataset from multiple diverse hospitals, and 2) to apply advanced analytical techniques that both accounted for multicollinearity and were able to produce highly accurate models.

I first addressed and interpreted the conflicting legal and ethical landscape pertinent to the collection of patient data across multiple hospitals. A thorough understanding of the legal framework was a key component of my success in collecting this multisite linked hospitalised patient dataset. As detailed in Chapter 2, the data were legally categorised as 'data for limited disclosure'. Robust linkage and anonymisation procedures were followed, and assurances were made to each NHS Trust, that the data would only be analysed 1) for the purposes of this study, and 2) by myself, the chief investigator.

I identified common sources for the data which were accessible from multiple NHS trusts, and from these I selected universally collected variables. The data sources were 1) CDS.APC (Commissioning Dataset for Admitted Patient Care) for all administrative data, and 2) 'Blood Test Results' using a combination of hospital LIMS (Laboratory Information Management System) and pathology databases. A lot of work went into acquiring this data, I overcame significant governance and technical challenges (Chapter 2, Section 2.3.1). The challenges continued with the pre-processing of the data, specifically their ingestion into a common database, their transformation to a canonical (standard) form, and their cleaning (identification and removal of incomplete, inaccurate and irrelevant data) (Chapter 2). This entire process took over 2 years.

My dataset captures >8 million admissions across 14 NHS Trusts between 2005 and 2015. Of these, blood tests were performed in hospital for ~2 million admissions. Over 500 million blood test results (both in and out of hospital) were collected for analyses.

My initial analysis focussed on the utility of using simple blood tests, such as urea (Ur) and creatinine (Cr), to identify patients at risk of in-hospital death. I found that whilst Ur and Cr have a non-linear relationship, the Ur:Cr ratio is a useful marker. A low or falling Ur:Cr ratio was associated with a reduced risk of in-hospital mortality. Similarly, a high or rising ratio increased the relative risk of death when compared with a normal Ur:Cr, ranging from 6.4 to 47.9. It was also apparent that, for the same ratio values, the raw values of urea and creatinine non-linearly influenced mortality risk. In addition to correlating with outcome, the Ur:Cr ratio is also a surrogate marker for dehydration; thus, a high or rising ratio also provides a clinical team with insight into a patient's condition, potentially informing their decisions for the ongoing management of the patient. Given my finding that dehydration was present on admission to hospital in 45.5% of patients, and additionally occurred in 28.3% of remaining

patients in-hospital, any means of improving their care could potentially save a large number of lives (Chapter 3).

AKI is a serious syndrome that costs the NHS £1 billion annually in England and Wales alone. I found AKI, as defined using the NHS England Algorithm, to be prevalent in hospitalised patients affecting 8.6% (170,596 / 1,972,130) of all admissions with an in-hospital creatinine result; this figure is lower than those reported previously (14.7% in northern Europe, 20% worldwide). Also, amongst patients with AKI, over one-third (38.6%) of these patients had AKI on admission to hospital (community-acquired AKI). AKI was twice as common in emergency (10.7%) than in planned (5.8%) admissions, and occurs more frequently in patients admitted during the winter (December and January: 9.2%) than in the summer (May to August: 8.4%).

AKI was associated with poor outcome: of the 170,596 of admissions who developed AKI, 21% died or underwent unplanned renal replacement therapy (Drrt) in hospital. I found an increased rate of Drrt in AKI patients with cardiac disease (heart failure: 36.9%, atrial fibrillation: 32.7%, chronic ischaemic heart disease: 26.2%, angina: 24%; or presence of electronic cardiac devices: 28.8%), chronic obstructive pulmonary disease (30.8%) and malignancy (36%). Increasing AKI stages (1 to 3) are designed to risk-stratify AKI patients by worsening outcomes. Although the Drrt rates increased with each AKI stage (AKI_{max} Stage 1: 13.6%; Stage 2: 27.3%; Stage 3: 42.4%), I found significant variation within each AKI stage when these were interrogated in the context of additional variables. For example, the variation in the Drrt rate of a patient diagnosed with AKI_{max} Stage 3 ranged from ~0.5% to ~70% ($p < 0.01$) when the actual creatinine value is considered in addition to the AKI stage (Figure 4.7). Similar variations in stage-specific Drrt rates were seen when patients were grouped by method of admission or by chronic kidney disease status, thus highlighting the inadequacy of the current stage classification in identifying patients at risk of Drrt (Chapter 4).

Analysing the data for all 1.87 million hospitalised patients who had both a full blood count and 'urea and electrolyte' blood tests (i.e., not just those who were dehydrated or who suffered from AKI), I observed an in-hospital mortality rate of 3.14% and a median length of stay of 2 days (interquartile range: 0–6 days). Emergency, elective and maternity admissions accounted for 57.6%, 34.6% and 5.7% of the total respectively. The ethnic category 'British' described the highest proportion of admissions at 42.8%, followed by the 'Not Stated' category at 30.4%. The top three co-morbidities were primary hypertension, type 2 diabetes mellitus, and chronic ischaemic heart disease, at 23%, 10.6% and 9.3% respectively.

In this cohort (Admission_Cohort), mortality by method of admission (Emergency: 4.9%; Elective: 0.4%; and Maternity: 0.07%; $p < 0.01$) substantiates the view that patients admitted as an emergency are significantly more likely to die than those admitted via either the elective or maternity methods. The 'Other' method of admission (the bulk of which were made up of non-emergency inter-hospital transfers) while accounting for only 2.1% of admissions had the highest mortality rate of all, at 9%. The months of admission with the highest mortality were January (3.5%) and December (3.9%). As expected, admissions at the weekend had a higher mortality than weekday admissions (4.3% vs 2.9%, $p < 0.01$) given the lack of elective admissions taking place at a weekend. Thus, mortality for weekend vs weekday admissions was less marked (5.13% vs 4.82%, $p < 0.01$), though still significantly higher, when only emergency admissions were examined. There is currently a major workforce restructuring underway led by the Department of Health and NHS England to increase the provision of doctors available over the weekends, citing that this may mitigate the so called weekend effect. My analyses, highlight that in-hospital mortality varies to a greater degree by the season (rather than the day of the week) of admission. There could be multiple causes for such an increase. Although not specifically analysed, I believe that a rise in acute infective respiratory diseases (e.g. viral and bacterial flu/pneumonia) during the winter months have a detrimental effect. Social isolation of elderly/chronically ill patients may exacerbate their illnesses prior to being admitted to hospital, and once they are admitted, pressure on already stretched hospital resources, may also all have an effect. All of these need to explored in more detail to precisely understand the causes.

I next explored the benefits of assessing patients at different time points in their hospital stay trajectory to better identify those at risk of deterioration in hospital. Patients change during their hospital stay; some get better and are discharged home, while others remain, due to either inadequate improvement or worsening in their condition. Understanding the characteristics of these patients is important to better characterise their risk of deterioration leading to mortality. Of the original 1.87 million admissions, 43.28% (811,268) remained in hospital to undergo a second set of blood tests (Two_Tests_Cohort). The mortality in this group rose from 3.14% to 6.1% (Emergency 7.6%; Elective: 1.5%; and Maternity: 0.04%), and the median length of stay rose from 2 to 6 days. The Two_Test_Cohort patients, when compared to the Admission_Cohort, were also older (median age increased from 60 to 66); comprised of more emergency admissions (prevalence increased from 57.6% to 71.9%, $p < 0.01$); and more patients had co-morbidities (the proportion of patients with hypertension, chronic ischaemic heart disease, heart failure, type 2 diabetes and COPD all increased, by an absolute 6.7%, 4.1%, 3.3%, 4% and 2.7% respectively ($p < 0.01$)).

Having highlighted the importance of variables such as Ur:Cr, dehydration or AKI, co-morbidities, age, method and timing of admission, on a patient's risk of mortality, I then went on to develop a detailed

system (the ML – Early Warning System (ML-EWS)) that enables accurate recognition of hospitalised patients likely to suffer an AE. I used existing and routinely electronically collected patient data: blood test results and administrative (demographics, comorbidity, method of admission, etc.) data. The development comprised two key parts. First, I created various ML models (using the RF and GB techniques), and assessed how effectively they discriminated between patients likely or unlikely to have an poor outcome. Excitingly, I then developed a technology platform called 'Patient Rescue' that can continuously apply these models to patient data as soon as they are generated, and communicate patient specific-reports to clinicians in real time.

