Risk in major surgery

Suneetha Ramani Moonesinghe
BSc. (Hons) MBBS MRCP FRCA FFICM

University College London

Doctor of Medicine (Research)

2013

I, Suneetha Ramani Moonesinghe, confirm that the work presented in this thesis is my own. Where work has been derived from other sources, I confirm that this has been indicated in the thesis (below)

Chapters One and Two

Nil

Chapters Three, Four and Five

Original patient data collected and entered into two databases by Sister Claire Matejowsky and Sister Marjory Mutch, while working as research nurses for the Surgical Outcomes Research Centre at the Middlesex Hospital

Chapter Six

Nil

Abstract

A qualitative systematic review was conducted to assess the performance of tools which have been validated for the prediction of morbidity and/or mortality, in heterogeneous cohorts of surgical (non-cardiac, non-neurological) patients. The Portsmouth-Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (P-POSSUM) and the Surgical Risk Scale (SRS) were found to be the most widely validated and accurate risk stratification tools.

The POSSUM, P-POSSUM and SRS were then validated in a population of patients who had major non-cardiac surgery in a metropolitan UK hospital. Their accuracy (discrimination) was compared against two novel predictors - the additive POSSUM score and the POSSUM physiology score. P-POSSUM and the additive POSSUM score predicted short-term mortality with high-moderate accuracy. The POSSUM Physiology score was moderately accurate and therefore worthy of further evaluation. Both POSSUM and P-POSSUM were poorly calibrated for this population.

The relationships between perioperative risk, postoperative morbidity (measured using the Post Operative Morbidity Survey, POMS), postoperative length of hospital stay, and short-term mortality, were explored in a series of univariate analyses. There were differences in short-term mortality, and the patterns and prevalence of POMS-defined morbidity between surgical specialities.

Cox Proportional Hazards Modelling, using time-dependent covariates, was undertaken to explore the independent relationship between perioperative risk, postoperative morbidity and long-term survival. POMS-defined neurological morbidity (prevalence 2.9%) was independently associated with reduced long-term survival. Prolonged postoperative morbidity, defined as the presence of POMS-defined morbidity on Day 15

post-surgery (prevalence 15.6%), conferred a relative hazard for death in the first 12 months post surgery of 3.52 ($p<0.001$; 95% C.I. 2.23-5.43) and for the next two years of 2.33 ($p<0.001$; 95% C.I. 1.56-3.50). Postoperative morbidity is a significant public health issue and poses a risk to longer-term survival; it would be an important measure of the quality of perioperative healthcare.

Acknowledgements

My sincere thanks to the following people and organisations:

My primary supervisor, Professor Mike Grocott, for support, friendship & his continued mentorship

My secondary supervisor, Professor Kathy Rowan, for encouragement & intellectual contribution

The Francis and Augustus Newman foundation for the grant which supported this work

My friends & colleagues in the Departments of Anaesthesia & Critical Care at University College Hospital, in particular Simon Clarke, Jim Down, Viki Mitchell, & Therese Parker, for giving me a job, letting me leave it, & allowing me to come back again

Professor Monty Mythen, for leading me down this road in the first place

Dr Steve Harris, for being a calming and educating voice on the other end of the phone when I couldn't do my sums

Claire Matejowsky & Maj Mutch, the original SOuRCe research nurses, for collecting the data which I would use so many years later

The patients of the Middlesex hospital, for making Chapters 3, 4 & 5 possible

My mother & sister, for lifelong support & encouragement; my late father, Sesil Dayananda Moonesinghe, who I hope would have been proud to see this book.

Most of all, my husband Ed - brilliant, kind & understanding - for just about everything

Contents

Chapter Two: Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review

Chapter Three: Validation of six perioperative risk stratification tools in a UK surgical population

Chapter Six: Conclusions and future work

Table of tables

Table of Figures

Abbreviations

ACC	American College of Cardiology
ACS	American College of Surgeons
AHA	American Heart Association
ANOVA	One-Way Analysis of Variance
APACHE II	Acute Physiology and Chronic Health Evaluation II
ASA-PSS	American Society of Anesthesiologists' Physical Status Score
AT	Anaerobic Threshold
AUROC	Area under the receiver operator characteristic curve
BHOM	Biochemistry Haematology Outcome Model
BNP	Brain Natriuretic Peptide
BUPA	British United Provident Association
CACI	Charlson Age-Comorbidity Index
CAM	Confusion Assessment Method
CI	Confidence Intervals
CPET	Cardio-Pulmonary Exercise Test
CQUIN	Commissioning for Quality and Innovation
CRP	C-Reactive Protein
DASI	Duke Activity Status Index
DH	Department of Health
DO_2	Oxygen delivery
ECG	Electrocardiograph
E-PASS	Estimation of Physiologic Ability and Surgical Stress
EQ5D	Euro-QOL
FMD	Final Morbidity Day
GA	General Anaesthesia

GI	Gastro-Intestinal
HbA1C	Glycosylated Haemoglobin
HES	Hospital Episode Statistics
HR	Hazard Ratio
HRQOL	Health Related Quality Of Life
hsCRP	High Sensitivity C-Reactive Protein
ICISS	International Classification of Disease Illness Severity Score
ICU	Intensive Care Unit
IQR	Inter-Quartile Range
IRIS	Identification of Risk In Surgical Patients
ISWT	Incremental Shuttle Walk Test
kg	Kilogram
MET	Metabolic Equivalent
MI	Myocardial Infarction
min	Minute
ml	Millilitres
MPM	Mortality Prediction Model
MRIS	Medical Research Information Service
NCEPOD	National Confidential Enquiry into Patient Outcome and Death
NHS	National Health Service
NSQIP	National Surgical Quality Improvement Program
NT-pro BNP	N-terminal pro-Brain Natriuretic Peptide
O:E	Observed: Expected
O_2	Oxygen
PACU	Post Anaesthetic Care Unit
PH	Proportional Hazards
POCD	Post Operative Cognitive Dysfunction

POMS	Post Operative Morbidity Survey
POSSUM	Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity
P-POSSUM	Portsmouth Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity
PROMs	Patient Reported Outcome Measures
QOF	Quality Outcomes Framework
RCRI	Revised Cardiac Risk Index
ROC	Receiver Operator Characteristic (curve)
s.d.	Standard deviation
SAPS	Simplified Acute Physiology Score
SF-36	Short Form 36
SOuRCe	Surgical Outcomes Research Centre
SRS	Surgical Risk Scale
UCH	University College Hospital
UCL	University College London
UK	United Kingdom
US	United States of America
VO_2	Oxygen consumption

“Systems awareness and systems design are important for health professionals, but are not enough. They are enabling mechanisms only. It is the ethical dimension of individuals that is essential to a system’s success. Ultimately, the secret of quality is love.”

Avedis Donabedian

Chapter One: Introduction

1.1 Overview

This thesis focuses on two distinct but closely related areas. The first is the prediction of perioperative risk. The second is the epidemiology of short and long-term outcome after surgery. In this introduction, these areas will be discussed in detail.

Initially, an overview of the significance of perioperative morbidity to patients and the healthcare system is provided. Then, the factors that influence outcome in the perioperative period will be discussed. These include both patient (pathophysiological) and organisational (structure and process) risk factors, and methods for predicting and measuring them. Finally, the various definitions of outcome will be reviewed and appraised, before setting out the aims and objectives of this thesis.

1.2 The scale of the problem

An estimated 234 million major surgical cases occur worldwide each year.[1] The impact of this burden on individual persons, healthcare providers, and society as a whole is difficult to estimate; however, there is convincing evidence that surgical morbidity and mortality is a major public health issue in both developed and developing world environments. Recent data suggest that in Europe, hospital mortality after inpatient surgery may be is as high as 4%, with wide variation in crude mortality rates between countries; the UK estimate of in-hospital mortality was 3.6%, which was considerably higher than had been previously thought.[2] Complications may occur in between 3 and 17% of patients, depending on the type of patients and surgery, and how complications are defined.[3;4] In a study of over 6000 UK civil servants, absence from work of greater

than 7 consecutive days due to surgery, was associated with the second highest hazard ratio for long-term mortality (after circulatory diseases) of any cause of sickness leave (Hazard Ratio [HR] 2.16; 95% C.I. 1.42-3.26, adjusted for age, gender and employment grade).[5] A Finnish study has found that the standardised mortality ratio for patients undergoing most types of surgery is higher than age and sex matched controls drawn from a non-surgical population; this increase in mortality is greater in patients who are high resource consumers while in hospital (and therefore likely to have had a complicated post-operative course).[6]

Data from a large epidemiological study looking at over 4 million surgical procedures in the United Kingdom, identified a 'high risk surgical population': the relatively small proportion of patients that fell into this group (12.3%) accounted for the majority of post-operative mortality (83.4%) and these patients had a significantly increased duration of hospital stay.[7] Perioperative risk was related to increasing patient age and the complexity and immediacy of the procedure: 88.5% of patients in the high risk category underwent emergency surgery, as opposed to just 21.3% in the 'standard risk' population. Of particular interest, less than 15% of the patients who fell into the high risk category were admitted to critical care directly from the operating theatre. The authors noted that cardiothoracic patients (excluded from this study), who by definition have significant co-morbidity and are undergoing complex surgery, have a relatively low population mortality[8] when compared to patients undergoing certain orthopaedic, general and vascular surgical procedures.[9-12] This may be partly due to process-related issues, as the cardiac surgical population is relatively homogenous, allowing for more protocolised and streamlined care. However, the authors also postulated that the more

intensive management afforded to cardiac surgical patients, as a result of routine admission to critical care post-operatively, may lead to better outcomes.

These data may also reflect inaccurate pre-operative risk stratification and insufficient available resources for this high-risk population. They highlight the need for the accurate identification of patients at the highest risk of surgical complications, and who may therefore benefit from higher levels of support, such as elective postoperative critical care admission. Equally, identifying patients for whom this level of support is not required is just as important, so that limited resources are directed at those who are most likely to benefit. In order to explore these possibilities, we need to identify risk factors for adverse outcomes, related to both variations in individual patient physiology and healthcare service provision.

1.3 Pathogenesis of surgical morbidity

The biological mechanism responsible for the development of postoperative morbidity involves an inflammatory response which may clinically manifest as a Systemic Inflammatory Response Syndrome (SIRS); this may lead onto the Multiple Organ Dysfunction Syndrome (MODS) or may be limited milder clinical variants.[13] As long ago as the 1970s, it was recognised that immune modulation occurred as a result of the surgical insult.[14] Although Slade and team originally reported that the immune changes they observed were not linked to clinically significant adverse outcomes,[14] since then perioperative inflammation has been linked to both generic morbidity and specific complications; furthermore, markers which reflect the severity of this response may be measured and have been found to correlate with post-operative outcome.[15-21] For example, Endotoxin, which is a component of gram-negative bacteria and is normally

found in high concentrations in the gastrointestinal tract, is thought to be a trigger for SIRS in the surgical setting.[13] Endotoxaemia may occur as a result of gut mucosal impairment during and after major non-gastrointestinal surgery or from direct manipulation of the gut during abdominal surgery.[22] Work looking at the level of anti-endotoxin core antibody (EndoCAb) in preoperative patients has found that patients who have lower preoperative serum EndoCAb concentrations are more likely to have postoperative complications[23-25] and reduced long term survival. [26] There are also links between surgical inflammatory responses and specific complications. Postoperative cognitive dysfunction is common and may lead to long-term harm; [27;28] in non-cardiac, non-neurological surgery it has been associated with increased levels of pro-inflammatory cytokines (interleukin-6 and prostaglandin E2) in the cerebrospinal fluid. [29] In animal models (elderly mice), the post-surgical inflammatory response has been shown to lead to cognitive decline associated with pathological changes similar to those of Alzheimer's disease (for example beta-amyloidosis and gliosis).[30]

Whether or not a patient enters into a clinically significant physiological decompensation after a surgical procedure is dependent on a number of factors. The physiological characteristics of the patient themselves will determine the likelihood of an adverse response to surgery. The ability of the patient to increase their oxygen delivery to meet the increased demands of the perioperative period is thought to be a fundamental determinant of outcome[31;32] and it is this principle that underpins the strategy of perioperative haemodynamic optimisation. The causes of the uncoupling of oxygen supply and demand in the perioperative period relate both to the patient's co-morbidities and the severity of the surgical insult. The combination of a failing heart and microcirculatory dsyfunction[33] which pre-date the surgical insult, but may then be

exacerbated by it, may lead to poor oxygen delivery and utilisation and the development of postoperative complications.[33] Goal directed fluid optimisation aimed at maximising oxygen delivery, using non-invasive monitors such as the oesophageal Doppler, has repeatedly been shown to be associated with a reduction in perioperative complications and hospital length of stay in certain types of surgical procedure.[34] [35]

1.4 Predicting patient risk for adverse perioperative outcomes

1.4.1 Overview: Why predict perioperative patient risk?

Accurate stratification of patients into 'risk categories' according to their own physiological characteristics may lead perioperative physicians to modify or abandon a planned surgical intervention if the risk is seen to outweigh the benefit, particularly in diseases where a more conservative alternative to major surgery is available. Alternatively, accurate preoperative risk stratification may allow clinicians to select those patients who may benefit from specific treatment strategies, such perioperative haemodynamic optimisation or an enhanced level of perioperative care such as in a critical care unit [36] or Post Anaesthetic Care Unit (PACU).[37] Risk (or case-mix) adjustment is a related technique which is used in healthcare to account for patient risk factors when measuring quality and / or outcomes. The Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM) was first developed for the purposes of risk adjustment, to enable comparative audit in the National Health Service (NHS).[38]

Hundreds of studies have shown both univariate and adjusted relationships between various risk factors and surgical outcome in different settings: these include advancing

age,[39] diabetes mellitus,[40] renal impairment,[41] and cardiac failure[41 42] to name but a few. In recent years, a number of different methods of predicting perioperative risk have been developed which attempt to take into consideration the contributions of all the different risk factors which patients might have; these include risk stratification tools, measures or estimates of functional capacity, and biomarkers. Each of these will now be discussed in turn.

1.4.2 Risk stratification tools

Three of the most commonly used risk stratification tools, the American Society of Anaesthesiologists' Physical Status Score (ASA-PSS), Charlson Age-comorbidity index (CACI), and the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM) are summarised in Table 1. There are also a number of tools in use which have been developed purely for the prediction of cardiovascular morbidity and mortality; the most widely used and recommended tool[43] of these is the Lee Revised Cardiac Risk Index (RCRI).[44;45] These tools, and the many others in existence, differ in their discriminant ability (accuracy) and reliability, and the types of patient, procedure and outcomes for which they are validated. A number of derivatives of these three systems have been validated in different surgical cohorts: the Surgical Risk Scale,[46] and Donati's Surgical Risk Score,[47] are both based on the ASA-PSS, but also include details of the proposed surgical procedure; both systems demonstrate improved predictive accuracy when compared with the ASA-PSS used alone. However, caution should be exerted when considering the use of any of these risk stratification tools for clinical decision making on whether to proceed with a proposed intervention, as they demonstrate variable predictive precision, particularly at the extremes of age[48;49] and calculated risk.[50-52]

Despite these caveats, there are a number of potential benefits to using risk stratification tools. They are usually openly available, may be used on patients undergoing both elective and emergency surgery and in many cases, they are relatively simple to apply. Risk adjustment models developed for the purposes of comparative audit may also be used for patient risk stratification, as has been demonstrated with several models used in the critical care setting.[53;54] While there are a large number of published studies reporting the validity of different surgical risk prediction rules, clinical experience tells us that they are not widely used, at least in the United Kingdom; there are several reasons why this may be so. First, many prediction rules have been developed for the prediction of specific morbid events after surgery, such as cardiac morbidity or renal dysfunction. While these complications are important, they are also relatively uncommon;[55;56] furthermore, generic complications such as infectious and gastrointestinal morbidity and wound complications may carry significant morbidity in the short and long-term.[55-57] It is therefore more useful for perioperative physicians to be able to predict generic morbidity and all-cause mortality, so that a patient's care may be modified if they are identified as high risk. Second, many models predicting generic adverse outcomes have been validated in single surgical speciality cohorts: for example, after vascular or colorectal surgery.[58;59] Most anaesthetists and perioperative physicians manage patients undergoing a wide variety of surgical procedures, and there would be clear advantages to a model which can be applied across surgical specialities. Finally, the reported accuracy of different risk stratification tools has varied in published studies, leading to uncertainty over their value. As a consequence, a number of modifications have been made to previously validated scores, such as the POSSUM system.