The aims of my models were to enable accurate discrimination between patients likely or unlikely to suffer an AE, thus enabling their use as the afferent limb of a rapid response system. The threshold for referral of to a RRT should ideally be based only on the patient's individual risk. However, in a resource-limited environment such as the NHS, efficiency of resource utilisation must also be taken into account. I therefore made an a priori selection of three positive predictive value (PPV) thresholds to test the performance of my models. These PPVs were 1:2 (ML50), 1:3 (ML33) and 1:5 (ML20). A PPV of 1 in 2 means that if the model identifies the patient as likely to die, then one out every two patients will subsequently do so. Similarly, a PPV of 1:5 means that one out of every five patients identified will die in hospital. Thus, the mortality rates of the patients identified by these thresholds are 50% (ML50), 33.33% (ML33) and 20% (ML20).

The four models I built were 1) ML-Dehydration, 2) ML-AKI, 3) ML-Admission, and 4) ML-Two-Tests. The performance of the models increased as additional variables were added. I calculated the area under the receiver operator curve (AUROC), as well as the individual PPV thresholds for each model. Each model is briefly summarised below.

1. **ML-Dehydration** (AUROC: 79.6%): this model was applicable only to emergency admissions, and utilised two types of blood test results (urea and creatinine), along with the most basic demographic data (age and sex). At a PPV of 1:5 (ML20: mortality rate of 20%), the model identified 49.1% (sensitivity) of admissions likely to die (specificity: 86.7%). Thus, if the model (at this threshold) identifies a patient as likely to die, he/she has a 3.1x increased risk of dying relative to the overall mean (the mean mortality in this group of 79,949 admissions was 6.4%).

2. **ML-AKI** (AUROC: 85.9%): this model was built both to replace the existing KDIGO AKI Stages, and also to be used upon a patient's first diagnosis of AKI (AKI_{first}). The mean Drrt rate among AKI admissions was already very high at 21%; hence a PPV of 1:2 (ML50: mortality rate 50%) was used. At this threshold, the model identified almost three out of every four admissions

likely to subsequently suffer from Drrt (sensitivity: 74.3%; specificity: 80.3%). Thus, compared with the AKI mean, a patient identified by my model (at this threshold) had a 2.4 relative risk of Drrt. If this threshold were used to direct escalation of care, as opposed to the existing KDIGO AKI staging, only 31.2% of admissions on first presentation with AKI would trigger the efferent limb (rapid response team or dedicated AKI nephrologist). The ML50 Drrt rate (50%) is also higher than that of any of the existing KDIGO AKI Stages, at AKI_{first} 1: 18.7%; AKI_{first} 2: 25.5%; and AKI_{first} 3: 32.8% (i.e. higher PPV and specificity), while at the same time identifying 1.1, 4.9, and 4.6 times as many patients likely to suffer from Drrt (higher sensitivity), compared with the respective KDIGO AKI Stages.

3. **ML-Admission** (AUROC: 93%): this model was built to be applicable to all hospitalised patients on receipt of their first set of blood results, regardless of method of admission or syndrome (e.g., AKI, etc.). The mean mortality rate of this combined cohort (Admission_Cohort) was 3.14%; thus, the performance of the lowest PPV of 1:5 (ML20: mortality rate 20%) is described. At this threshold, three-quarters of patients likely to die can be identified (sensitivity: 76.1%; specificity: 90.2%). Compared with the cohort mean, a patient identified by the model (at this threshold) has a 6.4 relative risk of dying. Escalating the care of these patients would result in 12% of all admissions being referred to the RRT. This may result in an unacceptably high and possibly inefficient workload, especially since the patient's own clinical team may not have had an opportunity to initiate treatment as the patient had just been admitted. Thus, rather than escalation to a RRT, I have suggested that for these identified admissions, an immediate consultant clinical (or senior clinician) review be undertaken, as opposed to the existing guidelines that arbitrarily recommends that a senior clinical review every hospital admission within 12 hours.

4. **ML-Two-Tests** (AUROC 90.6%): this model was built to track patients who have remained in hospital and have undergone at least two sets of blood tests (Two_Tests_Cohort). At this point, there are 56% fewer patients than the total who were admitted to hospital (Admission_Cohort), and these remaining patients are distinctly different, being older and sicker (mortality rate 6.1%). Additionally, as some time has elapsed between the first and second sets of blood tests, it is reasonable to expect that some treatment or intervention to improve the patient's outcome has been initiated. At this stage, a mortality risk of 33.3% should be unacceptable, and thus the performance of the PPV 1:3 (ML33: mortality rate 33.33%) is highlighted, as it provides the optimal balance between relative risk of death and RRT workload. At this threshold, the model identifies 65% of patients likely to die, each with a relative risk of dying 5.5 times higher than that of the cohort mean (sensitivity: 65%; specificity:

91.5%). Escalating the care of these admissions would result in only 12% of all these patients being referred to the RRT.

The predictive performance of my models surpass those of any current alternative in the published medical literature, and provides an opportunity to substantially improve the recognition of the deteriorating (or at risk of death) patient.

An exciting challenge encountered during the course of this PhD research was exploring the most effective way to apply these advanced patient AE detection models within clinical care, and thus positively affecting patient care. As no such technology platform exists that allows such models to be implemented across diverse hospitals with different clinical care systems, I created my own: '*Patient Rescue'*. Using a cloud-based platform, and complying with all legal and information governance requirements, Patient Rescue was developed by myself, leading a core team comprising a software developer (Simon Brown), an NHS information analyst (Ani Dwarakanath), and a project manager (Prashant Lele). To fund this innovative development, I successfully competed for the Regional Innovation Fund, and was awarded £183,000 from NHS England, for the team and the software. The challenges we faced and our solutions are described in Chapter 6, but in summary, we overcame the live access, analytic and communication problems, and successfully created a modular technology framework based on a mixture of open-source and Microsoft components. *Patient Rescue* was based on a 'Publisher-Subscriber' architecture, and was successfully integrated within three hospitals. By analysing millions of data points, milliseconds after they were generated, *Patient Rescue* generated rich patient-level reports that provided 'actionable insight' to clinicians, as well as hospital-wide dashboards that could be used to track AKI incidence and quality of care. As Patient Rescue was a proof of concept, and we did not have MHRA certification, it was not used to treat any patients, and only implemented the NHSE-AKI algorithm.

Since the start of my PhD, there have been a number of initiatives/solutions, in the UK and internationally, that are tackling the problem of real time integration of disparate healthcare repositories and enabling live analytics. My core team and myself continue to pursue relevant avenues and to forge partnerships for deploying such a system on a large scale. In parallel, one of the NHS trusts in which we tested Patient Rescue has partnered with Google DeepMind, taken the concept of my platform, copied its documentation, and is developing their version. A leader in this space, Enlitic, in 2015 started providing real time clinical decision support to radiologists in Australia, using advanced ML on x-ray images. The most prominent in the UK being Datawell, a Connected Health Cities project of the Northern Health Science Alliance. Datawell builds on a platform developed by LumiraDx, and focusses on connecting multiple providers of healthcare (GP's, hospitals, private providers, etc.) and

enabling patient centred care and analytics. In the USA, multiple hospital groups have built their own data and analytic solutions, and are carrying out machine learning analysis, two particular examples being the Mayo Clinic and the Children's Hospital of Orange County. The Watson Health division within IBM is also building on its cognitive computing solutions to gain insight from existing data, both within the hospital but also from public research repositories (e.g. pubmed). Overall, this is a rapidly advancing field, and I believe in just a few years, it will be commonplace for IT solutions to be augmenting the intelligence of doctors (and other healthcare professionals) and enabling better quality of patient care.

For future work, there are still numerous challenges to be faced. Medicine needs to be more data driven, rather than paying lip service to an evidence based medicine approach, healthcare data needs to be accessible (respecting privacy), to the patient and to healthcare professionals both on an individual basis but also as cohorts. Simple presentation of real time data would help identify issues that need further exploration, such as the unequivocal seasonal variation on mortality. Further down the line, issues relating to the effect of the new models themselves on the healthcare system need to be understood. From an analytic point of view, the actual deployment of any model directly affects its subsequent performance in predicting death in those patients previously identified. This is because in an ideal environment, a prediction of death and subsequent RRT referral should result in those patients not dying, i.e. a 100% successful intervention. However, as highlighted earlier, the 'Chain of Prevention' consists of multiple components; and understanding the influence of each of these components in an environment where complex predictive models are already deployed will require further advances in ML as applied to healthcare. In addition, understanding which information should be communicated to clinicians and what clinical decisions/interventions could be automated is key to the successful deployment of such a system. My models themselves can be improved as more patient data is incorporated. A recommendation for future work would be the addition of vital sign variables, OPCS-4 codes and routinely collected perioperative medicine variables.

In my opinion, one key feature that must be present in any future system is the ability to communicate health status (and risk) directly to the patient. The patient should be a partner in his/her medical care, rather than being merely the recipient of investigations and interventions. *Patient Rescue* has the capability to communicate directly with patients; however, this function was not used, due to both regulatory and clinical concerns. Direct and real-time communication with patients is, I believe, the key factor that will enable full-scale implementation of automated ML in health care.

Overall, my study has significantly advanced the field of detection of patients likely to suffer an SAE in hospital. I have curated a large dataset of UK hospitalised patients, and utilised the most current ML

techniques to create ML models that can be used 1) with limited data, 2) for specific syndromes, and 3) on all hospitalised patients at different time points in their hospital trajectory. I have also demonstrated a modular cloud-based technology platform that enables the implementation of such models and provides actionable insight to clinicians in real time.

During 2015-2016, 283,000 patients died in hospital (or within 30 days of hospitalisation) in England alone. Based on the reports previously described, almost half of these patients received poor quality care at some point in their hospital stay (141,500). This poor quality care is cited as a contributing factor to their ultimate poor outcome (in this case death), in at least half of all these patients. Implementation of such a system of advanced algorithms and real time identification with appropriate intervention, could theoretically save anywhere between 70,750 to 141,500 lives in England alone (the figure probably being closer to 70,750 as some patients would probably be palliative, or no interventions could change outcome). The potential impact, both nationally and internationally, of such an approach is clear and I hope my work motivates others to develop this field and make a significant positive impact on patient quality of care.

In June 2016, I was selected into the first cohort of NHS England's Clinician Entrepreneurs. This unique programme is giving me the opportunity to translate this research into clinical practice, and I am working with a range of partners to fulfil this vision.

8 References

1 *National Patient Safety Agency (NPSA)*. (National Confidential Enquiry into Patient Outcome and Death (NCEPOD), 2005).

2 Wachter, R. *Understanding Patient Safety*. (McGraw-Hill Education / Medical, 2012).

3 Levinson, D. Adverse events in hospitals: national incidence among medicare beneficiaries. Report No. OEI-06-09-00090, (2010).

4 *Knowing the Risk A review of the peri-operative care of surgical patients*. (National Confidential Enquiry into Patient Outcome and Death, 2011).

5 *Adding Insult to Injury: a Review of the Care of Patients who Died in Hospital with a Primary Diagnosis of Acute Kidney Injury (acute renal failure)*. (National Confidential Enquiry into Patient Outcome and Death, 2009).

6 Centre, N. C. G. Intravenous Fluid Therapy- Intravenous fluid therapy in adults in hospital. Report No. CG174, (National Institute for Health and Care Excellence, 2013).

7 NIfHaC, E. NICE clinical guideline 50 Acutely ill patients in hospital: recognition of and response to acute illness in adults in hospital. *London: National Institute for Health and Clinical Excellence* (2007).

8 *An Acute Problem ?* , (National Confidential Enquiry into Patient Outcome and Death, 2005).

9 Martin, I. C. *et al. Emergency Admissions: A Journey in the Right Direction?* , (NCEPOD, 2007).

10 Caring to the End? A review of the care of patients who died in hospital within four days of admissions. (2009).

11 Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of in-hospital cardiorespiratory arrest. (National Confidential Enquiry into Patient Outcome and Death, 2012).

12 Just Say Sepsis! A review of the process of care received by patients with sepsis. (National Confidential Enquiry into Patient Outcome and Death, 2015).

13 Federal Register Volume 68. 68 FR 3454 (U.S. Government Publishing Office, 2003). <https://www.gpo.gov/fdsys/pkg/FR-2003-01-24/content-detail.html>

14 *Commissioning for Quality and Innovation (CQUIN)*, <https://www.england.nhs.uk/nhs-standard-contract/cquin/> (

15 Smith, G. B. In-hospital cardiac arrest: is it time for an in-hospital 'chain of prevention'? *Resuscitation* **81**, 1209-1211, doi:10.1016/j.resuscitation.2010.04.017 (2010).

16 Jones, D. A., DeVita, M. A. & Bellomo, R. Rapid-response teams. *N Engl J Med* **365**, 139-146, doi:10.1056/NEJMra0910926 (2011).

17 *National Early Warning Score (NEWS) Standardising the assessment acute illness severity in the NHS*. (Royal College of Physicians, 2012).

18 (ed Department of Health) (2013).

19 Gardner-Thorpe, J., Love, N., Wrightson, J., Walsh, S. & Keeling, N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl* **88**, 571-575, doi:10.1308/003588406X130615 (2006).

20 Smith, G. B., Prytherch, D. R., Schmidt, P. E., Featherstone, P. I. & Higgins, B. A review, and performance evaluation, of single-parameter "track and trigger" systems. *Resuscitation* **79**, 11-21, doi:10.1016/j.resuscitation.2008.05.004 (2008).

21 Smith, G. B., Prytherch, D. R., Schmidt, P. E. & Featherstone, P. I. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation* **77**, 170-179, doi:10.1016/j.resuscitation.2007.12.004 (2008).

22 Salamonson, Y., Kariyawasam, A., van Heere, B. & O'Connor, C. The evolutionary process of Medical Emergency Team (MET) implementation: reduction in unanticipated ICU transfers. *Resuscitation* **49**, 135-141 (2001).

23 Harrison, G. A., Jacques, T. C., Kilborn, G. & McLaws, M. L. The prevalence of recordings of the signs of critical conditions and emergency responses in hospital wards--the SOCCER study. *Resuscitation* **65**, 149-157, doi:10.1016/j.resuscitation.2004.11.017 (2005).