Other methods, such as exercise testing or biomarker assays which are discussed below, may be able to provide a more accurate assessment of a patient's perioperative risk than currently available risk stratification tools. However, in certain patients, especially those undergoing emergency surgery, and who are therefore at particularly high risk of adverse outcomes, complex preoperative investigation is unlikely to be feasible. It may be unlikely that a 'one size fits all' risk prediction model would be able to accurately stratify patients worldwide, and over time, as a result of variation in structure and process between healthcare systems, differences in the type of surgery that may be offered to patients for a particular disease in different regions, and changes in how patients are looked after, as medical knowledge and management improves. This issue would be addressed by the development of large databases of patient demographics and risk factors, such as that of the United States' Veterans' Administration National Surgical Quality Improvement Program.[60;61] Such systems would enable the development of 'bespoke' risk stratification models which may be modified annually to reflect regional fluctuations in patient health and perioperative practice.

In the meantime, if we are to use risk models to guide patient management, or for the purposes of risk adjustment for comparative audit, a clear understanding of their strengths and limitations is necessary. In order to progress this understanding, and to summarize the evidence supporting the predictive precision of these tools, a systematic review of clinical prediction models used in major surgery is presented in Chapter 2.

Table 1 Comparison of commonly used risk-stratification scoring systems

Scoring system	Description	Background	Validation	Advantages	Disadvantages
American Society of Anaesthesiologists' Physical Status Score (ASA-PS)	Categorical numbered scale (I-V) based on the severity of patient comorbidities	First reported in 1948 as a description of 'anaesthetic risk' for the purposes of epidemiological study[62]	Multi-centre validation for categorisation of population risk;[63] not validated for prediction of individual patient risk[64]	Simple, easily applied bedside tool not requiring complex calculation.	Subjective; does not differentiate between different types of patient comorbidity; subject to wide inter-observer variability;[65] does not use information regarding the surgical procedure itself. Poor sensitivity and specificity for prediction of morbidity and mortality on an individual patient basis[64]
Charlson Age-Comorbidity Index (CACI)	Additive score based on the presence and severity of different patient comorbidities	First reported and validated in 1994 for the purpose of risk classification in longitudinal studies	Multi-centre; international. Originally validated for long-term population mortality prediction (3-5 years post-operatively) in diabetic and hypertensive surgical patients; has since been validated for prediction of in-patient mortality and morbidity in several cohorts	Relatively simple; bedside calculation possible; more accurate predictor of outcome than ASA-PS;[66] accurate predictor of population risk according to assigned score[67;68]	Uses no information regarding the surgical procedure itself; assessment of patient co-morbidity may be subjective
Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM)	12 physiological and 6 operative variables; each variable is assigned a score and the total scores entered into logistic regression equations which calculate a percentage mortality & morbidity risk	First reported in 1991 to facilitate comparative audit between surgical services (by adjusting surgical outcome data for calculated predictive peri-operative risk)	Multi-centre, international. Portsmouth POSSUM (P-POSSUM) system[69] uses the same dataset as the original POSSUM score, but using linear (rather than logistic) regression to calculate predicted mortality. Found to be an equivalent or superior predictor of mortality than the original POSSUM in different surgical settings;[70;71] Surgery-specific variants such as Cr-POSSUM (colorectal), have been validated as better predictors of outcome than both 'general' POSSUM systems.[72]	Validated for prediction of individual patient risk; POSSUM and its variants have been widely validated in different patient cohorts internationally; uses objective variables (such as blood test results) in dataset therefore reducing inter-observer variability	Uses some subjective variables (such as chest radiograph interpretation and jugular venous pressure measurement) which may be subject to inter-observer variability. The operative variables are not available until during or after surgery. Logistic regression calculation means that the lowest possible mortality risk is 1.08%; predictive ability limited at extremes of age and calculated risk. Linear regression used in the P-POSSUM modification may be more precise

1.4.3 Assessment of functional capacity

The estimation of cardiorespiratory fitness by simple direct questioning has long been a part of preoperative assessment by the anaesthetist, and the patient's self-estimation of exercise tolerance has been shown to correlate with post-operative outcome.[73] However, more sophisticated methods of assessing functional capacity are now being evaluated and more widely implemented: these include subjective but structured estimation (e.g. the Duke Activity Status Index [DASI]), or objective measurement (incremental shuttle walk test [ISWT] and cardiopulmonary exercise testing [CPET]).

The Duke Activity Status Index (DASI) is a simple questionnaire which categorises levels of exertion (for example, ability to climb stairs or complete household chores) according to the metabolic equivalent (MET) of oxygen consumption required to achieve the task. The DASI is correlated with peak oxygen uptake on exercise testing [74] and, as suggested in the American College of Cardiology / American Heart Association (ACC/AHA) guidelines on cardiovascular preoperative assessment, may be a useful element of a wider pre-operative evaluation, especially in the emergency situation where there is no opportunity for formal exercise testing.[43] However, it is a patient reported and therefore subjective measure, which may therefore not always correlate with an individual patient's true functional capacity.

The incremental shuttle walk test (ISWT) involves a patient walking back and forth between two fixed points to the limit of their exertion. It is a validated and highly reproducible measure of functional capacity, and is likely to be most valuable for the screening out of patients with sufficient level of fitness that they do not require further investigation.[75;76] Cardiopulmonary Exercise Testing (CPET) requires the patient to cycle on a bicycle ergometer at gradually increasing intensity during a 'ramp' exercise

protocol. The test results are summarized by a 'nine panel plot' which shows a number of variables including the peak oxygen consumption (VO2peak), the Oxygen Pulse (the amount of O_2 consumed from the volume of blood delivered to tissues by each heartbeat; O_2 pulse=VO_2/heart rate) and the anaerobic threshold (AT) (the point at which anaerobic metabolism starts to increase significantly, as oxygen delivery to muscles is surpassed by the metabolic demands placed by exercise).[77] Studies conducted in the 1990s identified a correlation between an objective measure of poor functional capacity (anaerobic threshold <11ml/kg/minute) and peri-operative mortality in different surgical populations.[78;79] Of note, exercise capacity was found to be a more significant correlate of outcome than the occurrence of exercise induced ischaemia; the prognosis was worst in patients who had both ischaemic exercise tests and a low anaerobic threshold. More recent work has also found VO_2peak to predict outcome in a variety of other patient cohorts.[80-82]

Recent work suggests that the DASI and the ISWT are highly sensitive and specific in the identification of patients at low risk of perioperative complications; however they may incorrectly categorise as high risk a significant number of patients who, on cardiopulmonary exercise testing, would be considered low risk (that is, with an anaerobic threshold of greater than 11mlO_2/kg/min or peak VO_2 >15mlO_2/kg/min).[83] Such data suggest that following an algorithmic approach to the assessment of functional capacity might be useful: the DASI ± ISWT may be used as inexpensive and quick screening tools to identify low risk patients who warrant no further investigation; patients identified as 'high risk' by these methods, may benefit from further evaluation using CPET. For those patients identified as 'high-risk', management on critical care post-operatively, and the use of haemodynamic optimisation perioperatively, may be of

benefit, although at present there are no published multi-centre studies examining the impact of specific management strategies on patients who have been risk- stratified using CPET. A study evaluating this issue is near completion (Cardiopulmonary exercise testing and pre-operative risk stratification: UKCRN ID 4187): patients are randomised to either having their place of postoperative care (Critical Care versus general ward care) determined by the results of the CPET (intervention limb) or by the perioperative team based on their best judgement (standard of care).

1.4.4. Biomarkers

A 'biomarker' has been defined as 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'.[84] Thus, there is a huge spectrum of investigations which fall within this definition, and which may be used in the perioperative setting. Abnormal results from some routine preoperative tests (for example, blood pressure, ECG and measures of renal function such as serum creatinine) have been linked to poor outcome in general populations, and have been incorporated into surgical risk stratification tools such as the Lee Revised Cardiac Risk Index and the POSSUM predictors.[38;44]

However, work continues to identify markers which may more specifically or accurately predict or quantify adverse surgical outcomes. For example, high sensitivity C-Reactive Protein (hsCRP) is well established as a marker of inflammation that may predict vascular and cardiac adverse events in the general population, independently of accepted risk factors such as smoking, hypertension and diabetes mellitus.[85] It has been found to inversely correlate with functional capacity as measured by VO2peak even in symptom free individuals, [86] and the addition of hsCRP level to existing risk models for

the prediction of cardiovascular risk in the general population improves accuracy.[87] Recent work has examined the relationship between preoperative hsCRP level and surgical outcome, and found an independent association between elevated hsCRP levels and adverse events in orthopaedic[88] and vascular surgical patients.[89-91] Other studies show preoperative hsCRP to be independently associated with long term outcome in a variety of cohorts with surgically treated malignancies,[92-94] and patients undergoing cardiac surgery.[95;96] However, heterogeneity in the hsCRP level that was defined as 'high' in these studies, means that, based on existing data, the generalizability of this assay for patient risk stratification remains limited.

Biomarker assays being evaluated for the prediction of adverse cardiac events include Brain Natriuretic Peptide. The pre-operative level of N-terminal pro-Brain Natriuretic Peptide (NT-pro BNP) has been found to independently predict mortality and cardiac events after both major cardiac[97-99] and major non-cardiac surgery.[100-103] There are, however, several limitations of NT-pro BNP as a biomarker which should be considered. In the general population, there are age, gender and assay-specific variations in cut-off values for risk stratification.[104] Renal impairment has been shown to impair the usefulness of the assay for the prediction of perioperative cardiac complications.[105] Nevertheless, in the future, NT-pro BNP (and similar biomarkers) may prove to be a useful part of a multi-factorial risk assessment: as with generic perioperative risk, the evaluation of cardiac risk is likely to involve a combination of the above strategies to achieve optimal predictive accuracy.

Looking ahead, rapid advances in genomic sequencing mean that personalised diagnostics may be become a reality sooner rather than later,[106;107] although cost is likely to remain an issue for some time. It is possible that in the future, the addition of

biomarker assays and measures of functional capacity to risk stratification scoring systems will lead to improvements in accuracy of perioperative risk prediction.

Patient risk is a key determinant of perioperative outcome. However, it is known that clinical management may either mitigate or exacerbate pre-existing risk, and it is also known that there is wide variation in surgical outcomes (risk adjusted) between institutions. This variation may be explained by difference in the care that patients receive once they enter the surgical pathway; this is most succinctly explained using the 'structure process outcome' model, which will be discussed now.

1.5 How quality of healthcare affects outcome: the structure / process / outcome model

1.5.1 Overview

Over 40 years ago, Avedis Donabedian, the Lebanese-born Professor of Public Health at the University of Michigan, described the structure / process / outcome (SPO) model of evaluating quality in healthcare.[108] 'Structure' refers to how healthcare is organised, 'process' is the method by which healthcare is provided, and 'outcome' is the state resulting from healthcare processes. There is a growing body of literature using this model to describe quality, and demonstrating the impact that structure and process have on population outcomes for patients.

1.5.2 Structure

The description of structure incorporates a wide variety of variables specifically related to the environment within which healthcare is provided. Physical characteristics of hospitals, such as the number of beds,[109] bed occupancy rate[110] and the presence of

particular services associated with 'high technology' (such as cardiac or organ transplantation surgery) [111] have all been shown to be associated with improved surgical outcomes. Staff characteristics, such as nurse-patient ratios,[112] compliance with training requirements, the proportion of nurses and doctors with higher qualifications or specialist experience,[110] and payroll expenses[110] have also been associated with hospital level outcomes. Then there are structural elements which might be considered more esoteric, such as those pertaining to the attitudes and behaviours of staff members; these include adherence to surgical checklists,[113] engagement with audits monitoring quality of care,[114] and critical incident reporting rates.[115]

There is a plethora of evidence demonstrating the association between healthcare structures and population outcomes. A retrospective analysis conducted by the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) examined the outcomes of over 180,000 surgical cases and found a significant increase in 30 day mortality for the population of patients who had their elective inpatient surgery performed on a Friday as opposed to those who were operated on between Monday and Wednesday. This difference was only evident in patients who were admitted post-operatively to general wards, rather than day-case patients or those admitted to critical care.[116] These data imply that reduced staffing levels on general wards at the weekends are having an adverse effect on patient outcomes, a phenomenon which has previously been reported in patients admitted via emergency departments with a wide variety of medical conditions.[117;118] By contrast, the maintenance of normal staffing levels at night and weekends on critical care wards may explain why patients admitted to ICU post-operatively have similar outcomes irrespective of the day of their surgery. This theory is supported by previous data

showing no association between the time of admission and the case mix-adjusted outcomes of unselected patients admitted to critical care,[119] but a significantly higher mortality in patients discharged from critical care to normal wards at night.[120] In the UK, the disruption to NHS organisational structure at the beginning of August, as a result of the changeover of junior hospital doctors (and possibly also the absence of senior medical staff due to the holiday season) has been shown to be associated with excess patient mortality, particularly for medical admissions (OR 1.08 for patients admitted on the first Wednesday in August, 95% CI 1.01 to 1.16, p = 0.03).[121]

Further work conducted by the NSQIP revealed a significant difference in surgical mortality between institutions, despite a similar case-mix adjusted morbidity rate.[122] This 'failure to rescue', in hospitals with higher mortalities, is likely to be associated with structure related issues: for example, medical and nursing staffing levels may influence the timely recognition and management of patients with post-operative complications. Indeed, in a follow-up study focussing only on patients who had undergone pancreatectomy, significant differences between high and low performing institutions were identified in nurse-patient ratios, bed occupancy rates, technology provision and hospital size.[109]

In many respects, structural metrics are an attractive method of assessing quality of healthcare, as the reporting and recording of these measures may be easier than process and outcome measures. In the UK, the Department of Health (DH) has adopted a strategy of incentivising healthcare institutions by rewarding then financially for adherence to the 'Commissioning for Quality and Innovation' or 'CQUIN' framework which was introduced in 2008.[123] There are a number of CQUIN indicators which are measures of structure, rather than process or outcome: for example, the indicator

related to preventing harm from venous thromboembolism (VTE). The National Institute for Clinical Excellence has issued guidelines on best practice related to preventing VTE. Consequent to this, a CQUIN payment was introduced nationally, the award of which is determined not by a measure of outcome (e.g. reduction in VTE rate) or even process (e.g. proportion of patients receiving appropriate VTE prophylaxis) but of structure (>90% of patients having a documented VTE risk assessment on admission to hospital). Other examples of CQUIN payments attached to structural changes include the implementation of enhanced recovery programmes,[124] and the Institute for Health Improvement Global Trigger Tool.[125]

The CQUIN targets are structure-related metrics of quality which are relatively easily within the influence of clinical and managerial leaders within individual institutions. There are however, a number of other structural elements which healthcare providers may not necessarily be able to address. A number of cross-sectional studies reveal a volume - outcome relationship in healthcare: that is, that the greater the number of patients that a particular surgeon, physician or institution manages, the better the outcome for the cohort.[126;127] This has been demonstrated consistently and in a wide variety of surgical specialities and healthcare systems worldwide.[128-131] While this may not seem surprising, longitudinal studies do not confirm a 'practice makes perfect' explanation for these observations, implying that there are intrinsic differences between high and low volume hospitals in terms of standards of care[132;133] - hence their appeal as measures of quality. However, one must be cautious about writing off 'low volume' institutions as poor quality, or considering that 'bigger is always better':[134] there are examples of outlying providers in most studies, the magnitude of the volume-outcome relationship varies widely depending on the procedure being considered, and

indeed the association may not be clinically (or statistically) significant in some conditions.[126] Despite these caveats, the volume – outcome relationship is a major driver for service reconfiguration in the NHS, even though most studies in this field have been conducted in the US;[135] it is true, however, that there is an emerging literature in the UK.[131;136-139]

The main advantage of using metrics related to structure in the evaluation of healthcare quality is simplicity. Many variables, such as procedural volume, staffing levels and bed occupancy are easily measured using administrative data. However, this must be balanced against a number of disadvantages. First, changing structure to improve quality is not always possible: for example, an NHS provider is unlikely to be able to transform itself into being a high-volume institution without the influence of regional service reconfiguration. Second, structural metrics may be subject to 'gaming'; an example of this was demonstrate in the NHS when data revealed a disproportionately high number of patients were being admitted into hospital towards the end of a four hour wait in the Emergency department when this was set as a government target.[140] Third, as already stated, structural variables reflect average figures for large numbers of institutions, not individualised results for specific providers – hence the likelihood that there will be outliers in the provision of healthcare quality which may be unjustly criticised or rewarded if measured against structure-associated standards alone.

1.5.3 Process

'Process' is the care which patients actually receive, and is not limited to describing pharmacological or procedural interventions. In primary care, the Quality Outcomes Framework (QOF) has been used to reward strategies aimed at early disease detection (blood pressure monitoring), disease prevention (provision of smoking cessation advice

or influenza vaccination) or disease control (referral to secondary care for management of ischaemic heart disease). In perioperative care, process measures may include the proportions of surgical patients who attended preoperative assessment clinics, or are reviewed postoperatively by specialist pain teams, or who are normothermic at the end of their surgical procedure.[141]

Process measurement has clear appeal as a driver of quality assessment and improvement. An evidence-based approach is possible, by measuring compliance with interventions which have been shown to be effective in clinical studies. Examples of evidence based process measures which are used across the world include correct prescription of perioperative antibiotic and VTE prophylaxis, or the proportion of patients who undergo Caesarean section under general anaesthesia (GA) – a higher proportion is viewed as unsatisfactory.[141] Importantly, changes in process are often more easily implemented than structural amendments: ensuring that >90% of elective surgical patients are seen in an established preoperative assessment clinic (process measure) may be somewhat easier than finding the physical space, financial and personnel resources required to set such a service up in the first place (structure measure). Process measures may also be viewed as being a more useful or somehow 'fairer' reflection of quality (the preoperative assessment clinic is only worthwhile if the majority of surgical patients who would benefit from it, are actually seen in it, rather than it being inefficient or dormant).

However, there are also disadvantages with the process measurement approach. First, the denominator for measurement needs to be clear and accurate in order for fair assessment to be made. For example, if the GA Caesarean section rate is higher than expected in a particular maternity unit, might this not be because in the specific

circumstances of those individual patients, a GA section was the safest approach? Second, compliance with process measures does not ensure better patient outcome, even if the intervention has been shown to be of benefit in clinical studies. An example of this is the use of the oesophageal Doppler monitoring device for perioperative fluid management. Simply inserting a Doppler probe into a patient in all relevant cases (easily measured process variable), does not ensure that the device is used correctly (or indeed used at all) in order to optimise patients' fluid status appropriately, and therefore reduce postoperative morbidity as the studies suggest it should.[142-144]

1.6 Outcome measures after major surgery

In this section, I shall discuss the various options for measuring surgical outcome. Outcome can be defined as 'what happens to the patient as a result of their treatment'. In surgery, the most commonly measured outcome is mortality.

1.6.1 Mortality

Mortality has several advantages as a quality metric, including that it is a dichotomous variable which is objective, clinically important, and not subject to inter-observer variability. Nevertheless, there are a number of issues that may limit its use as an outcome measure, or at least argue for the concurrent recording of other outcomes. Interpretation of mortality as an outcome is highly dependent on the timeframe of measurement: in-patient, 28 or 30 day or 1 year mortality (or survival) are commonly reported definitions but are not reliably related to each other; therefore, it is scientifically unsatisfactory to compare the results of studies or audits in which different mortality definitions are used. Inpatient mortality has the advantage that the data should be easier to collect, with loss to follow-up being an infrequent problem;

however, confounding may occur as a result of different criteria for patient discharge. In studies which take longer term mortality as their outcome (28 day or longer), loss to follow up post-hospital discharge may be a concern, as may the influence that the natural course of co-morbid diseases has on patient survival.

Another important issue, both for the assessment of quality in surgical care, and for clinical trials interventions, is the incidence of mortality in perioperative patients. As surgical and anaesthetic techniques have improved over time, this is reflected in published audit data and in longitudinal studies using short term patient mortality as an outcome. [145-147] Studies looking at the influence of haemodynamic optimisation on perioperative outcomes frequently fail to show an improvement on perioperative mortality as a result of intervention, although they consistently show an improvement in morbidity and also resource utilisation measures such as hospital or intensive care unit length of stay.[35;148-150] The failure of such studies to demonstrate improvement in hospital survival may, in part, be responsible for the previously poor uptake of optimisation strategies by perioperative physicians. However, accumulating evidence of an association between perioperative morbidity and long-term survival, suggests that mortality data relating only to the acute surgical admission may not fully reflect the healthcare impact of the surgical episode.[151]

Analysis of over 105000 patients in the NSQIP database revealed that the occurrence of any one of 22 peri-operative complications reduced median life expectancy by 69%.[151] This impact of morbidity on long-term mortality was found to be independent of preoperative comorbidities (multivariable analysis with Cox Proportional Hazards modelling) and was still significant after deaths within 30 days were excluded from analysis. The extent to which morbidity affected long term survival was related both to

the seriousness of the adverse event and the type of operative procedure. Nevertheless, even complications which may be considered to be relatively minor (such as urinary tract or wound infection) occurring after procedures that may be considered to be relatively benign (for example, laparoscopic cholecystectomy) were shown to be associated with reduction in long term survival. This study is the largest to link long-term outcome with perioperative events, and is supported by a growing body of evidence in heterogeneous surgical populations.[152-156] While there is currently no mechanistic explanation for these findings, there seem two likely possibilities. The first, is that the methodology for risk adjustment is flawed, and that the patients who developed postoperative complications were 'predestined' to do so on the basis of prior ill-health, which may not be overt at the time of surgery. The second possible explanation is that residual effects on functional capacity or the persistence of an ongoing inflammatory process associated with a postoperative complication is responsible for long term adverse outcomes. Whether such an inflammatory response occurs as a result of the postoperative complication, or whether the complication occurs as a result of an already existing (and unidentified) pro- inflammatory state, is unclear. There are data from other studies which support this theory. For example, the development of an asymptomatic rise in cardiac troponin levels within seven days of vascular surgery is associated with reduced long term survival ($p<0.0001$; median follow up 4.3 years, 25^{th}-75^{th} percentile 2.8-5.3 years).[157] The question of whether this troponin rise results from myocardial ischaemia, or myocardial inflammation, is being investigated currently in the international Vascular Events in Non Cardiac Surgery Patients Cohort Evaluation (VISION) study.[158]

A further observation relates back to the issue of when after surgery mortality should be recorded. Data from the NSQIP study described above suggest that most informative time to measure mortality may differ between surgical procedures.[151] Survival curves for patients undergoing each of the 9 different surgical procedures analysed, and who incurred at least one postoperative complication, all show an 'inflection point' where the mortality rate appears to change and flatten out. One interpretation of the inflection point is that this is the time point when the true impact of the surgical episode ends. Of interest, this point varies hugely between surgical procedures (from 27 days for total hip replacement to 180 days for carotid endarterectomy) and for all procedures except for hip replacement, the inflection point occurs beyond 30 days postoperatively. This is notable because 30 day mortality is one of the most commonly used endpoints for measuring surgical outcome, both in clinical practice and research. Furthermore, these data again highlight the inadequacy of inpatient mortality as an endpoint for quality measurement or research purposes. It should be noted that these inflection points were derived by plotting survival curves only for patients who developed postoperative complications; although this was not explicitly discussed in Khuri's paper.

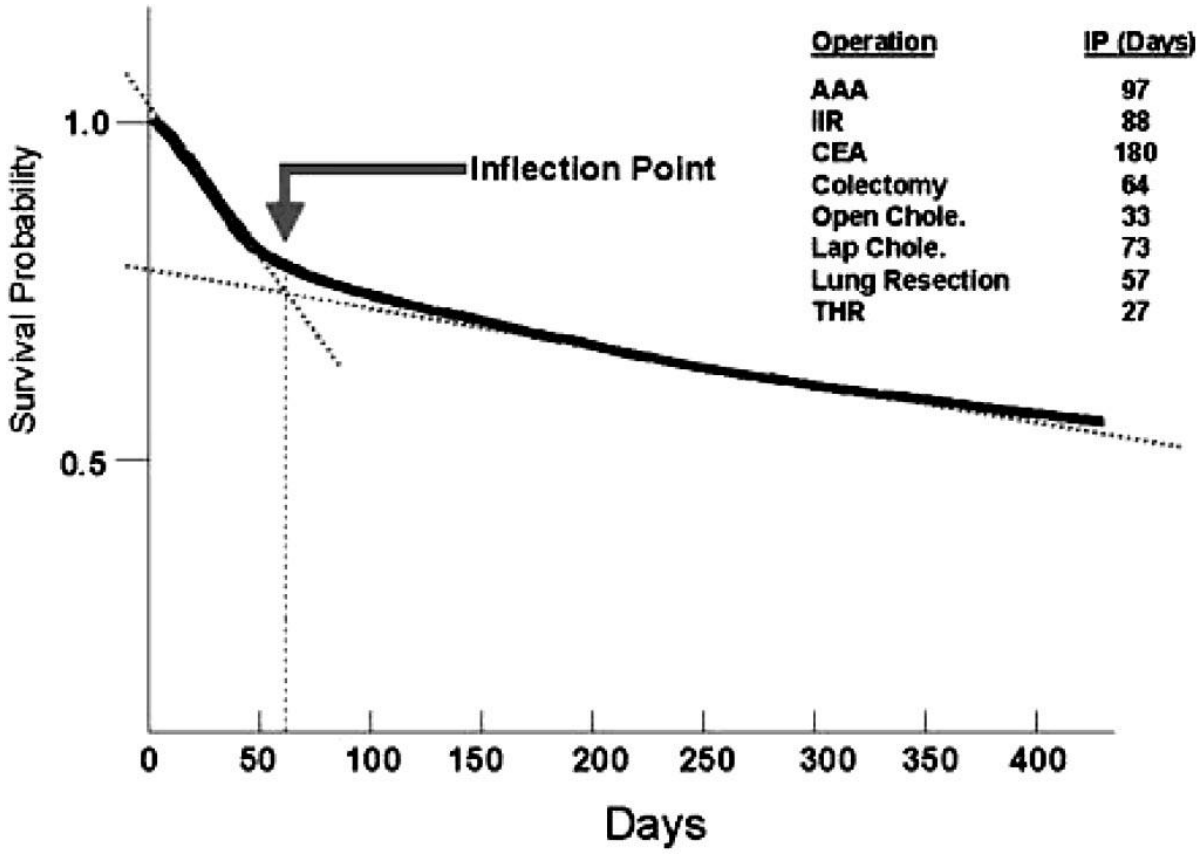

Figure 1: Survival curve of patients who developed complications within 30 days of colectomy, showing the 'inflection point'

Inflection points (days) for 8 other procedures is tabulated at the top right of the graph. *(Taken from Khuri et al: Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. Ann.Surg. 2005; 242: 326-41)*

1.6.2 Morbidity

In general, clinical outcome measures may be classified as disease specific or generic.[159;160] Perioperative morbidity may be considered to be a type of disease specific measure, where the 'disease' is undergoing major surgery, and 'morbidity' is defined as the occurrence of any clinically significant non-fatal complication. However, the literature is inconsistent in its definition and reporting of complications. Traditionally, postoperative morbidity has been classified as either local (relating to the operative site) and general (relating to other organ systems) or specific (relating to the particular operation) and general (relating to any type of operation).[161;162] Morbidity may also be classified as early, intermediate or late, defined by arbitrary time-points. However, the classification of adverse events by any of these methods has a number of difficulties. For example, there may be significant interaction between the occurrence of a local complication (for example, wound infection) and a general outcome (for example, pyrexia). There may also be difficulties in attribution: while post-operative gastrointestinal dysfunction or ileus may be considered as a specific consequence of gastro-intestinal (GI) surgery, it is also a common complication of non-GI surgery with a multi-factorial aetiology,[163] and therefore it may also be considered as a general complication.

1.6.3. Which complications matter? Organ specific morbidity

1.6.3.1 Cardiac morbidity

Considerable attention has been focussed on the incidence of adverse cardiac outcomes following surgery. Cardiac complication rates may be as high as 13.5% in vascular surgery, [164] and 12.9% after major abdominal surgery, [165] although studies vary in their definitions of cardiac outcome. Comprehensive guidelines have been issued both in the

US[43] and in Europe,[166] aimed at guiding clinicians in the detection and management of cardiac risk and morbidity in the perioperative period. The ACC/AHA guidelines recommend a stepwise approach to pre-operative investigation of cardiac disease, based on the presence of clinical risk factors, the patient's functional capacity and the severity of the proposed surgical intervention. It is suggested that clinical risk factors may be assessed using the Lee Revised Cardiac Risk Index (Lee RCRI), which is a widely validated and simple composite scoring system, looking at 5 risk factors for cardiac disease, and the risk of the proposed surgery.[44] Functional capacity may be assessed using the Duke Activity Status Index, as previously discussed. However, once a patient is identified as being at high cardiac risk using these simple screening tools, there remains controversy over the choice of non-invasive test which will most reliably identify significant coronary artery disease. The ACC / AHA guidelines recommend that the choice of investigation be determined by local availability and expertise, as there is no clearly superior test.[43]

Whichever investigation is undertaken, the key decision in this process is determining whether the test results will influence subsequent management: that is, whether the results will help to identify high risk patients who may benefit from specific pre- or perioperative management interventions, such as coronary revascularisation or beta-blockade.[167] Furthermore, even in the case of patients in whom significant coronary artery disease is identified, the evidence in favour of particular treatment strategies is subject to debate. In patients with stable one, two or three vessel disease, for example, recent randomized controlled trials suggest that outcome is not improved by prophylactic revascularisation.[168;169] The optimal management strategy in patients in whom more significant cardiac impairment is identified (for example, left ventricular

dysfunction, unstable angina or aortic stenosis) is not yet clear.[168;169] Similarly, recent data suggest that in many patients, the perioperative introduction of beta-blockade may be more harmful than beneficial.[167] It is a core role of the perioperative team to weigh up the potential risks and benefits to an individual patient of embarking on a cardiac optimisation programme, which may involve either surgery, percutaneous intervention or pharmacological manipulation, and all of which may necessitate at least a 6 month delay to surgery with the inherent risk of disease progression, particularly in cancer patients. Central to this decision making, is the understanding that ischaemic heart disease per se, is less likely to impact on overall postoperative outcome than cardiac failure, which may lead to multi-organ dysfunction.[170;171]