24 Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E. & Featherstone, P. I. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* **84**, 465-470, doi:10.1016/j.resuscitation.2012.12.016 (2013).
25 Bleyer, A. J. *et al.* Longitudinal analysis of one million vital signs in patients in an academic medical center. *Resuscitation*, doi:10.1016/j.resuscitation.2011.06.033 (2011).
26 Romero-Brufau, S. *et al.* Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation* **85**, 549-552, doi:10.1016/j.resuscitation.2013.12.017 (2014).
27 Ranson, J. H. *et al.* Prognostic signs and the role of operative management in acute pancreatitis. *Surg Gynecol Obstet* **139**, 69-81 (1974).
28 Ozarda, Y. Reference intervals: current status, recent developments and future considerations. *Biochem Med (Zagreb)* **26**, 5-16, doi:10.11613/BM.2016.001 (2016).
29 Grasbeck, R. & Fellman, J. Normal values and statistics. *Scand J Clin Lab Invest* **21**, 193-195 (1968).
30 Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition C28-A3c EP28-A3c* 3/16/2012. Clinical and Laboratory Standards Institute, Wayne, PA, USA
31 Asadollahi, K., Hastings, I. M., Gill, G. V. & Beeching, N. J. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. *Emerg Med Australas* **23**, 354-363, doi:10.1111/j.1742-6723.2011.01410.x (2011).
32 Loekito, E. *et al.* Common laboratory tests predict imminent medical emergency team calls, intensive care unit admission or death in emergency department patients. *Emerg Med Australas* **25**, 132-139, doi:10.1111/1742-6723.12040 (2013).
33 Weinberg, A. D. & Minaker, K. L. Dehydration. Evaluation and management in older adults. Council on Scientific Affairs, American Medical Association. *JAMA* **274**, 1552-1556 (1995).
34 Hooper, L., Bunn, D., Jimoh, F. O. & Fairweather-Tait, S. J. Water-loss dehydration and aging. *Mech Ageing Dev*, doi:10.1016/j.mad.2013.11.009 (2013).
35 González-Alonso, J., Mora-Rodríguez, R., Below, P. R. & Coyle, E. F. Dehydration reduces cardiac output and increases systemic and cutaneous vascular resistance during exercise. *J Appl Physiol (1985)* **79**, 1487-1496 (1995).
36 Popkin, B. M., D'Anci, K. E. & Rosenberg, I. H. Water, hydration, and health. *Nutr Rev* **68**, 439-458, doi:10.1111/j.1753-4887.2010.00304.x (2010).
37 Kelly, J. *et al.* Dehydration and venous thromboembolism after acute stroke. *QJM* **97**, 293-296 (2004).
38 Briongos Figuero, S. *et al.* Dehydration and serum hyperosmolarity as new predictors of mortality after acute coronary syndrome. *Int J Cardiol*, doi:10.1016/j.ijcard.2014.01.033 (2014).
39 Lin, H.-L. *et al.* High preoperative ratio of blood urea nitrogen to creatinine increased mortality in gastrointestinal cancer patients who developed postoperative enteric fistulas. *Kaohsiung J Med Sci* **28**, 418-422, doi:10.1016/j.kjms.2012.02.011 (2012).
40 Beier, K. *et al.* Elevation of blood urea nitrogen is predictive of long-term mortality in critically ill patients independent of "normal" creatinine. *Crit Care Med* **39**, 305-313, doi:10.1097/CCM.0b013e3181ffe22a (2011).
41 Mortensen, P. B., Nohr, M., Moller-Petersen, J. F. & Balslev, I. The diagnostic value of serum urea/creatinine ratio in distinguishing between upper and lower gastrointestinal bleeding. A prospective study. *Danish Medical Bulletin* **41**, 237-240 (1994).
42 Witting, M. D. *et al.* ED predictors of upper gastrointestinal tract bleeding in patients without hematemesis. *Am J Emerg Med* **24**, 280-285, doi:10.1016/j.ajem.2005.11.005 (2006).
43 Beaufrere, B. *et al.* Glucocorticosteroids increase leucine oxidation and impair leucine balance in humans. *The American journal of physiology* **257**, E712-721 (1989).
44 Combe, C. *et al.* Influence of nutritional factors and hemodialysis adequacy on the survival of 1,610 French patients. *Am J Kidney Dis* **37**, S81-88 (2001).

45 Cartin-Ceba, R., Afessa, B. & Gajic, O. Low baseline serum creatinine concentration predicts mortality in critically ill patients independent of body mass index. *Crit Care Med* **35**, 2420-2423 (2007).
46 Lieberman, H. R. Hydration and cognition: a critical review and recommendations for future research. *J Am Coll Nutr* **26**, 555S-561S (2007).
47 Maughan, R. J. Impact of mild dehydration on wellness and on exercise performance. *Eur J Clin Nutr* **57 Suppl 2**, S19-23, doi:10.1038/sj.ejcn.1601897 (2003).
48 Armstrong, L. E. *et al.* Mild dehydration affects mood in healthy young women. *J Nutr* **142**, 382-388, doi:10.3945/jn.111.142000 (2012).
49 Maughan, R. J. Hydration, morbidity, and mortality in vulnerable populations. *Nutr Rev* **70 Suppl 2**, S152-155, doi:10.1111/j.1753-4887.2012.00531.x (2012).
50 Murray, B. Hydration and physical performance. *J Am Coll Nutr* **26**, 542S-548S (2007).
51 Namdar, T. *et al.* Impact of hypernatremia on burn wound healing: results of an exploratory, retrospective study. *Ostomy Wound Manage* **57**, 30-34 (2011).
52 Kateros, K. *et al.* Analysis of kidney dysfunction in orthopaedic patients. *BMC Nephrol* **13**, 101, doi:10.1186/1471-2369-13-101 (2012).
53 Uchino, S., Bellomo, R. & Goldsmith, D. The meaning of the blood urea nitrogen/creatinine ratio in acute kidney injury. *Clinical Kidney Journal* **5**, 187-191, doi:10.1093/ckj/sfs013 (2012).
54 Oksa, H., Pasternack, A. & Pasanen, M. Serum urea-creatinine ratio as a prognostic index in hemodialysis patients. *Clinical Nephrology* **27**, 125-130 (1987).
55 De Lima, J. J., da Fonseca, J. A. & Godoy, A. D. Baseline variables associated with early death and extended survival on dialysis. *Renal Failure* **20**, 581-587 (1998).
56 Lin, L. C. *et al.* Predictors of early clinical deterioration after acute ischemic stroke. *Am J Emerg Med* **29**, 577-581, doi:10.1016/j.ajem.2009.12.019 (2011).
57 Schrock, J. W., Glasenapp, M. & Drogell, K. Elevated blood urea nitrogen/creatinine ratio is associated with poor outcome in patients with ischemic stroke. *Clin Neurol Neurosurg* **114**, 881-884, doi:10.1016/j.clineuro.2012.01.031 (2012).
58 Rowat, A., Graham, C. & Dennis, M. Dehydration in hospital-admitted stroke patients: detection, frequency, and association. *Stroke* **43**, 857-859, doi:10.1161/STROKEAHA.111.640821 (2012).
59 Feinfeld, D. A., Bargouthi, H., Niaz, Q. & Carvounis, C. P. Massive and disproportionate elevation of blood urea nitrogen in acute azotemia. *International urology and nephrology* **34**, 143-145 (2002).
60 Gotsman, I. *et al.* The significance of serum urea and renal function in patients with heart failure. *Medicine* **89**, 197-203 (2010).
61 Aronson, D. *et al.* Serum blood urea nitrogen and long-term mortality in acute ST-elevation myocardial infarction. *International Journal of Cardiology* **127**, 380-385 (2008).
62 Stookey, J. D., Pieper, C. F. & Cohen, H. J. Is the prevalence of dehydration among community-dwelling older adults really low? Informing current debate over the fluid recommendation for adults aged 70+years. *Public Health Nutrition* **8**, 1275-1285 (2005).
63 Mukand, J. A., Cai, C., Zielinski, A., Danish, M. & Berman, J. The effects of dehydration on rehabilitation outcomes of elderly orthopedic patients. *Arch Phys Med Rehabil* **84**, 58-61, doi:10.1053/apmr.2003.50064 (2003).
64 Fine, M. J. *et al.* A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* **336**, 243-250, doi:10.1056/NEJM199701233360402 (1997).
65 Lim, W. S. *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* **58**, 377-382 (2003).
66 Wu, B. U. *et al.* Blood urea nitrogen in the early assessment of acute pancreatitis: an international validation study. *Arch Intern Med* **171**, 669-676, doi:10.1001/archinternmed.2011.126 (2011).
67 Hartz, A. J., Kuhn, E. M., Kayser, K. L. & Johnson, W. D. BUN as a risk factor for mortality after coronary artery bypass grafting. *The Annals of thoracic surgery* **60**, 398-404 (1995).