1.6.3.2 Neurological

Postoperative neurological complications have a wide disease spectrum, and include delirium, Post-Operative Cognitive Dysfunction (POCD) and cerebrovascular events (transient ischaemic attack or cerebrovascular accident). Delirium is defined as an acutely altered and fluctuating mental status, including features of inattention and fluctuation in conscious level and may be diagnosed using the 'Confusion Assessment Method' tool (CAM, or CAM-ICU for use in Critical Care settings). Delirium should be distinguished from the more subtle POCD, in which an alteration in thought processes leads to disturbances of cognition which may include visual and verbal memory, attention and language comprehension. Delirium may affect between 0% and 78% of patients depending on the population characteristics,[172] and may be predicted by pre-existing cognitive impairment and advanced age amongst other factors.[172-176] Perioperative management, such as the use of tertiary ammonium compound anti-cholinergic drugs (for example, atropine) may also put patients at risk of delirium.[172]

Delirium is associated with increased length of hospital stay and increased mortality.[176;177] A large international prospective study of POCD after non-cardiac surgery, found the incidence to be over 25% at one week postoperatively (25.8%; 95% C.I. 23.1-28.5) and 9.9% (8.1-12.0) at 3 months.[27] A follow-up study of a sub-group of Danish patients found reduced long-term survival (median follow-up time 8.5 years) in patients who had POCD at 3 months postoperatively (HR 1.63 [95% C.I., 1.11–2.38], adjusted for sex, age, and cancer).[178]

1.6.3.3 Renal

Acute kidney injury occurring after surgery is a leading cause of renal impairment for hospitalised patients.[179] While the aetiology may be multifactorial, the mechanism is usually acute tubular necrosis (ATN) as a result of hypoxic injury to medullary nephrons, secondary to hypotension, hypoxemia or hypovolaemia.[180] The risk factors for acute perioperative kidney injury include pre-existing renal dysfunction, cardiac failure and iatrogenic causes, particularly non-steroidal anti-inflammatory drugs and contrast used for radiological investigation.[181] After cardiac interventions, the occurrence of acute kidney injury is associated with reduced long-term survival.[182]

1.6.3.4 Gastrointestinal

Postoperative gastrointestinal (GI) dysfunction is common and is associated with prolonged length of hospital stay.[56;183] While GI complications are also multifactorial in aetiology, a significant contribution may come from gut ischaemia, which may be sub-clinical and not necessarily associated with surgery where the bowel is handled. Gut perfusion may be critically affected by changes in volaemic status, and if ischaemia occurs, cell death and necrosis may ensue, leading to clinical manifestations such as nausea, vomiting and intolerance of enteral nutrition.[184] Furthermore, gut ischaemia

may herald more widespread morbidity, leading to increased costs, length of stay and increased perioperative mortality. [22;185;186] In addition, neurally mediated dysfunction (activated either via central mechanisms or local release of neurotransmitters), opiate and surgically driven gut stasis, bacterial translocation and endotoxaemia all combine to bring about gut dysfunction which may lead to a postoperative inflammatory syndrome and multiple organ dysfunction or failure.[163;185-187]

1.6.4 The case for using a generic measure of morbidity

1.6.4.1 Overview

While it is important to understand the risk factors and underlying mechanisms of postoperative dysfunction in individual organ systems, a different approach would be to consider post-operative morbidity as a 'syndrome', in a similar manner to how Multiple Organ Dysfunction Syndrome (MODS) or sepsis are described and now defined by specific scoring systems.[188-190] A syndrome is defined as a pathological condition associated with a cluster of co-occurring symptoms, usually three or more.[191] Also central to the definition of a syndrome is the existence of an underlying common pathological process. While each of the individual organ system dysfunctions described above may have particular risk factors (for example, renal impairment from contrast-induced nephropathy), it is also true that the postoperative SIRS response is a common feature in their aetiologies. It may therefore also be argued that a composite measure of postoperative morbidity may be a useful tool in the description of perioperative outcome for the purposes of audit and studies of clinical effectiveness and prognosis.

Furthermore, there is a need to report postoperative complications in a uniform manner, so enabling academic comparison between studies, and clinical comparison between teams and institutions. Systematic [192] and structured [193] reviews have found

lack of consistency of definition of morbidity, and therefore wide variation in reported incidence of complications after surgery. Furthermore, even when complications are defined using a recognised system, the level of training, and indeed engagement, of those collecting the data, have a major influence on the quality of reporting and recording of outcomes.[194]

1.6.4.2 Clavien, Dindo & Strasberg: Severity grading of surgical complications

Clavien and Dindo described their system for reporting surgical complications in 2004,[195] which was a revised version of an earlier grading system (sometimes known as T92) which Clavien had developed using a cohort of patients undergoing cholecystectomy, with another surgeon, Strasberg.[196]

The 2004 Clavien-Dindo system very clearly attempts to quantify the severity of postoperative adverse events but does not attempt between types of complication (for example by organ system) or aetiology (for example infectious); nor does it appear to consider morbidity as an over-arching construct. It is a summary measure which is recorded at the end of the surgical episode for the purposes of later audit, and has also been used as an outcome measure in many published surgical studies. In 2009, the authors published a five year review of their system; this incorporated a literature search for citing papers, and some elements of a validation process, such as inter-rater reliability (they posed several clinical scenarios to multiple assessors in different continents and reported the answers) and face validity (by asking patients, nurses and doctors to record their perception of the severity of complications in these clinical scenarios on a visual analogue scale [VAS]).[197] By that time, 214 papers had cited or used their classification system to grade postoperative complications. Inter-rater reliability was not formally (statistically) tested but was reported in narrative form to

be reasonable. Interestingly, while all three groups were able to distinguish the difference between grades of complications appropriately, patients' quantification (using the VAS) of complication severity was consistently and significantly higher than that of doctors or nurses; in addition, nurses generally perceived complications as more serious than doctors.[197] This observation raises some interesting questions about the communication of risk to patients, and importantly, their perception of outcome, both of which will be discussed later in this introduction.

Table 2: Clavien-Dindo Classification of Postoperative Complications

Grade	Definition
Grade I	Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic or radiological interventions
	Allowed interventions include drugs as anti-emetics, antipyretics, analgesics, diuretics, electrolytes and physiotherapy. This grade also includes wound infections opened at the bedside.
Grade II	Requiring pharmacological treatment with drugs other than those allowed for Grade I complications Blood transfusions and Total Parenteral Nutrition (TPN) also included
Grade III	Requiring surgical, endoscopic or radiological intervention
Grade IIIA	Intervention not under general anaesthesia
Grade IIIB	Intervention under general anaesthesia
Grade IV	Life threatening complications (including CNS complications) requiring critical care management
Grade IVA	Single organ dysfunction (including dialysis)
Grade IVB	Multi-organ dysfunction
Grade V	Death of a patient
Suffix 'd'	If the patient suffers from a complication at the time of discharge, the suffix 'd' (for disability) is added to the respective grade of complication. This label indicates the need for a follow-up to fully evaluate the complication

The Clavien-Dindo classification was reviewed by its developers as a reliable tool for grading the severity of surgical complications.[197] However, a contrary view was taken in a paper published in the same edition of *Annals of Surgery*. [192] Clavien's partner in the development of the original T92 severity scale, Strasberg, led a review of the literature looking for evidence of the accurate utilisation of the Clavien-Dindo classification. One of their key findings was that the number of complications reported in studies using the Clavien-Dindo system varied between three and 720.[198] Their recommendation was to adapt the T92 and Clavien-Dindo classifications into a modified system known as the 'Accordion Grading System' (see Tables 3, 4 and 5). The Accordion system is so named as it can contract and expand to suit the requirements of the study or clinical setting in which it is being used.

Grade	Description
1. Mild	Requires only minor invasive procedures that can be done at the bedside such as insertion of intravenous lines, urinary catheters, and nasogastric tubes, and drainage of wound infections. Physiotherapy and the following drugs are allowed: anti-emetics, antipyretics, analgesics, diuretics, electrolytes, and physiotherapy.
2. Moderate	Requires pharmacologic treatment with drugs other than such allowed for minor complications, for instance antibiotics. Blood transfusions and total parenteral nutrition are also included.
3. Severe	All complications requiring endoscopic or interventional radiologic procedures or re-operation as well as complications resulting in failure of one or more organ systems.
4. Death	Post-operative death

Table 3: Accordion Severity Classification of Postoperative complications – Contracted Classification

Grade	Description
1. Mild	Requires only minor invasive procedures that can be done at the bedside such as insertion of intravenous lines, urinary catheters, and nasogastric tubes, and drainage of wound infections. Physiotherapy and the following drugs are allowed: anti-emetics, antipyretics, analgesics, diuretics, electrolytes, and physiotherapy.
2. Moderate	Requires pharmacologic treatment with drugs other than such allowed for minor complications, for instance antibiotics. Blood transfusions and total parenteral nutrition are also included.
3. Severe	Invasive procedure without General Anaesthesia: Requires management by an endoscopic, interventional procedure or re-operation* without general anaesthesia.
4. Severe	Operation under General Anaesthesia
5. Severe	Organ System Failure
6. Death	Post-operative death

Table 4: Accordion Severity Classification of Postoperative complications – Expanded Classification

It can be seen that the contracted and expanded systems are based on the T92 and Clavien-Dindo classifications; the definitions of organ failure are based on the gradings used in the Sequential Organ Failure Assessment (SOFA) score used in Critical Care.

Organ system	Description
Cardiac	Need for any of the following medications in the following doses Norepinephrine: >0.1 mcg. kg. min^{-1}; Epinephrine: >0.1 mcg. kg. min^{-1} Dopamine: >15 mcg. kg. min^{-1}
Central Nervous System	Glasgow coma scale equal to or less than 6.
Haematology	Platelet count less than 20 _ 109/L
Liver	Need for FFP to correct INR in patient with serum bilirubin >12 mg/dL; OR: INR >2.5 in patient with serum bilirubin >12 mg/dL
Renal	Need for dialysis in patient not on dialysis preoperatively
Respiratory	Need for mechanical ventilation for greater than 24 h in a patient who requires re-intubation after surgery; OR: Need for mechanical ventilation of greater than 72 h in a patient who is not extubated on the day of surgery. Does not include patients already on a mechanical ventilator for respiratory failure

Table 5: Definitions of Organ Failure for Accordion Severity Classification

The 2009 papers were accompanied by an Editorial written jointly by Pierre Clavien and Steven Strasberg,[199] where despite the conflict between their proposed systems, they agreed that surgeons should be involved in determining how to take these systems forward. An invitation was made for surgeons to comment on the two systems, with the aim of the authors reaching a consensus for publication within six to twelve months. To date (April 2013) no consensus has been published. One may deduce that surgeons and perioperative physicians in general have' voted with their feet': while Strasberg's 2009 Accordion Classification paper has been cited 20 times to date, Clavien's five-year

review of the Clavien-Dindo system paper has been cited 88 times in the same timeframe; furthermore, the Clavien-Dindo system has been adopted as an outcome measure by the Safe Surgery Saves Lives study group which developed the World Health Organisation Safer Surgery Checklist,[113] and a number of other highly publicised initiatives which have led to widely accessed papers. However, it remains the case that none of these systems have undergone a formal development and validation process, using clinimetric and / or psychometric strategies.[200]

1.6.4.3 The National Surgical Quality Improvement System Classification

While the T92, Clavien-Dindo and Accordion classifications all deal with grading the *severity* of complications, there is also a clinical and academic need to be able to define the *type* of complications that patients endure. The National Surgical Quality Improvement Program (NSQIP) have developed their own system for classifying complications[201] which has also been adopted by the Safe Surgery Saves Lives campaign.[113] An example of the NSQIP complications classification is shown in Table 6, taken from a paper published in 2005.

The data presented in Chapter 2 of this thesis show that there are subtle differences in the definitions of morbidity between NSQIP papers; however, broadly speaking, they have kept to this classification since 1995. No formal development or validation process has been published for this system.

System	Description
Cardiovascular	Cardiac arrest requiring cardiopulmonary resuscitation, myocardial infarction, pulmonary oedema
Respiratory	Pneumonia, unplanned intubations, pulmonary embolism, failure to wean from the ventilator >48h post-op
Gastrointestinal	Prolonged ileus
Renal	Progressive renal insufficiency, renal failure requiring dialysis, urinary tract infection
Neurological	Cerebrovascular accident (stroke), coma persisting more than 24 hours postoperatively, other neurologic deficits
Infectious	Sepsis
Wound	Superficial and deep wound infections, wound dehiscence
Thrombosis	Deep vein thrombophlebitis
Other	Bleeding requiring >4 U RBC transfusion, graft or prosthesis failure

Table 6: NSQIP definitions of postoperative complications taken from Alvord et al, 2005. [202]

1.6.4.4 The Post-Operative Morbidity Survey

The Post-Operative Morbidity Survey (POMS) was developed in the United States as a composite measure of clinically significant short-term postoperative harm,[183] and has been validated for this purpose in orthopaedic, general and urological surgical patients in the UK.[56] It classifies post-operative morbidity into nine domains; these were derived on the basis of being complications of a type and severity that would delay hospital discharge,[183] and could be defined by data that would be easy to collect and subject to minimal inter-observer variability.[56] The POMS is being used increasingly in clinical effectiveness studies,[165] and has been adapted for use in cardiac surgery.[203] The original POMS was not validated for use as a cumulative score, as it lacked the necessary internal consistency between domains for such use.[56] However, the C-POMS has been found to be valid as a score, with internal consistency of 0.7 (Cronbach's alpha) on days 5, 8 and 15 postoperatively; in addition, construct validity was demonstrated by a step-

wise rise in subsequent length of stay for each additional POMS domain on each of the postoperative days of measurement.[203] As this thesis is focussed on major non-cardiac surgery, the C-POMS will not be discussed further.

There are potential disadvantages of the POMS. The use of administered treatments (for example, antibiotics and parenteral opioids) as morbidity-defining criteria poses difficulties when using POMS to compare outcomes between different care providers. Enhanced recovery programmes, for example, encourage clinicians to limit parenteral opioid use; it may therefore be argued that direct comparison between two surgical teams, one who has implemented enhanced recovery protocols, and one who has not, may be inappropriate using the POMS. In addition, provision of 'as required' medication may be seen as an indication of good quality care; so a patient on one ward who is left in pain and who might have benefitted from parenteral opioids will be recorded as morbidity-free on the pain domain whereas a patient on a 'better' ward who is administered appropriate analgesia will be recorded as POMS positive. Finally, while the validation study demonstrated high inter-rater reliability, it is possible that this was because the data were collected by two dedicated research nurses: these nurses were highly trained, and it is therefore possible that without this training – and the resources required to implement this – that the reliability of the survey may be jeopardised. However, this potential problem is not limited to the POMS. In a study conducted in their own institution, Clavien and Dindo found that the quality of recording of complications according to their 2004 system was extremely poor when conducted by surgical residents (compared with the 'gold standard' of research nurses), even after the residents received training in how to collect the data. [194] Lack of motivation and lack of time are cited as likely reasons for this observation. The use of trained dedicated

nursing staff is therefore advocated; the NSQIP has made the required investment to adopt this approach.