68 Fonarow, G. C. *et al.* Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA* **293**, 572-580, doi:10.1001/jama.293.5.572 (2005).
69 Commission, C. Q. Dignity and nutrition inspection programme: national overview. *Newcastle upon Tyne: Care Quality Commission* (2011).
70 Association, P. Listen to patients, Speak up for change. *Harrow: Patients Association* (2010).
71 Parliamentary & Ombudsman, H. S. *Care and Compassion?: Report of the Health Service Ombudsman on Ten Investigations Into Nhs Care of Older People*. Vol. 778 (The Stationery Office, 2011).
72 Francis, R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary*. Vol. 947 (TSO Shop, 2013).
73 Feinmann, J. *How hundreds of patients are dying of thirst in hospital: The story of a 22-year-old who died in hospital from dehydration shocked Britain. But his tragedy is horrifyingly common*, <http://www.dailymail.co.uk/health/article-2177969/How-hundreds-patients-dying-thirst-hospital-The-story-22-year-old-died-hospital-dehydration-shocked-Britain-But-tragedy-horrifyingly-common.html> (2012).
74 Borland, S. *Thousands dying of thirst on NHS: Watchdog forced to issue guidelines on giving patients water*, <http://www.dailymail.co.uk/news/article-2403501/Thousands-dying-thirst-NHS-Watchdog-forced-issue-guidelines-giving-patients-water.html> (2013).
75 Hodgekiss, A. *1,000 patients a month are dying in hospital from 'avoidable' kidney problems caused by dehydration*, <http://www.dailymail.co.uk/health/article-2609542/1-000-patients-month-dying-hospital-avoidable-kidney-problems-caused-dehydration.html> (2014).
76 Ellicott, C. *Patient dying of thirst rang 999: Inquest hears of mother's fury at nurses who neglected son*, <http://www.dailymail.co.uk/news/article-2167643/Patient-dying-thirst-rang-999-Inquest-hears-mothers-fury-nurses-neglected-son.html> (2014).
77 Borland, S. *Four elderly patients a week are dying of hunger and thirst 'utterly scandalous' figures reveal*, <http://www.dailymail.co.uk/news/article-2516692/Four-elderly-patients-week-dying-hunger-thirst-utterly-scandalous-figures-reveal.html> (2013).
78 (GDG), G. d. g. (ed National Institute for Health and Care Excellence (NICE)) (National Clinical Guideline Centre, http://guidance.nice.org.uk/CG169, 2013).
79 Kerr, M., Bedford, M., Matthews, B. & O'Donoghue, D. The economic impact of acute kidney injury in England. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, doi:10.1093/ndt/gfu016 (2014).
80 England, N. *Hospital Episode Statistics, Admitted Patient Care, England - 2012-13 [NS]*, <http://www.hscic.gov.uk/searchcatalogue?productid=13264&q=title%3a%22Hospital+Episode+Statistics%2c+Admitted+patient+care+-+England%22&sort=Relevance&size=10&page=1#top> (2013).
81 Crary, M. A. *et al.* Dysphagia, nutrition, and hydration in ischemic stroke patients at admission and discharge from acute care. *Dysphagia* **28**, 69-76, doi:10.1007/s00455-012-9414-0 (2013).
82 Ftouh, S., Thomas, M. & Acute Kidney Injury Guideline Development, G. Acute kidney injury: summary of NICE guidance. *BMJ* **347**, f4930, doi:10.1136/bmj.f4930 (2013).
83 KDIGO, A. K. I. Work Group: KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl* **2**, 1-138 (2012).
84 Susantitaphong, P. *et al.* World incidence of AKI: a meta-analysis. *Clinical journal of the American Society of Nephrology : CJASN* **8**, 1482-1493, doi:10.2215/CJN.00710113 (2013).
85 Wang, H. E., Muntner, P., Chertow, G. M. & Warnock, D. G. Acute kidney injury and mortality in hospitalized patients. *Am J Nephrol* **35**, 349-355, doi:10.1159/000337487 (2012).
86 Aitken, E. *et al.* Acute kidney injury: outcomes and quality of care. *QJM* **106**, 323-332, doi:10.1093/qjmed/hcs237 (2013).
87 Metcalfe, W. *et al.* Acute renal failure requiring renal replacement therapy: incidence and outcome. *QJM* **95**, 579-583 (2002).

88 Chawla, L. S., Amdur, R. L., Amodeo, S., Kimmel, P. L. & Palant, C. E. The severity of acute kidney injury predicts progression to chronic kidney disease. *Kidney Int* **79**, 1361-1369, doi:10.1038/ki.2011.42 (2011).
89 Selby, N. M., Hill, R., Fluck, R. J. & Programme, N. H. S. E. T. K. A. Standardizing the Early Identification of Acute Kidney Injury: The NHS England National Patient Safety Alert. *Nephron* **131**, 113-117, doi:10.1159/000439146 (2015).
90 Moonesinghe, S. R., Mythen, M. G., Das, P., Rowan, K. M. & Grocott, M. P. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. *Anesthesiology* **119**, 959-981, doi:10.1097/ALN.0b013e3182a4e94d (2013).
91 Prytherch, D. R. *et al.* The use of routine laboratory data to predict in-hospital death in medical admissions. *Resuscitation* **66**, 203-207, doi:10.1016/j.resuscitation.2005.02.011 (2005).
92 Hucker, T. R. *et al.* Identifying the sick: can biochemical measurements be used to aid decision making on presentation to the accident and emergency department. *Br J Anaesth* **94**, 735-741, doi:10.1093/bja/aei122 (2005).
93 Silke, B., Kellett, J., Rooney, T., Bennett, K. & O'Riordan, D. An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. *QJM* **103**, 23-32, doi:10.1093/qjmed/hcp149 (2010).
94 Loekito, E. *et al.* Common laboratory tests predict imminent death in ward patients. *Resuscitation*, doi:10.1016/j.resuscitation.2012.07.025 (2012).
95 Mohammed, M. A. *et al.* Which is more useful in predicting hospital mortality--dichotomised blood test results or actual test values? A retrospective study in two hospitals. *PLoS One* **7**, e46860, doi:10.1371/journal.pone.0046860 (2012).
96 Jarvis, S. W. *et al.* Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation* **84**, 1494-1499, doi:10.1016/j.resuscitation.2013.05.018 (2013).
97 Baxmann, A. C. *et al.* Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. *Clinical journal of the American Society of Nephrology : CJASN* **3**, 348-354, doi:10.2215/CJN.02870707 (2008).
98 Kalyani, R. R., Corriere, M. & Ferrucci, L. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. *Lancet Diabetes Endocrinol* **2**, 819-829, doi:10.1016/S2213-8587(14)70034-8 (2014).
99 Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development* **3**, 210-229 (1959).
100 Mitchell, T. M. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill* **45**, 37 (1997).
101 *Enlitic*, <www.enlitic.com> (2017).
102 Ho, T. K. in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on.* 278-282 (IEEE).
103 Breiman, L. Random forests. *Machine learning* **45**, 5-32 (2001).
104 C Click, M. M., V Parmar, H Roark, A Candel. (2016).
105 *h2o*, <www.h2o.ai>
106 HSCIC. *Commissioning Data Sets*, <http://www.datadictionary.nhs.uk/web_site_content/cds_supporting_information/commissioning_data_sets_overview.asp?shownav=1> (2016).
107 Nouraei, S. *et al.* A study of clinical coding accuracy in surgery: implications for the use of administrative big data for outcomes management. *Annals of surgery* **261**, 1096-1107 (2015).
108 England, N. (ed National Pathology Programme) (NHS England, 2014).
109 Regenstrief Institute, I. *Logical Observation Identifiers Names and Codes*, <http://loinc.org/background> (2016).
110 HSCIC. *National Laboratory Medicine Catalogue (NLMC)* <http://systems.hscic.gov.uk/pathology/projects/nlmc> (2016).
111 McDonald, C. J. *et al.* LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry* **49**, 624-633 (2003).

112 Association, W. M. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama* **310**, 2191 (2013).
113 Health, D. o. (ed Department of Health) 54 (2005).
114 Authority, H. R. *Research requiring NHS review but not ethical review*, <http://www.hra.nhs.uk/resources/before-you-apply/research-requiring-nhs-rd-review-but-not-ethical-review/> (
115 Health, D. o. (ed Department of Health) (www.dh.gov.uk, 2011).
116 ICO. (ed Information Commissioner's Office) (https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/, 2012).
117 The Caldicott Committee - Chair: Dame Caldicott, F. (ed Department of Health) 137 (http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4068403, 1997).
118 Caldicott, F. (ed National Data Guardian) 139 (https://www.gov.uk/government/publications/the-information-governance-review, 2013).
119 Guardian, N. D. (ed Department of Health) (https://www.gov.uk/government/consultations/the-role-of-the-national-data-guardian-for-health-and-social-care, 2015).
120 Carter, P., Laurie, G. T. & Dixon-Woods, M. The social licence for research: why care. data ran into trouble. *Journal of medical ethics*, medethics-2014-102374 (2015).
121 *NHS Care.data PR Fiasco Continues as Google Pulls Out of Secret Deal*, <http://www.infosecurity-magazine.com/news/nhs-caredata-pr-fiasco-continues/> (
122 Martin, A. J. *Care.data is a complete omnishambles, says study into hated scheme*, <http://www.theregister.co.uk/2015/08/12/caredata_case_study/> (
123 medconfidential.org. *The Care.data shambles*, <https://medconfidential.org/> (
124 Ingraham, A. M., Cohen, M. E., Raval, M. V., Ko, C. Y. & Nathens, A. B. Comparison of hospital performance in emergency versus elective general surgery operations at 198 hospitals. *J Am Coll Surg* **212**, 20-28.e21, doi:10.1016/j.jamcollsurg.2010.09.026 (2011).
125 Commission, C. Q. *Annual Health Check Report 2008-09.xlsx*, 2009).
126 *dr foster intelligence*, <http://myhospitalguide.drfosterintelligence.co.uk/#/mortality> (2014).
127 *Admission Method: NHS Data Dictionary*, <(http://www.datadictionary.nhs.uk/data_dictionary/attributes/a/add/admission_method_de.asp)>
128 Yoo, W. *et al.* A study of effects of multicollinearity in the multivariable analysis. *International journal of applied science and technology* **4**, 9 (2014).
129 *Jeffreys Interval*, <https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval> (
130 Porter, C. J. *et al.* A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, doi:10.1093/ndt/gfu082 (2014).
131 Barochia, A. V. *et al.* Bundled care for septic shock: an analysis of clinical trials. *Crit Care Med* **38**, 668-678, doi:10.1097/CCM.0b013e3181cb0ddf (2010).
132 Westphal, G. A. *et al.* Reduced mortality after the implementation of a protocol for the early detection of severe sepsis. *J Crit Care* **26**, 76-81, doi:10.1016/j.jcrc.2010.08.001 (2011).
133 Pestana, D. *et al.* Compliance with a sepsis bundle and its effect on intensive care unit mortality in surgical septic shock patients. *J Trauma* **69**, 1282-1287, doi:10.1097/TA.0b013e3181c4539f (2010).
134 Chen, Q. H. *et al.* [The impact of sepsis bundles on mortality in patients with sepsis shock: a prospective clinical study]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue* **20**, 534-537 (2008).
135 Castellanos-Ortega, A. *et al.* Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study. *Crit Care Med* **38**, 1036-1043, doi:10.1097/CCM.0b013e3181d455b6 (2010).

136 *Discharge destination. NHS Data Dictionary*, <http://www.datadictionary.nhs.uk/data_dictionary/attributes/d/disc/discharge_destination_de.asp?shownav=1?query=%22discharge+destination%22&rank=100&shownav=1> (2017).
137 Association, T. R. *CKD Stages*, <http://www.renal.org/information-resources/the-uk-eckd-guide/ckd-stages> (2016).
138 Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, 604-612 (2009).
139 Harasymiv, V. Lessons from 2 Million Machine Learning Models on Kaggle.
140 Friedman, J. H. Greedy function approximation: A gradient boosting machine. 1189-1232, doi:10.1214/aos/1013203451 (2001).
141 Yang, L. *et al.* Acute kidney injury in China: a cross-sectional survey. *Lancet* **386**, 1465-1471, doi:10.1016/S0140-6736(15)00344-X (2015).
142 Challiner, R., Ritchie, J. P., Fullwood, C., Loughnan, P. & Hutchison, A. J. Incidence and consequence of acute kidney injury in unselected emergency admissions to a large acute UK hospital trust. *BMC Nephrol* **15**, 84, doi:10.1186/1471-2369-15-84 (2014).
143 Roberts, G. *et al.* Acute kidney injury risk assessment at the hospital front door: what is the best measure of risk? *Clin Kidney J* **8**, 673-680, doi:10.1093/ckj/sfv080 (2015).
144 Pankhurst, T. *et al.* Acute kidney injury following unselected emergency admission: role of the inflammatory response, medication and co-morbidity. *Nephron Clin Pract* **126**, 81-89, doi:10.1159/000357845 (2014).
145 Mehta, R. L. *et al.* Recognition and management of acute kidney injury in the International Society of Nephrology 0by25 Global Snapshot: a multinational cross-sectional study. *Lancet* **387**, 2017-2025, doi:10.1016/S0140-6736(16)30240-9 (2016).
146 *HL7*, <www.hl7.org>
147 Dixon, J. *Pentaho, Hadoop, and Data Lakes*, <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/> (2010).
148 Woods, D. *Big data requires a big architecture*, <http://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/> (2011).
149 *Mirth*, <https://www.mirth.com/>
150 *Rabbit MQ*, <https://www.rabbitmq.com/>

-

Appendices

Appendix 1: Table: Commissioning Data Sets

Table: Commissioning Data Sets [106]	
Commissioning Data Set	***Overview***
Accident and Emergency Commissioning Data Set Type	
CDS V6-2 Type 010 - Accident and Emergency CDS	*An Accident and Emergency Attendance is an individual visit by one PATIENT to an Accident and Emergency Department to receive treatment from the accident and emergency service.*
Out Patient Commissioning Data Set Types	
CDS V6-2 Type 020 - Outpatient CDS	*Carries the data for an Outpatient Attendance or a cancelled/missed APPOINTMENT. It covers all NHS and private Outpatient ACTIVITY taking place in any: acute, community, mental health NHS Trust or NHS Foundation Trust, other NHS hospital, non-NHS hospitals or institutions where the care delivered is NHS-funded; under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
Admitted Patient Care Commissioning Data Set Types	
CDS V6-2 Type 120 - Admitted Patient Care - Finished Birth Episode CDS	*Carries the data for a Finished Birth Episode. This is required when a delivery has resulted in a REGISTRABLE BIRTH which has taken place in either an NHS Hospital or in a non-NHS ORGANISATION funded by the NHS. The information is taken from the birth notification for each baby born.*
CDS V6-2 Type 130 - Admitted Patient Care - Finished General Episode CDS	*Carries the data for a Finished General Episode. It covers all NHS and private Admitted Patient Care (day case and inpatient) ACTIVITY taking place in any: acute, community, mental health NHS Trust or NHS Foundation Trust, other NHS hospital, non-NHS hospitals or institutions where the care delivered is NHS-funded; under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
CDS V6-2 Type 140 - Admitted Patient Care - Finished Delivery Episode CDS	*Carries the data for a Finished Delivery Episode which is required when a delivery has resulted in a REGISTRABLE BIRTH. This may take place in either NHS Hospitals or in non-NHS ORGANISATIONS funded by the NHS. The information is taken from the birth notification for each baby born.*
CDS V6-2 Type 150 - Admitted Patient Care - Other Birth Event CDS	*Applies to: NHS funded home births and all other birth events which are not NHS-funded, either directly or under an NHS SERVICE AGREEMENT. The data in these records originates from birth notification records and requires only a limited data set to be completed*
CDS V6-2 Type 160 - Admitted Patient Care - Other Delivery Event CDS	*Applies to: NHS funded home deliveries and all other delivery events which are not NHS-funded, either directly or under an NHS SERVICE AGREEMENT. The data in these records originates from birth notification records and requires only a limited data set to be completed.*
CDS V6-2 Type 170 - Admitted Patient Care - Detained and/or Long Term Psychiatric Census CDS	*a record for every PATIENT admitted as at 31 March each year for which the PATIENT is detained or the Episode is part of a Hospital Provider Spell which has lasted longer than one year and for which the majority of time has been spent under the care of a CONSULTANT in one of the psychiatric specialties.*