Morbidity Type	Criteria
Pulmonary	New requirement for oxygen or respiratory support
Infectious	Currently on antibiotics and / or has developed a temperature of >/= 38°C in past 24 hours
Renal	Presence of oliguria (<500ml urine /24 hours) Increased serum creatinine (>30% from pre-operative level) Urinary Catheter in situ
Gastrointestinal	Unable to tolerate enteral diet for any reason including nausea, vomiting and abdominal distension Use of anti-emetic
Cardiovascular	Diagnostic tests and /or treatment for any of the following in the past 24 hours: New myocardial infarction or ischaemia Hypotension (requiring fluid therapy >200ml/h or pharmacological therapy) Atrial or ventricular arrhythmias Cardiogenic pulmonary oedema Thrombotic event requiring anticoagulation
Neurological	New focal neurological deficit, confusion, delirium or coma
Haematological	Requirement for any of the following within the past 24 hours Packed erythrocytes; Fresh frozen plasma; Cryoprecipitate
Wound	Wound dehiscence requiring surgical exploration or drainage of pus from the operation wound with or without isolation of organisms
Pain	New post-operative pain significant enough to require parenteral opioids or regional analgesia

Table 7: The Post-Operative Morbidity Survey (POMS)

Despite these caveats, the POMS has a number of advantages as an outcome measure. Unlike other previously mentioned systems for reporting surgical outcome, the POMS has been robustly scientifically validated, which in itself is important and supports its use both as an audit and quality improvement tool, and as an outcome measure in clinical studies. The validation study showed that the POMS was acceptable to patients, had acceptably low inter-observer variability and demonstrated construct validity, as the POMS was shown to discriminate between patients at different risk levels, and also

to predict postoperative length of stay.[56] Face validity was also achieved, in part by using simple objective criteria to define morbidity within each domain. The use of therapeutic interventions as criteria to define morbidity also has a number of advantages, despite the concerns raised above: namely high inter-rater reliability and face validity. While accepting that training may be required to ensure accurate data acquisition, the high inter-rater reliability reported is particularly important given the potential uses of a postoperative morbidity measure – namely, as an outcome measure in clinical effectiveness studies, for the purposes of comparative audit between teams and institutions, and to guide clinical decision making.

Experience at the Surgical Outcomes Research Centre (SOuRCe) has shown that while the POMS is a useful and valid measure of postoperative harm, some Trusts may not be able to provide the financial resources (personnel) required to implement it on four postoperative days. In addition, even in clinical trials, for the purpose of power calculations and clarity of results to the reader, there may be benefit to naming a single postoperative day on which to focus for POMS defined outcome. The original 1999 POMS development paper[183] has been cited 83 times and the 2007 validation paper[56] 27 times. On reviewing these citations, 11 studies using original data were found which had used the POMS as an outcome measure in the form it was originally developed; these are listed in Table 8. (There are more studies which have used the POMS but which are either analyses of the same data as presented in previous papers,[55] or have significantly modified it).[204] It is clear from this brief evidence synthesis, that there is disparity between studies in the endpoints at which POMS is measured, and it would be beneficial to have clear guidance to offer future investigators.

Lead Author	Publication year	Study Design	POMS days recorded
Bennett-Guerrero[24]	2001	Interventional	11
LeBuffe[205]	2004	Observational	2, 8
Ackland[206]	2007	Observational	3, 5, 8, 15
Scollay[207]	2009	Observational	5, 8, 15
Snowden[165]	2010	Interventional	7
Ackland[41]	2010	Observational	3, 5 8, 15
Davies[208]	2011	Interventional	5
Phypers[209]	2011	Interventional	5
Hennis[210]	2012	Observational	5
Ausania[211]	2012	Observational	Throughout hospital stay
Junejo[212]	2012	Observational	Throughout hospital stay

Table 8: Published papers using the Post Operative Morbidity Survey

The POMS is an outcome measure in several major perioperative medicine studies currently underway in the UK. (UKCRN IDs 4187, 8132, 6307, 9750, 10526 and 10666; see http://public.ukcrn.org.uk/search/)

1.6.5 Patient Reported Outcome Measures (PROMS)

Patient reported outcome measures (PROMs) are questionnaires which are completed by patients and measure health status or health related quality of life (HRQOL). PROMs are measured before and after a clinical intervention so as to measure the change in the patient's health status or HRQOL as a result of the intervention. PROMs may be disease-specific or generic and a large number of these tools have been developed and validated in different surgical cohorts. Generic questionnaires such as the Short Form -36 (SF-36) and EuroQOL (EQ-5D) are designed to reflect quality of life and may be particularly useful when assessing outcome in types of surgery which are principally aimed at

improving symptoms, such as joint replacements.[213] Disease-specific PROMs such as the Oxford Hip and Knee Scores, aim to assess patient health status, and demonstrate improved responsiveness in the detection of treatment effects when compared to generic measures.[214] Rasch analysis is a 'goodness of fit' estimation which may be used to evaluate instruments such as PROMs.[215]

In general terms, there are a number of biases which limit the usefulness of PROMs for comparative audit or in clinical effectiveness studies. For example, experiences that patients may report as complications may be viewed as 'normal' or 'expected' by physicians (for example wound pain); and process issues such as the mode of administration of the questionnaire (e.g. telephone versus mailing) may affect the responses that patients give.[216] It is issues such as these which may explain why feedback from PROMs to clinicians and managers has been shown to significantly influence process in healthcare, but has limited impact on patient health status. [217] Equally however, it may be that PROMs are being used inappropriately as a result of a lack of knowledge and experience of the healthcare providers who are implementing them. [217]

Despite these reservations, the implementation of PROMs was a key recommendation of Lord Darzi's 2008 report 'High Quality Care for All' in the United Kingdom.[218] The mandatory use of PROMs as a measure of outcome throughout the National Health Service has now been implemented for four operations, all of which are aimed at improving symptoms as opposed to saving or prolonging life: these are total hip replacement, total knee replacement, varicose vein repair and inguinal hernia repair. Healthcare providers are incentivised to participate, as higher rates of data completion are rewarded with a reduction in NHS Litigation Authority (NHSLA) tariff payments.

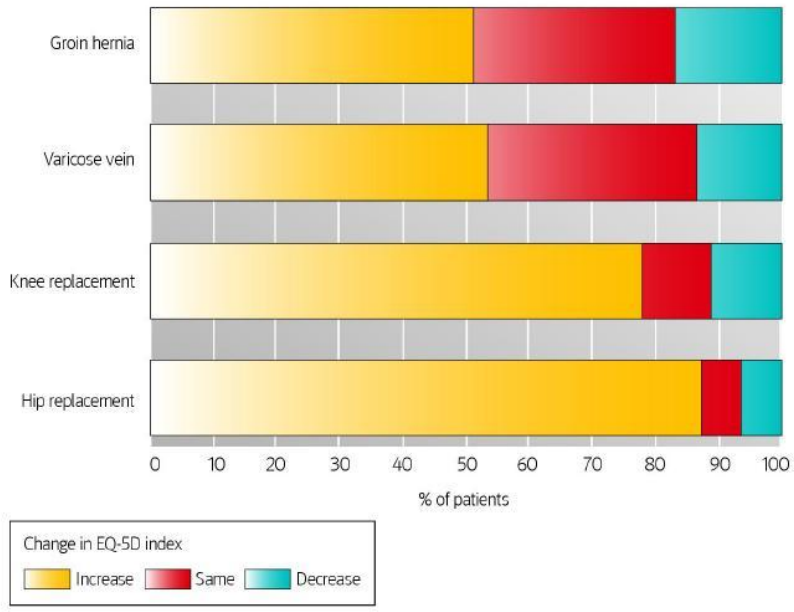

Figure 2: Change in Health Related Quality of Life (EQ5D score); 2009-10

(from Appleby ; BMJ 2011;343:d8191)

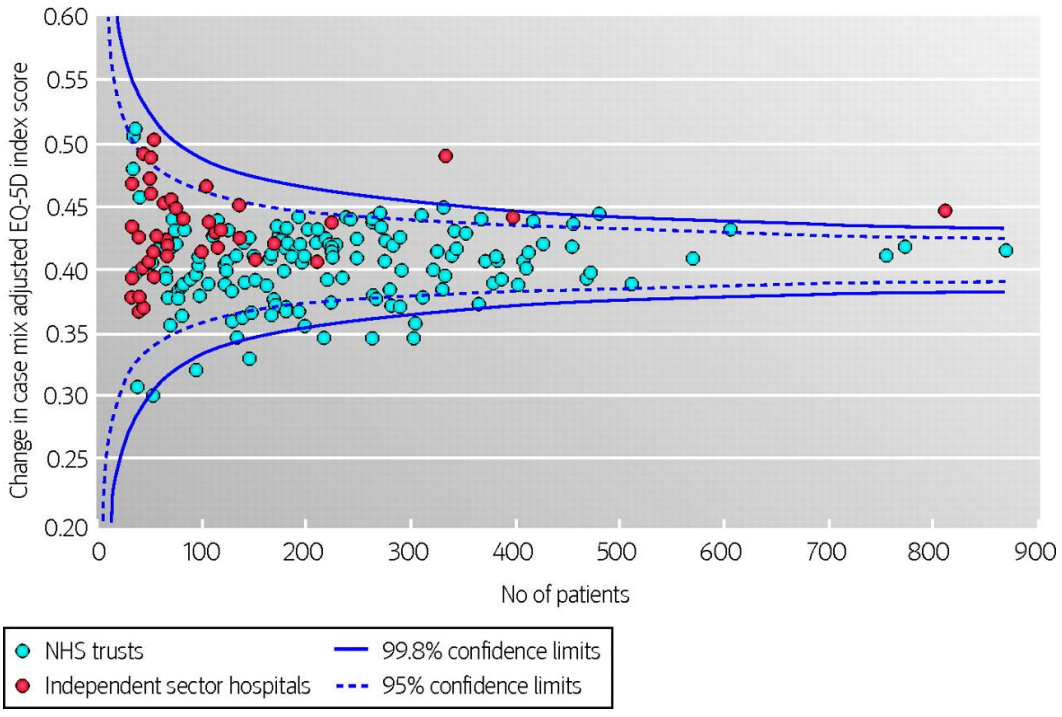

Figure 3: Changes in case-mix adjusted health related quality of life (EQ5D): independent and NHS providers (2009-10)

(from Appleby ; BMJ 2011;343:d8191)

The UK PROMS programme is now in its fourth year and interesting trends are apparent when examining the results. Taken at face value, outcomes after hernia and varicose vein repairs appear to be less beneficial to the overall patient population than joint replacements (see Figure 2). Independent sector hospitals also appear to be achieving better patient outcomes than some NHS providers (see figure 3). However, behind these figures are a number of confounding factors. First, there may be legitimate concerns raised over the case-mix adjustment methodology. Case-mix (or risk) adjustment is by linkage with Hospital Episode Statistics (HES) data; linkage rates vary between 75% for hernia repairs to 85% for hip replacements.[219] Adjustment is made for age, gender, socioeconomic status, general health status (preoperatively), whether the operation is a primary or revision procedure, and HES defined comorbidities. Concerns have been raised over the quality of administrative data in the UK, and the use of such data for risk-adjustment;[220;221] however, other researchers have devised complex models which have similar predictive precision to models using clinical data.[222] Second, there is evidence of response bias. Non-responders tend to be male, younger (under the age of 55 years), of non-white ethnicity, from the lowest quintile of social deprivation, live alone, require assistance with questionnaire completion and have more comorbidities.[219] There is also evidence of bias in terms of equity of use and patient reported outcome: patients who have more severe disease at the time of surgery and / or have had symptoms for longer, again tend to be from more deprived or ethnic minority (predominantly black) backgrounds.[219] All of these biases require consideration when interpreting data which appear to show higher improvements in patient reported outcomes in independent sector hospitals than NHS Trusts.

Some parts of the medical profession have expressed concern at the interpretation of PROMs data by the government and Department of Health, and alarm bells have been raised, particularly given the financial difficulty in which the NHS finds itself at the moment. Selective reporting of generic PROMs data, rather than disease specific for patients undergoing joint replacements has infuriated the British Orthopaedic Association, as unsurprisingly, generic PROMs show less improvement than the Oxford Hip and Knee scores, and the surgeons are concerned that these data will be used to rationalise healthcare spending.[223] Further alarm was raised by comments made by the Secretary of State for Health, Andrew Lansley, suggesting that in half of patients undergoing knee surgery, neither pain nor mobility was substantially improved.[224] Later clarification by the Department of Health seemed to suggest that his comments were based on generic PROMs outcome data: "The latest data from the NHS shows that only half of patients report any improvement in their health after a knee operation...."[224]

The counter argument to the orthopaedic surgeons' perspective, is that if the patient's general quality of life does not improve (or even deteriorates) after an operation aimed at improving quality of life, can the surgery really be viewed as a success, either by the patient, the surgeon or the wider healthcare system? While the surgical profession may be focussed on specific outcomes, perhaps the patient and the NHS are rightly more concerned with generic? A similar analogy may be drawn when considering clinical effectiveness studies which use disease-specific mortality rates and all-cause mortality rates as separate outcome measures. While the difference between these two might be of academic interest and importance to clinicians, it is likely that patients are less concerned with the detail of which treatment might lead to which cause of death or morbidity: avoiding all-cause death and morbidity is likely to be the main focus. The

sour old joke 'The operation was a success but the patient died" satirises this eloquently. Understanding, what the patient thinks is important, and keeping this foremost in one's thinking, is key to this area. Using data responsibly, and understanding its limits, particularly with relatively new outcome measures such as PROMs, is also paramount. The PROMs programme in the UK continues to expand, with pilot studies currently on-going in coronary artery bypass surgery and in various long-term medical conditions.

1.6.6 Patient satisfaction

The NHS increasingly focuses on patient satisfaction and patient experience as measures of the quality of healthcare provision.[218] The NHS Inpatient Survey is a measure of patient experience, which is conducted in every NHS Trust on an annual basis, and the results of which form part of the Care Quality Commission's assessment process. Patient experience measures focus on the patient's experience of the care environment: for example, staff attitudes, ward cleanliness, efficiency of transit through the hospital and so on. Patient satisfaction is a complex concept, and may encompass elements of both patient experience and patient reported outcome. The role of expectations, and the balance between these expectations and outcomes are key determinants of satisfaction for individual patients.[225] Determinants of patient satisfaction include 'patient related factors' (for example age, gender, ethnicity), 'provider care' (interactions with staff) and 'processes of care' (efficiency of transit through the hospital). Although it seems instinctive to believe that the competence of the provider is fundamental in the quality of care the patient receives, and therefore the patient's subsequent satisfaction, this is not always the case, as patients can find this a difficult concept to grasp.[226]

Given the complexity of the concept, it is clear that validated, psychometrically developed measures of patient satisfaction should be used to measure this outcome both in clinical practice and in research studies. However, a systematic review of the anaesthesia literature which I led, found that despite patient satisfaction being measured as an outcome in over 3000 published studies, fewer than 10% of these used a validated psychometrically developed instrument: most simply posed the question: 'are you satisfied?'.[227] Despite this, we also identified over thirty psychometrically developed and validated instruments which could be used to measure patient satisfaction with anaesthesia in a variety of clinical settings, including after the preoperative assessment process, during the perioperative period, in the maternity suite and in paediatric patients. We hope that this evidence synthesis will to help to guide future researchers and clinicians in their choice of instrument to use, so that the results of patient satisfaction surveys can be reliably and confidently used to measure and compare outcomes, and to improve quality of care.

1.6.7. Surrogate outcome measures

Structure and process measures have already been mentioned as potential indicators for the quality of patient care. In addition to these, surrogate outcome measures are used widely particularly in primary and secondary medical care. Surrogate outcome measures have be defined by the US Institute of Medicine as "biomarker[s] intended to substitute for a clinical endpoint [and] expected to predict clinical benefit (or harm . . .) based on epidemiologic, therapeutic, pathophysiological, or other scientific evidence."[228]

The dominance of surrogate outcome measures in diabetes mellitus care was highlighted in a recent review, and the pitfalls of this approach discussed.[229] For example, glycosylated haemoglobin (HbA1C) is used as a surrogate for glycaemic

control in the management of diabetes. However, two major concerns arise from this. First, HbA1C level may not truly reflect glycaemic control in patients with big swings of blood glucose level.[230] Secondly, as an outcome measure HbA1C is of little relevance to patients, compared with for example, renal failure, blindness and amputation which may all occur as a result of poor diabetic control.