CDS V6-2 Type 180 - Admitted Patient Care - Unfinished Birth Episode CDS	*Carries the data for an Unfinished Birth Episode. This is required when a delivery has resulted in a REGISTRABLE BIRTH which has taken place in either an NHS Hospital or in a non-NHS ORGANISATION funded by the NHS. The information is taken from the birth notification for each baby born. Unfinished Birth Episode Commissioning Data Set records are required for all Unfinished Birth Episodes as at midnight on 31st March each year*
CDS V6-2 Type 190 - Admitted Patient Care - Unfinished General Episode CDS	*Carries the data for an Unfinished General Episode. It covers all NHS and private Admitted Patient Care (day case and inpatient) ACTIVITY taking place in any: acute, community, mental health NHS Trust or NHS Foundation Trust, other NHS hospital,* *non-NHS hospitals or institutions where the care delivered is NHS-funded; under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
CDS V6-2 Type 200 - Admitted Patient Care - Unfinished Delivery Episode CDS	*Carries the data for an Unfinished Delivery Episode. This may take place in either NHS Hospitals or in non-NHS ORGANISATIONS funded by the NHS. The information is taken from the birth notification for each baby born. Unfinished Birth and Delivery Episode Commissioning Data Set records are required for all Unfinished Birth and Delivery Episodes as at midnight on 31 March each year.*
Elective Admission List Commissioning Data Set Types - End Of Period Census Types	
CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) CDS	*Carries the data for all booked, planned and waiting list admissions. This consists of records for PATIENTS waiting for Elective Admission at a specified date and should be sent to the Secondary Uses Service within one month of the end of the period to which they relate unless a shorter time-scale has been agreed with the recipient. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
CDS V6-2 Type 040 - Elective Admission List - End Of Period Census (Old) CDS	*Is used to report to the previous (old) Commissioner that the ELECTIVE ADMISSION LIST ENTRY is now the responsibility of another Commissioner. This CDS Type should be sent within one month of the end of the period to which it relates unless a shorter time-scale has been agreed with the recipient. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
CDS V6-2 Type 050 - Elective Admission List - End Of Period Census (New) CDS	*Is used to report to a new Commissioner an ELECTIVE ADMISSION LIST ENTRY that had previously been the responsibility of another Commissioner. This CDS Type should be sent within one month of the end of the period to which it relates unless a shorter time-scale has been agreed with the recipient. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists.*
Elective Admission List Commissioning Data Set Types - Event During Period Types	
CDS V6-2 Type 060 - Elective Admission List - Event During Period (Add) CDS	*Is used to make an initial report that an ELECTIVE ADMISSION LIST ENTRY has been added to the Health Care Provider's ELECTIVE ADMISSION LIST. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.*
CDS V6-2 Type 070 - Elective Admission List - Event During Period (Remove) CDS	*Is used to report that the ELECTIVE ADMISSION LIST ENTRY has been removed from the Health Care Provider's ELECTIVE ADMISSION LIST. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are*

	transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.
CDS V6-2 Type 080 - Elective Admission List - Event During Period (Offer) CDS	*Is used to report that an OFFER OF ADMISSION has been made to the PATIENT. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.*
CDS V6-2 Type 090 - Elective Admission List - Event During Period (Available or Unavailable) CDS	*Is used to report changes in the PATIENT's availability for treatment. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.*
CDS V6-2 Type 100 - Elective Admission List - Event During Period (Old Service Agreement) CDS	*Is used to report to the previous Commissioner that the ELECTIVE ADMISSION LIST ENTRY is now the responsibility of a new Commissioner. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.*
CDS V6-2 Type 110 - Elective Admission List - Event During Period (New Service Agreement) CDS	*Is used to make an initial report to a new Commissioner of an ELECTIVE ADMISSION LIST ENTRY that had previously been the responsibility of another Commissioner. It covers ELECTIVE ADMISSION LIST ENTRIES under the care of a CONSULTANT, MIDWIFE or NURSE, where an appropriate MAIN SPECIALTY CODE and TREATMENT FUNCTION CODE exists. Elective Admission List Event During Period CDS Types are intended for those ORGANISATIONS who have the capability to implement transaction-based processing, and are transmitted using Net Change Commissioning Data Set Submission Protocol. They should be supplemented where required by an annual (or other agreed time-cycle) submission of the CDS V6-2 Type 030 - Elective Admission List - End of Period Census (Standard) Commissioning Data Set.*

Appendix 2: ML-EWS Research Protocol

Machine Learning Analysis of

Electronically Stored Blood Test Results

Chief Investigator

Dr Vishal Nangalia

Supervisors

Dr Reecha Sofat (UCL Division of Medicine)

Dr David Barber (Reader, UCL Department of Computer Science)

Collaborators

Dr Tony Wolff (Consultant, Barnet and Chase Farm Hospital)

Sponsor

University College London

Reference No: 13/0407

Protocol Number: 1.0

Protocol Version: 2.0

Date: 9th August 2013

Summary

The aim of this study is to employ Machine Learning methods of analysis to predict in-hospital deterioration of patients.

Patients who deteriorate, may face the need for emergency resuscitation, surgery, intensive care unit admission or if all else fails they may die. Stopping this happening requires identifying those at risk as early as possible. I aim therefore to develop a computer based system, that is constantly analysing individual patient data to identify these patients prior to them deteriorating. Implementation of a working system (not part of this study), will help clinicians make the right decisions on further investigations, referral to specialist teams and specific treatment earlier than what currently occurs.

The study will analyse anonymised data from all patients who have been admitted and subsequently discharged from a range of hospitals. The data (without patient name, date of birth, address, hospital number, etc) will be extracted from existing hospital databases, and will comprise electronic patient administration information (age, sex, illnesses, etc) and electronically stored blood test results. The information teams at each individual NHS trust will be responsible for providing the anonymised data from their respective existing databases.

Background Information

Serious adverse events (SAEs, such as cardiac arrest, unexpected intensive care unit admission, the need for emergency surgery, or death) are commonplace in hospitalised patients, and reducing these is a national priority[1-5]. Physiological deterioration generally precedes in-

hospital SAEs[1-5]. Recognition of such decline, with appropriate intervention, can reduce SAE frequency6 and is broadly advocated by National Institute of Clinical Excellence (NICE)[2], National Confidential Enquiry into Patient Outcome and Death (NCEPOD)3-5 and The Royal College of Physicians (RCP)[1]. In response, the RCP developed an aggregated scoring system of vital signs: the NHS Early Warning Score (NEWS)1 (based on the ViEWS[7] system). I have previously shown that applying novel multivariate approaches to this dataset can improve sensitivity for detecting SAEs by 5% (while conserving specificity) for each trigger value of NEWS. This approach, though, is inherently limited by the use of 'snapshot' vital signs, without reference to separate prospective risk classification. As such, a patient is only identified as being 'at risk' when physiological derangement is already causing significant compromise- a point at which the window of opportunity for effective clinical intervention may already be closing. Here, I propose a research programme focused on a new and complementary approach.

Haematological tests are routinely conducted in hospital patients both on and during admission. Their use might, in theory, offer a universal and simple means of stratifying risk or predicting decline, that would require little or no significant additional resource. Efforts have thus been made to develop automatic interpretation and prediction systems based on blood results [8-12]. However, their scope has been limited by the use of conventional analyses of small single-centre datasets of single-timepoint blood tests. Furthermore, none use dynamic change over time in blood results, marry such data with co-morbidities (ICD codes) or modify their predictions when new procedures are performed on the patient: yet, as clinicians, we take all these factors into account every time we interpret, request and repeat a blood test.

I therefore hypothesise that serious adverse events (and other indicators of morbidity) will be predictable by applying novel multivariate analyses to large scale time-series data of hospital inpatient blood tests, married with data relating to co-morbidities. My preliminary data demonstrate that using reference ranges to dichotomise routine blood test results into 'normal' and 'abnormal' groups is less accurate in predicting in-hospital mortality than using actual (continuous) test values [12].

I propose to further develop and refine these algorithms, whilst concurrently developing entirely new machine learning models to significantly enhance predictive power, incorporating both multivariate and time series analyses.

Study Aims and Objectives

Aim

To develop a computer programme that analyses routinely collected electronic patient data to identify those patients at risk of deterioration.

Objectives

1. To investigate the mortality associated with blood results within the normal reference range and determine those interrelationships that represent minimal mortality i.e. optimal health.
2. To identify clusters of patients with above-average mortality and characterize the specific relationships between their blood results and their associated clinical conditions.
3. To map the temporal pattern of biochemical and haematological derangement, in specific clusters of patients, that precede serious adverse events.
4. To build and validate a computer-based inference system that predicts serious adverse events, in all hospitalized patients.