Surrogates for morbidity and patient satisfaction are regularly used in the perioperative literature. Opioid consumption[231] and blood transfusion rates[232] are both examples of surrogate outcomes which have been used in clinical studies, but usually as secondary end-points, with primary outcomes of greater clinical importance (that is, morbidity or mortality). Surrogates in the perioperative field, as with general medicine, have some appeal as they are often more easily measured, and the results are usually available more quickly than true outcome measures. However, the relationship between surrogates and measures of true clinical outcome which are important to patients and doctors alike, is not always clear, and therefore surrogate outcomes should be carefully considered before acting on them. [233]

1.6.8 Resource utilisation measures

Length of inpatient or critical care stay, are measures which are sometimes used as surrogates of postoperative recovery and therefore clinical outcome; however, the validity of using such measures for this purpose is limited for a number of reasons. First, there are inherent assumptions made when using length of stay as a surrogate of clinical outcome: namely, that every patient is discharged from hospital or critical care at the same level of 'fitness', and that their discharge is not affected by factors unrelated to physiological status (such as provision of social services or availability of ward beds). Such assumptions may lead to both intra- and inter- institutional bias, as non-clinical

issues affecting length of stay will vary between and within institutions depending on factors such as the provision of convalescent facilities and the patient's own social support network. Rates of readmission to hospital or critical care are also used as surrogates of clinical outcome and are similarly subject to confounding, as the clinical threshold for admission and discharge will also vary between services.

Even as a measure of resource utilisation alone, length of stay is of limited validity, as there are different costs associated with different levels of intensity of treatment (such as critical care versus ward care). Furthermore, using hospital length of stay as a resource utilisation measure, assumes that the cost of treatment is consistent throughout the patient episode. This approach is unlikely to accurately reflect the true cost: for example, the cost of patient care on the first day post-operatively is likely to be much higher than the cost on the day before discharge. Finally, the actual cost of a hospital episode will differ between healthcare systems based on the method by which hospitals charge for their services. Despite these limitations, length of stay (both hospital and critical care) continues to be a widely used outcome measure in clinical effectiveness studies, predominantly because of the ease of measurement and the lack of inter-observer variability in recording.

1.6.9. Patient – centred outcomes

Patient centred care is a buzz phrase of the 21st century NHS. The US Institute of Medicine defined patient centred care as: *"Healthcare that establishes a partnership among practitioners, patients, and their families (when appropriate) to ensure that decisions respect patients' wants, needs, and preferences and that patients have the education and support they need to make decisions and participate in their own care."* [234]

A central tenet of delivering patient centred care must be to ensure that we measure patient-centred outcomes, so that the quality of the care provided can be reported in a way that is both meaningful and important to patients. However, while the principle seems obvious, the implementation may be problematic. Which outcomes matter will vary between patients, procedures, clinical and sociodemographic settings. For example, the relative importance of particular outcomes is likely to differ widely between patients having potentially life-saving cancer surgery (but which might leave the patient with a stoma) and a patient having cosmetic surgery or a varicose vein repair. These variances are merely the tip of the iceberg: the situation is further complicated by the potential conflict between which outcomes are considered important by the various members of the perioperative team, and what elements of 'quality care' each of these outcomes represent. For example, the surgeon may want a safe (safety) fast (efficiency) anaesthetist who never cancels operations (??efficiency); the patient wants a safe (safety), kind (patient experience) anaesthetist who will alleviate their pain and suffering (patient satisfaction /patient experience). Novel methods of reporting outcome data are emerging which aim to consider and resolve some of these conflicts,[235] but there remains a fundamental problem with precise risk adjustment, robust and accurate outcome data collection, and providing feedback to clinicians in a timely manner. One might assume that all members of the extended surgical team (including the patients) want to know morbidity and mortality data, although we know that these are not necessarily routinely reported or even recorded.

It is common sense that patients would wish to have efficient, safe, clinically effective care delivered to them by courteous and kind staff who communicate well, all of which

will leave them feeling satisfied with their experience – the central principles of quality in healthcare.[218] Moving beyond these basic principles, it is also likely that patients are interested in knowing the chances of 'success', how long they are likely to stay in hospital or be off work, the risk of complications, and the risk of death within a meaningful timeframe from their surgery. How we communicate these data to the patients is another area worthy of lengthy discussion, but beyond the scope of this thesis.

1.7 Summary: The importance of predicting risk and measuring outcome

Perioperative morbidity and mortality are significant public health issues.

Accurate risk prediction is a fundamental component of individualised patient-centred care, as it enables the perioperative pathway to be appropriately planned, and perhaps most importantly, it provides information required for high quality informed patient consent. Despite the recommendations of numerous reports, published evidence and clinical experience both suggest that perioperative risk stratification is poorly understood and implemented.

Routine, systematic outcome measurement is required to be able to measure the quality of healthcare, understand the relationship between risk and outcome, and to improve the quality of healthcare delivery. Risk adjustment of clinical data is required to meaningfully compare outcomes between teams and institutions, and to measure changes in performance over time. In the UK, the reporting of surgical outcomes is non-uniform and sporadic, and this is in part due to the lack of a national reporting system in the UK, and a lack of understanding of the types of complications which occur after surgery, and their short and long-term implications.

1.8 Aim of this Thesis

The aim of this thesis is to advance the understanding of the available risk stratification tools for perioperative morbidity and mortality, and the relationship between patient risk factors, short-term and longer-term surgical outcomes.

The first study is an evidence synthesis: a systematic review of validation studies of risk stratification tools in heterogeneous cohorts of surgical patients. This summarises the reported accuracy of the available measures, and makes recommendations for which tools are most appropriate for use in the UK.

The predictive accuracy and calibration of several of the risk stratification tools evaluated in the systematic review are then tested in a cohort of patients who underwent elective major non-cardiac non-neurosurgical procedures at the Middlesex Hospital between 2001 and 2005.

The epidemiology of short and long-term outcomes in this cohort of patients is then described, using the Post Operative Morbidity Survey as an objective validated measure of postoperative outcome, and mortality at several time-points.

Finally, a multivariable analysis of the relationship between perioperative risk, inpatient morbidity and long-term survival in this cohort of patients is presented, with discussion on the implications of these findings.

Chapter Two: Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review

2.1 Introduction

Accurate prediction of perioperative risk is an important goal – to enable informed consent for patients undergoing surgery and to guide clinical decision-making in the perioperative period. In addition, by adjusting for risk, an accurate risk stratification tool enables meaningful comparison of surgical outcomes between providers for the purposes of service evaluation or clinical audit. Some risk stratification tools have been incorporated into clinical practice, and indeed, have been recommended for these purposes. [236] [237]

Risk stratification tools may be sub-divided into risk scores and risk prediction models. Both are usually developed using multivariable analysis of risk factors for a specific outcome.[238] Risk scores assign a weighting to factors identified as independent predictors of an outcome, with the weighting for each factor often determined by the value of the regression coefficient in the multivariable analysis. The sum of the weightings in the risk score then reflects increasing risk. Risk scores have the advantage that they are simple to use in the clinical setting. However, while they may score a patient on a scale on which other patients may be compared, they do not provide an individualised risk prediction of an adverse outcome.[239] Examples of risk scores are the American Society of Anesthesiologists' Physical Status score (ASA-PSS)[62] and the Lee Revised Cardiac Risk Index (RCRI).[44]

By contrast, risk prediction models estimate an individual probability of risk for a patient by entering the patient's data into the multivariable risk prediction model. While risk prediction models may be more accurate predictors of an individual patient's risk than risk scores, they are more complex to use in the day-to-day clinical setting.

Despite increasing interest in more sophisticated risk prediction methods, such as the measurement of functional capacity by exercise testing,[240] risk stratification tools remain the most readily accessible option for this purpose. However, clinical experience tells us that they are not commonly used in everyday practice.[237;241] Lack of use may be due to poor awareness amongst clinicians of the available options, and concerns regarding their complexity and accuracy.[242] In other clinical settings, low uptake of risk stratification tools has been ascribed to a lack of clarity on the precision of available tools, resulting from perhaps unnecessary efforts to make minor refinements to existing methods, or to developing novel methods, with the aim of achieving greater predictive accuracy. [243]

Thus the focus of this chapter is an evidence synthesis of the available risk stratification tools in perioperative care, in order to make recommendations about which methods are appropriate for use both in clinical practice and in research.

2.1.2 Aim and objectives

2.1.2.1 Aim

To conduct a qualitative systematic review of the available evidence to answer the following specific question: 'What is the performance of risk stratification tools, validated for morbidity and/or mortality, in heterogeneous cohort of surgical (non-cardiac, non-neurological) patients?'

2.1.2.2 Objectives:

1. To summarize the available risk prediction methods
2. To report on their performance (discrimination and calibration)
3. To comment on their strengths and weaknesses, with particular focus on accuracy and ease of application.

2.2 Methods

Previously published standards for reporting systematic reviews of observational studies were adhered to when undertaking this study.[244]

2.2.1 Definitions for the purposes of this study

A 'risk stratification tool' was defined as a scoring system or model used to predict or adjust for either mortality or morbidity after surgery, and which contained at least two different risk factors. 'Major surgery' was defined as a procedure taking place in an operating theatre and conducted by a surgeon; thus, studies of cohorts of patients undergoing endoscopic, angiographic, dental, and interventional radiological procedures were excluded. A 'heterogeneous patient cohort' was defined as a cohort of patients including at least two different surgical specialities. Studies of gastrointestinal surgery, which included hepatobiliary surgery, were included. We excluded studies that consisted entirely of cohorts undergoing ambulatory (day case) surgery, and cohorts that included cardiac or neurological surgery.

2.2.2 Search strategy and study eligibility

A search for papers published between 1 January 1980 and 6 August 2011 was undertaken using Medline, Embase and ISI Web of Science. No language restriction was applied. The search strategy, and inclusion / exclusion criteria are detailed in Appendix 1. Of note, papers reporting development studies were excluded, unless the paper included validation in a separate cohort.

2.2.3 Data extraction and quality assessment of studies

Data extraction was undertaken using standardised tables relating to the study characteristics, quality and outcomes. Study characteristics extracted from each paper

included the number of patients, the country where the study was conducted, the outcome measures and end-points of each study, and the risk stratification tools being assessed. Data were also extracted regarding the most detailed description of the types of surgery included in each study cohort reported in the manuscripts. Clinical outcome data (morbidity and mortality) for the cohorts in each study were also extracted from the manuscripts.

Assessment of study quality was based on the framework for assessing the internal validity of articles dealing with prognosis developed by Altman.[245;246] The following criteria were used: the number of patients included in analyses; whether the study was conducted on a single or multiple sites; the timing of data collection (prospective vs. retrospective); whether a description of baseline characteristics for the cohort was included (including comorbidities, type of surgery and demographic data); and selection criteria for patients included in the study (to assess for selection bias). Selection bias was judged to be present if a study restricted the type of patient who could be enrolled based on age, ethnicity, gender, premorbid condition, urgency of surgery or post-operative destination (e.g. critical care). In addition, the setting of each validation study was recorded–that is, whether the validation was internal (that is, conducted in a split sample of the original development cohort), temporal (conducted in the same institution(s) as the development study but in a later cohort of patients) or external (that is, that the validation cohort was entirely different to that in which the tool was developed).[247] Finally, as a measure of the clinical usability and reproducibility of each risk stratification tool, an assessment was made of whether each tool used variables which were objective (for example blood results), subjective (for example chest radiograph interpretation) or both.[248]

2.2.4 Data analysis and statistical considerations

The performance of each risk stratification tool was evaluated using measures of discrimination and, where appropriate, calibration. Discrimination (how well a model or score correctly identifies a particular outcome) was reported using either the area under the receiver operator curve (AUROC) or the concordance (c-) statistic. An AUROC or c-statistic of < 0.7 was to taken to indicate poor performance, 0.7-0.9 to be moderate, and >0.9 to reflect high performance.[249] Calibration is defined as how well the prognostic estimation of a model matches the probability of the event of interest across the full range of outcomes in the population being studied. Where reported, either Hosmer-Lemeshow or Pearson chi-squared statistics were extracted as an evaluation of calibration; $p>0.05$ was taken to indicate that there was no evidence of lack-of-fit.

2.3 Results

2.3.1 Search results

139,775 articles on Medline and 71,841 on Embase were listed in the initial search and the titles and abstracts of these were screened to identify articles which described risk stratification tools used in any adult non-cardiac, non-neurological surgery. 751 articles then underwent manuscript review. Hand searching of reference lists and citations identified a further 432 studies which were also reviewed in detail.

Three studies were identified which graphically displayed ROC curves in their results but did not report AUROCs.[38;250;251] The authors of these studies were contacted for additional information; none responded so these studies were excluded from the analysis. Six foreign language studies which may have been eligible for inclusion based on review of the abstracts, but for which we were unable to obtain translations, were also omitted from the analysis.[252-257] The flow chart for the review is detailed in Figure 4.

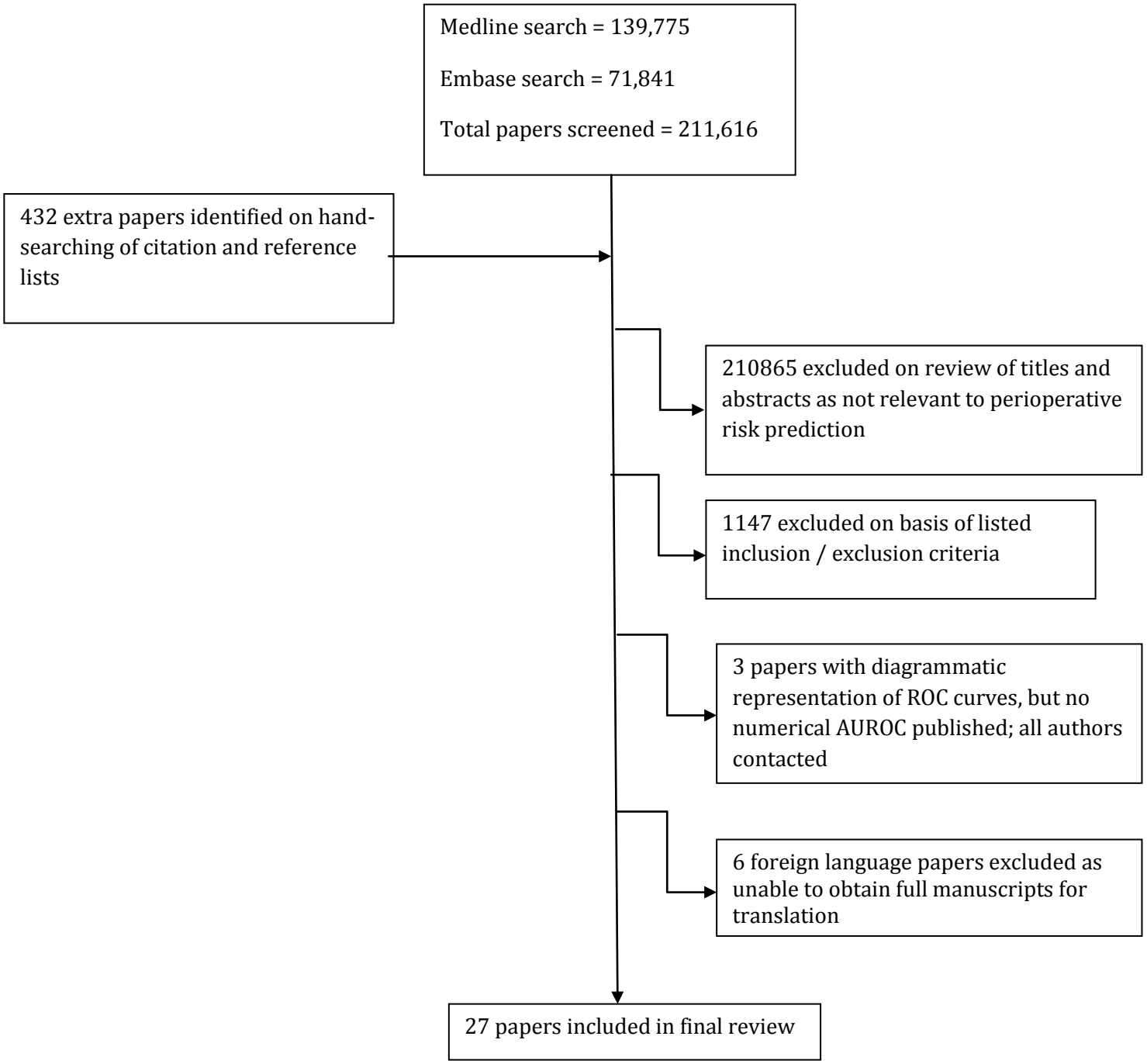

Figure 4: Flow diagram for the systematic review

A total of 27 studies, evaluating 34 risk stratification tools were included in the analysis. All were cohort studies. Eight tools were validated in multiple studies; the most commonly reported were the American Society of Anesthesiologists' Physical Status Score[62] (ASA-PS) (4 studies, total number of patients, n =4014), the Acute Physiology and Chronic Health Evaluation II scoring system,[258] (APACHE II) (4 studies, n=5897) the Physiological and Operative Score for the enUmeration of Mortality and Morbidity [38] (POSSUM) (3 studies, n=2915); the Portsmouth variation of POSSUM[259] (P-POSSUM) (5 studies, n=10648; mortality model only); the Surgical Risk Scale[46] (SRS), (3 studies, n=5244; mortality model only), the Surgical APGAR Score[260] (3 studies, n= 10795), the Charlson Co-morbidity Index [261] (2 studies, n=2,463,997) and Donati's Surgical Risk Score [47] (2 studies, n=7121). The accuracy of a further 26 tools was evaluated in single validation studies. A comparison of tools that were validated in multiple studies is detailed in Tables 9 and 10. The general characteristics of all included studies are summarised in Table 11.

Table 9: Mortality models validated in multiple studies

Model	No of varia-bles	Pre-intra- or post op data used	Original derivation cohort and outcome	Studies (n)	Author and year	Patients (n)	Type of surgery	Surgical urgency	End-point	AUROC (CI)
APACHE II	16	Post	Critical care patients; all diagnoses (not just surgical); Hospital mortality [258]	3	Jones 1992 [262]	117	GI, vascular, renal and urology	All	30 day	HDU admission score: 0.539 (+/- 0.083)
					Osler 1998 [263]	5322	Non-cardiac	All	Hospital discharge	ICU admission score: 0.806
					Stachon 2008	271	Ortho, spinal, trauma, visceral surgery, limb surgery	All	Hospital discharge	First 24h worst score: 0.777
ASA-PS	1	Pre	General surgical patients[62]	2	Sutton 2002 [46]	1946	GI, vascular, trauma	All	Hospital discharge	0.93 (0.90-0.97)
					Donati 2004 [47]	1849	Abdominal, vascular, ortho, urology, endocrine, ENT, neuro, Gynae, eye, thoracic, other	All	Hospital discharge	0.810 (0.792-0.828)
Charlson	17	Pre	Medical patients; 10 year mortality[264]	2	Atherly 2004 [265]	2167	General, vascular		30 day	0.52
					Sundarajan 2004 [266]	2,461,830	All in-patient surgery	All	Hospital discharge	0.85-0.87 (varied with year & if ICD-9 or ICD-10 used)
POSSUM	18	Pre and intra	General surgery; 30-day mortality[38]	3	Jones 1992 [262]	117	GI, vascular, renal and urology Elective & emergent	All	30 day	0.753 (+/-0.081)
					Donati, 2004 [47]	1849	See above	All	Hospital discharge	0.915 (0.884-0.947)
					Brooks 2005 [267]	949	General, colorectal, upper GI, urology, head and neck	All	30 day	0.92 (0.90-0.95)

P-POSSUM	18	Pre and intra	General surgery; 30 day mortality[69;259]	5	Organ 2002 [268]	229	General, vascular, ENT, plastics, thoracic, urology, other	All	30 day	0.68 (0.57-0.78)
					Donati, 2004 [47]	1849	Abdominal, vascular, ortho, urology, endocrine, ENT, neuro, Gynae, eye, thoracic, other	All	Hospital discharge	0.912 (0.898-0.924)
					Brooks 2005 [267]	949	General, colorectal, upper GI, urology, head and neck	All	30 day	0.92 (0.90-0.95)
					Neary 2007 [269]	2349	General, vascular, ENT, urology, ortho, other	Emergent & urgent	30 day	0.90 (0.87-0.93)
									1 year	0.90 (0.8-1.0)
					Haga 2011 [270]	5272	GI and HPB	Elective	30 day	0.74 (0.63–0.86)
									Hospital discharge	0.81 (0.75–0.88)
Surgical APGAR	3	Intra	Colorectal; 30 day mortality[260]	2	Regenbogen 2009 [271]	4119	General and vascular	All	30 day	0.81
					Haynes 2011 [272]	5909	Any non-cardiac	All	Inpatient	0.77
Surgical Risk Scale	3	Pre	General surgery; inpatient mortality[46]	3	Sutton 2002 [46]	1946	GI, vascular, trauma	All	Hospital discharge	0.95 (0.93-0.97)
					Brooks 2005 [267]	949	General, colorectal, upper GI, urology, head and neck	All	30 day	0.89 (0.86-0.93)
					Neary 2007 [269]	2349	General, vascular, ENT, urology, ortho, other	Emergent & urgent	30 day	30 day: 0.85 (0.82-0.89)
									1 year	1 year: 0.84 (0.75-0.94)
Surgical Risk Score (Donati)	3	Pre	General surgery; inpatient mortality[47]	2	Donati 2004 [47]	1849	Abdominal, vascular, ortho, urology, endocrine, ENT, neuro, Gynae, eye, thoracic, other	All	Hospital discharge	0.888 (0.838-0.937)
					Haga 2011 [270]	5272	GI and HPB	Elective	Hospital discharge	0.73 (0.63-0.83)

Table 10: Morbidity models validated in multiple studies

Model	Number of validation studies	Pre-intra or post- op varia-bles	Original derivation cohort and outcome	Author	N	Type of surgery and urgency	Surgical urgency	Endpoint	AUROC for outcome
ASA-PS	3	Pre	General surgery[62]	Goffi 1999 [273]	187	General	All	30 day (mortality and morbidity combined)	0.777
				Hightower 2010 [274]	32	Major abdominal (GI, urology)	Elective	7 day	0.688 (0.523 - 0.851)
				Makary 2010 [275]	594	Unselected inpatient	All	Hospital discharge	0.626
APACHE II	1	Post	Critical care patients; any diagnosis (not just surgical); Hospital mortality [258]	Goffi 1999 [273]	187	General	All	30 day (mortality and morbidity combined)	Hospital admission score: 0.866
									Pre-op score: 0.894
POSSUM	2	Pre & intra	General surgery; 30-day morbidity [38]	Jones 1992 [262]	117	GI, vascular, renal & urology; elective & emergent	All	30 day	0.82
				Brooks 2005 [267]	949	General, colorectal, upper GI, urology, head and neck	All	30 day	0.92
Surgical APGAR	3	Intra	Colorectal; 30 day mortality [260]	Gawande 2007 [260]	767	General and vascular	All	30 day (Mortality and morbidity combined)	0.72
				Regenbogen 2009 [271]	4119	General and Vascular	All	30 day	0.73
				Haynes 2011 [272]	5909	Any non-cardiac	All	Inpatient	0.70

2.3.2 Quality assessment

The quality assessment of included studies is summarised in Table11. Seven studies were multi-centre and 21 were single-centre. The data collection was prospective in 19 studies, retrospective in seven, and based on administrative data in two studies. 16 studies used mortality as an outcome measure, four used morbidity, and eight used both. The study endpoints included 30-day outcome in 12 papers, hospital discharge in 15 papers and three papers also included shorter or longer follow-up times ranging from one day to one year. 19 studies out of the total 28 reported baseline patient characteristics of physiology / comorbidity, surgery and demographics; selection bias was evident in 12 studies.

Table 11: Characteristics of all included studies

First Author	Year	Region	N	No of centres	Data acquisition	Selection bias	Subject description	Type of surgery	Surgical Urgency	Models used	Validation cohort: internal vs. external vs. temporal*	Outcome	End point
Atherly[265]	2004	US	2167	M	Administrative (ICD-9 codes)	N	N	General, vascular	All	Charlson Comorbidity Index based on ICD-9 codes	External	Mortality	30 day
Brooks[267]	2005	UK	949	S	Prospective	N	N	General, colorectal, upper GI, urology, head and neck	All	POSSUM P-POSSUM Surgical Risk Scale	Temporal	Mortality	30 day
DasGupta[276]	2009	Canada	125	S	Prospective	Y: >70y only	Y	General, abdominal, ortho, neuro-surgery, carotid surgery	Elective	Detsky Index Edmonton Frail Scale	External	Morbidity	Hospital discharge
Davenport[277]	2006	US	5878	S	Prospective	N	Y	General, neuro-surgery, ortho,	All	ASA-PS	External	Morbidity Mortality	30 day

								plastic, thoracic, vascular					
Donati[47]	2004	Italy	1849	M	Prospective	N	Y	Abdominal, vascular, ortho, urology, endocrine, ENT, neuro, Gynae, eye, thoracic, other	All	Surgical Risk Score POSSUM P-POSSUM	Temporal	Mortality	Hospital discharge
Gawande[260]	2007	US	767	S	Retrospective	N	Y	General and vascular	All	Surgical APGAR score	Temporal	Major complications or Mortality (Combined end-point)	30 day
Goffi[273]	1999	Italy	187	S	Prospective	N	N	General	All	ASA APACHE II on hospital admission APACHE II immediately	External	Combined end-point: Mortality Morbidity	30 day

										pre-op			
Hadjianastassiou[278]	2004	UK	4494	S	Retrospective	N	Y	Maxillo-facial, general, ortho, renal, urology, neuro	All	Surgical Mortality Score	Internal	Mortality	Hospital discharge
Haga[270]	2011	Japan	5272	M	Prospective	N	Y	GI and HPB	Elective	E-PASS mE-PASS P-POSSUM Surgical Risk Score (Donati)	External	Mortality	Hospital discharge
Haynes[279]	2011	International	5909	M	Prospective	N	Y	Any non-cardiac	All	Surgical APGAR	External	Mortality Morbidity	Hospital discharge
Hightower[274]	2010	US	32	S	Prospective	Y: major abdominal & fit enough for CPET	Y	Major abdominal (GI, urology)	Elective	ASA-PS	External	Morbidity	7 day
Hobson[280]	2007	UK	163	S	Prospective	Y: emergent surgery	N	General, gynae, renal,	Emergent	POSSUM	External	Mortality	30 day

						only		urology, vascular		P-POSSUM			
Jones[262]	1992	UK	117	S	Prospective	Y: HDU admissions only	N	GI, vascular, renal and urology	All	POSSUM APACHE II	External	Morbidity Mortality	30 day
Kuzu[281]	2006	Turkey	460	S	Prospective	N	Y	GI, vascular, HPB, gynae	Elective	Nutritional Risk Index Maastricht Index Subjective Global Assessment Mini Nutritional Assessment	External	Mortality Morbidity	Hospital discharge or 30day (whichever later)
Liebman[282]	2010	Nether-lands	33224	S	Prospective	N	Y	General and trauma	Emergent	IRIS (Identification of Risk in Surgical Patients)	Internal	Mortality Morbidity	Hospital discharge
Makary[275]	2010	US	594	S	Prospective	Y: elective only	Y	Un-selected inpatient	Elective	ASA, Lee RCRI and Eagle Scores alone	External	Morbidity	Hospital discharge

										and in combination with Frailty Index			
Nathanson[283]	2009	US	13,417	M	Retrospective	Y: post-op ICU admission only	Y	All excluding cardiac, neuro and trauma	Elective & emergent in separate cohorts	MPM_0-III	External	Mortality	Hospital discharge
Neary[269]	2007	UK	2349	S	Prospective	Y: emergent & urgent only	N	General, vascular, ENT, urology, ortho, other	Emergent and urgent	RCRI P-POSSUM Surgical Risk Scale BHOM	External	Mortality	30day and 1 year
Organ[268]	2002	Aus-tralia	229	S	Retrospective	Y: ICU only	N	General, vascular, ENT, plastics, thoracic, urology, other	All	P-POSSUM	External	Mortality	30day
Osler[263]	1998	US	5322	S	Retrospective	Y: ICU only	N	Non-cardiac	All	APACHE II International	External	Mortality	Hospital discharge

										Classification of Disease Illness Severity Score (ICISS)			
Pillai[284]	1999	New Zealand	6492	M	Retrospective	N	Y	GI, breast, endocrine vascular, gynae ortho, HPB	All	Otago Surgical Audit Score	External	Morbidity	Hospital discharge
Regenbogen[271]	2009	US	4119	S	Prospective	N	Y	General and vascular	All	Surgical APGAR Score	External	Mortality Morbidity	30 day
Stachon[285]	2008	Germany	271	S	Prospective	Y: ICU only	Y	Ortho, spinal, trauma, visceral surgery, limb surgery	All	APACHE II SAPS II APACHEN SAPSN	External	Mortality	Hospital discharge
Stachon[286]	2008	Germany	283	S	Prospective	Y: ICU only	Y	Ortho, spinal, trauma, visceral surgery, limb surgery	All	DELAWARE (Dense Laboratory Whole Blood Applied Risk Estimation) APACHE II	Temporal / external	Mortality	Hospital discharge

										SAPS II			
Story[287]	2009	Australia	256	S	Retrospective	Y: >70y only	Y	General, colorectal, ortho, plastics, urology, vascular, other	All	Perioperative Mortality Risk Score	Internal	Mortality	30 day
Sunda-rarajan[266]	2004	Australia	2,461,830	M	Administrative (ICD-9 & 10 codes)	N	Y	All in-patient surgery	All	Charlson Comorbidity Index using administrative data (ICD-9 and ICD-10AM coding)	External	Mortality	Hospital discharge
Sutton[46]	2002	UK	1946	S	Prospective	N	N	GI, vascular, trauma	All	Surgical Risk Scale; ASA-PS	Temporal	Mortality	Hospital discharge

Single centre= S; Multicentre = M; Y=Yes; N=No; ICU = Intensive Care Unit

* Definitions of validation cohorts: External = validation in new cohort unrelated to derivation study; Internal = Validation in split sample of same study population as derivation cohort; Temporal = validation in new cohort from derivation study but same institution(s)

2.3.3 Outcomes reporting

Outcomes are summarised in Table 12. Surgical mortality at 30 days varied between 1.3% [265] and 12.2%,[268] and at hospital discharge between 0.8% [276] and 24.7%.[286]

All but one[262] of the six studies which separately tested the discrimination of stratification tools for morbidity and mortality reported that morbidity prediction was less accurate. There was considerable heterogeneity in the definition of morbidity in the 12 studies that reported this outcome (see Appendix 2 for summary), and in keeping with this, there was wide variation in complication rates in different studies (between 6.7%[277] and 50.4%).[262]

2.3.4 Calibration

Calibration was poorly reported: 16 studies did not report calibration at all; of the remaining 11 papers, two reported only whether the models were of 'good fit', without reporting the appropriate statistics. One paper did not report calibration in their results, despite stating in the methods that they would calculate it.[275]