Study Design

Investigator

Dr Vishal Nangalia will be the chief investigator and will be supervised by Dr Reecha Sofat and Dr David Barber.

Local Collaborators

- A local collaborator will be identified at each individual NHS trust. The local collaborators will normally be from the Biochemistry and/or Intensive Care Departments.

The Data

- The anonymised data (without patient name, date of birth, address, hospital number, etc) will be extracted from existing hospital databases, and will comprise electronic patient administration information (age, sex, illnesses, etc) and biological data (e.g. blood results).
- Data anonymisation will be confirmed prior to the investigators receiving the data. All data will be anonymised by the respective organisations providing the data. The researchers will therefore not have access to any patient identifiable data.
- The anonymised data that will be analysed is all in electronic format on existing hospital databases and is collected as part of the routine care of patients.

Data Collection

- A local collaborator will be identified at each individual NHS trust. The local collaborators will normally be from the Biochemistry and Intensive Care Departments.
- The data extraction and anonymisation will be carried out by the information teams at the respective NHS trusts using existing hospital systems. The process is simple and straightforward within existing hospital IT infrastructure.
- Each NHS Trust's Research and Development, Information Governance and Caldicott Guardian teams will approve the anonymised data extraction.
- No patient identifiable data will leave the NHS trust or be shared with the researchers.
- The data will be collected by the Chief Investigator from each individual trust, using a NHS approved encrypted USB stick "IRONKEY".

Data Cleaning

- Once the anonymised dataset is obtained, missing and abnormal values will be removed under the guidance of the project team, who have relevant expertise.

Data linkage

- We will not attempt to link the data obtained from the various organisations with any other research database.

Data Sharing

- The data will not be shared with any other investigators.
- The data will be used solely for this project.

Data Storage and Analysis

- Appropriate computational technologies will be used to store and handle the data, and all will be compliant (in terms of security and so forth) with data protection needs.

Study Population

- All adult patients admitted to hospital, who have had a blood test, will be included in the database analysis.

Inclusion Criteria

- The total sample size will consist of all patients who have at some point been admitted to hospital and subsequently discharged from hospital and who have had a blood test.
- Age>=18 years old.

Exclusion Criteria

- Patients not admitted to hospital.
- Patients who have not had a blood test.
- Age<18 years old.

Sample Size

- 500,000 to 1 million anonymised patient admission episodes only.
- The investigational plan has been co-written with world leading statistical experts in the field, and has been reviewed by MRC experts and found valid. This Machine Learning approach relies on very large datasets, of the magnitude described.

Analysis

The data to be interrogated consist of millions of data points, distributed over a high number of dimensions. The dimensions themselves represent a mixture of discrete and continuous variables, some of which have temporal components. This volume and complexity requires a combination of three methodologies to achieve the inferential goals:

1. Building on my preliminary service evaluation/audit highlighting the problems with the current "normal range", I shall use probabilistic pattern-recognition (*supervised learning*) to identify the inter-relationships between routine haematological/biochemical data that result in minimum mortality and morbidity. Patients will be subdivided into demographic-, co-morbidity- and

intervention-based groups. Support Vector Machines will then be used for classification and regression.

2. Utilising neural networks (*unsupervised learning*) and the subsequent identified clusters, I will determine the relationship of the co-morbidities/diagnoses (ICD codes) and interventions (procedure codes) to these clusters and their association with mortality.
3. My next two objectives are to determine if there are patterns in the repeated blood results and associated data that are highly predictive of serious adverse events. This requires a different class of techniques than standard abnormality detection: *latent models with time series analysis.* A simple approach is to take along with each measurement, v_t, the end-class label, c_T, and attempt to form a classifier of the end-time class based on the measurements at time t. This is conceptually straightforward but suffers from being potentially inconsistent – that is, the method may predict that the patient is likely to die; yet based on subsequent results, may predict the patient is likely to be discharged. I therefore need a way to make temporally consistent predictions. I will achieve this by initially dividing the datasets into the previously identified clusters and then use a generative model, (Hidden Markov Model), of blood results $v_{1:t}$ with the outcome label c_T . Non-standard approaches will be utilised to train these models. In addition, separate models will be based on a continuous latent h(t) to track the 'health' of the patient. At all stages, medical knowledge will be used to supplement and direct the mathematical approach.

Performance endpoints

A range of cross validation techniques and independent testing between the various samples of the database will be used to assess the accuracy, sensitivity, specificity and performance of the models generated.

Research Environment and Developmental Opportunities

Dr David Barber (Reader, **Computer Science**) has over 70 published works in the field of the application and theories of Bayesian methods to large-scale time series data. He has recently explored switch-reset models (models where the underlying processes can be both discrete and continuous), concave Gaussian variational approximations for inference in large-scale Bayesian linear models as well as investigating variational methods for reinforcement learning. His involvement provides a medically unbiased and mathematically rigorous approach to inference of complex datasets.

Additionally, there exists a framework that supports such interdisciplinary research, as evidenced by the existence of the **Computational Life and Medical Sciences Network** and the **Systems Medicine**

Consortia for Research Computing, of which I hold membership. This infrastructure comprises high performance computing systems (e.g. LEGION and GPU clusters), secure data transmission and storage, some of the best machine learning courses in the world and a culture among experts of interdisciplinary research. All these make UCL, a centre of international excellence in biomedical fields and computer science, the perfect location to base this research.

Adverse Event Reporting

Not applicable as only anonymised and retrospective data is to be acquired and analysed.

Peer Review

The project and protocol have already undergone external peer review in an internationally competitive forum (the Medical Research Council).

Ethics

NHS Ethics has been granted: REC Ref: 13/WS/0243

Monitoring and Audit

The principal investigator will be responsible for the day-to-day monitoring and management of the study. The Joint UCLH/UCL/RFH Biomedical Research (R&D) Unit, on behalf of the Sponsor, UCL, will monitor and conduct random audits on a selection studies in its clinical research portfolio. Monitoring and auditing will be conducted in accordance with the Department of Health Research Governance Framework for Health & Social Care, and in accordance with the Sponsor's monitoring and audit policies and procedures.

Financing

The study is funded as part of a Medical Research Council Clinical research Training Fellowship, that has been awarded to Dr Vishal Nangalia. Medical Research Council Reference: MR/K024051/1.

Sponsorship

University College London will be the sponsor

Publication Policy

Multiple manuscripts to leading peer reviewed General Medical and Computer Science journals will be submitted.

References

1. National early warning score (NEWS) standardising the assessment acute illness severity in the NHS. London, UK: Royal College of Physicians; 2012.
2. NICE clinical guideline 50: Acutely Ill Patients in Hospital: Recognition of and response to acute illness in adults in hospital. London: National Institute for Health and Clinical Excellence 2007.
3. Emergency admissions: A journey in the right direction? NCEPOD; 2007.
4. Recognising and responding appropriately to early signs of deterioration in hospitalised patients. National Patient Safety Agency; 2007.
5. Knowing the risk: A review of the perioperative care of surgical patients. NCEPOD; 2011.
6. Chen J, et al. The relationship between early emergency team calls and serious adverse events. Crit Care Med 2009, Jan;37(1):148-53.
7. Prytherch DR, et al. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation 2010, Aug;81(8):932-7.
8. Prytherch DR, et al. The use of routine laboratory data to predict in-hospital death in medical admissions. Resuscitation 2005, Aug;66(2):203-7.
9. Froom P. et al. Prediction of hospital mortality rates by admission laboratory tests. Clin Chem 2006, Feb;52(2):325-8.
10. Asadollahi K, et al. Prediction of hospital mortality from admission laboratory data and patient age: A simple model. Emerg Med Australas 2011, Jun;23(3):354-63.
11. Loekito E, et al. Common laboratory tests predict imminent death in ward patients. Resuscitation - 02 August 2012. DOI: 10.1016/j.resuscitation.2012.07.025.
12. Mohammed AM, Rudge G, Wood G, Smith G, Nangalia V, et al. Which is more useful in predicting hospital mortality - dichotomised blood test results or actual test values? A retrospective study in two hospitals. PLoS ONE ;PONE-D-12-20159R1.