Table 12: Outcomes, discrimination and calibration

Author	Models used	End point	Morbidity (%)	AUROC morbidity (95% C.I.)	Mortality (%)	AUROC mortality (95% C.I.)	Calibration (p value for HL statistic unless otherwise stated)
Atherly, 2004[265]	Charlson Comorbidity Index using ICD-9 coding	30 day	NR	NR	1.3	0.47	NR
Aust, 2005[288]	UH formula based on NSQIP formula	30 day	NR	NR	1.4	UH formula: 0.915	NR
	UH formula with op severity added					UH with op severity: 0.941	NR
	Bedside UH formula					Bedside UH formula: 0.816	NR
Brooks, 2005[267]	POSSUM	30 day	NR	NR	8.4	POSSUM: 0.92	NR
	P-POSSUM					P-POSSUM: 0.92	NR
	Surgical Risk Scale					SRS: 0.89	NR
DasGupta, 2009[276]	Detsky Index	Hospital discharge	25	Detsky: 0.51 (0.39-0.63)	0.8	NR	NR
	Edmonton Frail Scale			EFS: 0.69 (0.58-			NR

				0.79)			
Davenport, 2006[277]	NSQIP	30 day	6.7	NSQIP: 0.769	1.5	NSQIP: 0.958	NR
	ASA			ASA-PS: 0.722		ASA-PS: 0.889	NR
	ASA and NSQIP combined			NSQIP RF with ASA-PS: 0.782		NSQIP RF with ASA-PS: 0.960	NR
Donati, 2004[47]	Surgical Risk Score	Hospital discharge	NR	NR	1.9	SRS: 0.888 (0.838-0.937)	0.744
	POSSUM					POSSUM: 0.915 (0.884-0.947)	0.0004
	P-POSSUM					P-POSSUM: 0.912 (0.898-0.924)	0.1528
	ASA					ASA: 0.810 (0.792-0.828)	NR
Gawande, 2007[260]	Surgical APGAR score	30 day	9.1	NR	1.4	Combined outcome of mortality and morbidity: 0.72	Pearson's goodness of fit: 0.57
Goffi, 1999[273]	ASA Preoperative APACHE II	30 day	Overall: 26.7 Elective: 15.9 Emergent:	NR	Overall: 8.6 Elective: 4.3 Emergent:	Combined outcome of mortality and morbidity:	NR
						ASA: 0.777	NR
						Hospital Admission APACHE II:	NR

			57.1		20.4	0.866			
						Immediate pre-op APACHE II: Overall: 0.894 Elective surgery: 0.826 Emergent surgery: 0.873 Cancer surgery: 0.915 Non-cancer surgery: 0.869			NR
Hadjuanastassiou, 2004 [278]	Surgical Mortality Score	Hospital discharge	NR	NR	4.1	0.82 (0.78-0.85)			0.10
Haga, 2011 [270]	E-PASS, mE-PASS, P-POSSUM, Surgical Risk Score (Donati)	Hospital discharge 30 day	NR	NR	NR		Hospital discharge	30d	
						E-PASS	0.86 (0.79–0.93)	0.82 (0.69–0.95)	NR
						mE-PASS	0.86 (0.79–0.92)	0.81 (0.66–0.96)	
						P-POSSUM	0.81 (0.75–0.88)	0.74 (0.63–0.86)	

						Surgical Risk Score	0.73 (0.63–0.83)	-	
Haynes, 2011[279]	Surgical APGAR	Hospital discharge	9.2 (major)	0.70	1.4	0.77			NR
Hightower, 2010[274]	ASA	7 day	50	0.688 (0.523 - 0.851)	NR	NR			NR
Hobson, 2007[280]	POSSUM P-POSSUM	30 day Hospital discharge	NR	NR	30 day: 9.2 Hospital discharge: 12.9	30 day: POSSUM: 0.946 PPOSSUM: 0.940			NR
Jones, 1992[262]	POSSUM APACHE II	30 day	50.4	POSSUM: 0.82	11.1	POSSUM: 0.75 APACHE II: 0.54			NR
Kuzu, 2006[281]	Subjective Global Assessment Nutritional Risk Index Maastricht Index	Hospital discharge or 30 day (which-ever later)	28.47	SGA:0.669 NRI:0.659 MI:0.671	4.34	SGA: 0.687 NRI:0.797 MI:0.743			NR

Liebman, 2010[282]	IRIS	Hospital discharge	13.3	0.77	2.2	0.90	NR
Makary, 2010[275]	ASA , Lee and Eagle with and without Frailty Index added	Hospital discharge	Not stated for entire cohort	ASA: 0.626 ASA + Frailty: 0.699 Lee: 0.618 Lee + Frailty: 0.669 Eagle: 0.678 Eagle + Frailty: 0.714	NR	NR	NR (but reported that this would be calculated in methods)
Nathanson, 2009[283]	MPM_0-III	Hospital discharge	NR	NR	Elective: 5.3 Emergent: 14.4	Elective:0.79 Emergency: 0.79	Good fit
Neary, 2007[269]	RCRI	30 day	NR	NR	30 day:	RCRI:	NR

		and 1 year			6.0 1 year: 10.8	30 day: 0.73 1 year: 0.71	
	P-POSSUM					P-POSSUM: 30 day: 0.90 1 year: 0.90	Good fit
	Surgical Risk Scale					SRS: 30 day: 0.85 1 year: 0.84	Good fit
	BHOM					BHOM: 30day: 0.84 1 year: 0.86	Good fit
Organ, 2002[268]	P-POSSUM	30 day	NR	NR	12.2	0.68	<0.001
Osler, 1998[263]	APACHE II	Hospital discharge	NR	NR	13.9	APACHE II: 0.806	0.002
	International Classification of Disease Illness Severity Score (ICISS)					ICISS: 0.892	0.15
	APACHE and ICISS combined					Combined: 0.903	0.038

Pillai, 1999[284]	Otago Surgical Audit Score	Hospital discharge	NR for validation cohort	0.86	NR	NR	Good fit
Regenbogen, 2009[271]	Surgical APGAR Score	30 day	14.1	0.73	2.3	0.81	NR
Stachon, 2008[285]	APACHE II	Hospital discharge	NR	NR	24.7	APACHE II: 0.777	NR
	SAPS II					SAPS II: 0.785	
	APACHEN					APACHEN: 0.829	
	SAPSN					SAPSN: 0.823	
Stachon, 2008[286]	Dense Laboratory Whole Blood Applied Risk Estimation (DELAWARE)	Hospital discharge	NR	NR	23.3	DELAWARE: 0.813	0.44
	APACHE II					0.777	NR
	SAPS II					0.785	NR
Story, 2009[287]	Perioperative Mortality Risk Score	30 day	NR	NR	6.0	0.79	0.35
Sundararajan, 2004[266]	Charlson Comorbidity Index using administrative data (ICD-9 and ICD-10AM coding)	Hospital discharge	NR	NR	Overall mortality not reported	ICD-9 1996-7: 0.87 ICD-9 1997-8: 0.86 ICD-10 1998-9: 0.85 ICD-10 1999-2000: 0.86	NR

						ICD-10 2000-1: 0.86 ICD-10 2001-2: 0.85	
Sutton, 2002[46]	Surgical Risk Scale	Hospital discharge	NR	NR	2.41	SRS: 0.95	0.65
	ASA-PSS					ASA: 0.93	NR

NR = Not reported

2.3.5 Risk stratification tools using only preoperative data

Four entirely preoperative risk stratification tools (ASA-PSS, Surgical Risk Scale [SRS], Surgical Risk Score, and the Charlson co-morbidity index, [CACI]) were validated in multiple studies. The Surgical Risk Scale and the Surgical Risk Score both contain the ASA-PSS, and the urgency and severity of surgery; both have also been multiply validated. The Surgical Risk Score[47;270] was developed and originally validated in Italy[47] and contains the ASA-PSS, a 3-point scale modification of the Johns Hopkins surgical severity criteria and a binary definition of surgical urgency (elective vs. emergency). The only published study evaluating the Surgical Risk Score after its initial validation found it to be poorly predictive of inpatient mortality.[270] The Surgical Risk Scale[46;267;269] uses the ASA-PSS alongside UK definitions of operative urgency (a 4-point scale defined by the UK National Confidential Enquiry into Post Operative Death and Outcome) and severity (the British United Provident Association [BUPA] classification which is used to rank surgical procedures for the purposes of financial billing in the private sector). Both studies validating this system after its initial development found it to be a moderately discriminant tool (AUROC>0.8).[267;269]

A further 18 different risk stratification tools using solely preoperative data were validated in single publications. These tools are described in Appendix 3.

2.3.6 Risk stratification tools incorporating intra- and post-operative data

The POSSUM and P-POSSUM scores were the most frequently used in heterogeneous surgical cohorts. The POSSUM score was derived by multivariable logistic regression

analysis and contains 18 variables, of which 12 are measured preoperatively and six at hospital discharge; two separate equations, for morbidity and mortality, were developed and validated.[38;289] After recognition that the POSSUM model over-predicted adverse outcome, the Portsmouth variation (P-POSSUM), was developed to predict mortality, using the same composite variables but a different calculation.[259] P-POSSUM has been used in a larger number of more recent studies[47;267-270;270] than the original POSSUM[47;262;267] and with the exception of one UK study, [40] has been found to be of moderate to high discriminant accuracy (AUROC varying between 0.68 and 0.92).

2.3.7 Models developed for purposes other than perioperative risk stratification

The APACHE II scoring system was originally developed as a case-mix adjustment tool for use in critical care;[258] the score consists of 12 physiological variables and an assessment of chronic health status. Only one of the four studies reporting the APACHE II score's predictive precision used it in the way originally intended: by incorporating the most deranged physiological results within 24 hours of critical care admission.[258] One small study[273] scored APACHE II on admission to hospital and immediately pre-operatively and found it to be highly predictive of outcome. This approach has face validity as APACHE II is a summary measure of acute physiology and chronic health, both of which may influence surgical outcome.

The Charlson comorbidity score was developed to predict 10-year mortality in medical patients.[264] A combined age-comorbidity index was subsequently validated for the

prediction of long-term mortality in a population of patients who had essential hypertension or diabetes and were undergoing elective surgery. [261] It is the original Charlson score, however, which is used in two studies identified in this search to stratify risk of short-term outcome. [265;266] These two studies reported very different predictive accuracy for the Charlson score; however the largest single study included in this entire review found the Charlson score (measured using administrative data) to be a moderately accurate tool.[266]

One study evaluated three scoring systems which were originally developed to classify patient nutritional status.[281] All of these models demonstrated moderate or high precision for predicting postoperative mortality. It is perhaps unsurprising that these measures predicted outcome, as different measures of nutritional status are associated with adverse outcomes in general populations of hospitalised patients.[290;291]

Two further papers reported the precision of scores incorporating measures of patient frailty.[275;276] The Edmonton Frail Scale had moderate precision in the prediction of surgical morbidity.[276] The addition of the Frailty Index to existing risk predictors such as the ASA-PSS, Lee Revised Cardiac Risk Index and the Eagle score improved their performance.[275] Again, frailty has face validity as a predictor of adverse outcome given the known association with poor outcome in medical patients[292] and after cardiac surgery.[293] Both frailty[294] and poor nutritional status[295] are associated with reduced functional capacity which in turn is associated with poor perioperative outcome.[296] The

incorporation of measures of frailty and nutritional status into risk prediction rules for adverse surgical outcome may merit further evaluation.

2.4 Discussion

The purpose of this systematic review was to identify all risk stratification tools which have been validated in heterogeneous patient cohorts, and to report and summarize their discrimination and calibration. A plethora of instruments have been identified, which have been developed and validated in single studies; this therefore unfortunately limits any assessment of their usefulness and generalizability. A smaller number of tools have been multiply validated which could be used universally for perioperative risk prediction; of these, the P-POSSUM and Surgical Risk Scale have been demonstrated to be the most consistently accurate systems.

2.4.1 Risk stratification tools in practice: complexity versus parsimony

There are two key considerations when assessing the clinical utility of the various risk stratification tools reviewed in our study. First, what level of predictive accuracy is fit for the purposes of risk stratification? Second, what is the likelihood that each of the described instruments may be used in everyday practice by clinicians? While the answer to the first question may be to aim as 'high' (accurate) as possible, this must also be balanced against the issues raised by the second question. Risk models incorporating over 30 variables may be highly accurate but are less likely to be routinely incorporated into pre-operative assessment processes than scores of similar performance that use

only a few data points. Furthermore, clinical experience shows that the clinician is less likely to use complex mathematical formulae, as opposed to additive scores, when attempting to risk-stratify patients at the bedside or in the preoperative clinic.[236]

2.4.1.1 P-POSSUM

The P-POSSUM model was developed in the UK, and has since been validated in Japan, Australia and Italy. While this is the most frequently and widely validated model identified by this study, it has some limitations. First, it includes both preoperative and intra-operative variables, and therefore cannot be used for preoperative risk prediction. Second, several of the variables are subjective (e.g. chest radiograph interpretation), carrying the risk of measurement error; furthermore, many of these investigations are no longer routinely implemented as part of the preoperative assessment process. Third, in common with the original POSSUM, the P-POSSUM tends to over-estimate risk in low-risk patients. Fourth, it contains 18 variables, which must be entered into a regression equation to obtain a predicted percentage risk value: clinicians may not wish to use such a complex system. Finally, the inclusion of intra-operative variables, particularly blood loss, which may be influenced by surgical technique, runs the risk of concealing poor surgical performance, therefore jeopardising its face validity as a risk adjustment model for comparative audit of surgeons or institutions.

2.4.1.2 Surgical Risk Scale (SRS)

The SRS consists entirely of variables that are available before surgery, making it a useful tool for preoperative risk stratification for the purposes of clinical decision-

making. However, there are also some limitations. First, it incorporates the ASA-PSS, which may be subject to inter-observer variability and therefore measurement error.[297-299] Second, the surgical severity coding is not intuitive, and some familiarity with the BUPA system would be required for bedside estimation, unless a reference manual was available. Finally, it has only been validated in single centre studies within the UK, therefore its generalizability to patient populations in the US and worldwide is unknown.

2.4.1.3 Other options

The ASA-PSS is widely used as an indicator of whether or not a patient falls into a high, medium or low risk population, but it was not originally intended to be used for the prediction of adverse outcome in individual subjects.[62] It is perhaps surprising that the ASA-PSS was reported as having good discrimination for predicting postoperative mortality, as it is a very simple scoring system, which has been demonstrated to have only moderate to poor inter-rater reliability.[297-300] Nevertheless, the ASA-PSS has face validity as an assessment of functional capacity, which is increasingly thought to be a significant predictor of patient outcome, as demonstrated by more sophisticated techniques such as cardiopulmonary exercise testing.[165] While it is possible that this provides some explanation for the high discriminant accuracy for ASA-PSS found in this systematic review, it is possible that publication bias, favouring studies with 'positive' results, may also be a factor.

The BHOM (Biochemistry Haematology Outcome Model) is a parsimonious version of POSSUM, which omits the subjective variables such as chest radiography and ECG results. It also has the advantage of consisting of variables which are all available preoperatively, with the exception of operative severity. Given the BHOM's similarity in predictive accuracy to P-POSSUM in the one study identified which made a direct comparison,[269] this system warrants further evaluation. Finally, the Identification of Risk In Surgical Patients score (IRIS) was developed in the Netherlands and consists of 4 variables (age, acuity of admission, acuity of surgery and severity of surgery). In the study which developed and validated it on separate cohorts, the validation AUROC was 0.92.[282] Again, further investigation of this simple system would be useful.

2.4.2 Generalizability of findings

2.4.2.1 Clinical and methodological heterogeneity

Clinical heterogeneity (both within- and between-cohort patient heterogeneity) and methodological heterogeneity (between-study differences in the outcome measures used) are both likely to have had a significant influence on some of these findings. For example, between-cohort heterogeneity, and variation in how morbidity is defined (see Appendix 2), may explain the wide range of morbidity rates reported in different studies. Heterogeneity of morbidity definitions may also in part explain the lower accuracy of models for predicting morbidity compared with mortality. On a different note, this study included all populations of patients that were determined to be heterogeneous, using the definitions described in the methods. However, the degree of

heterogeneity varied between studies, including whether or not patients of all surgical urgency categories were included: this may have affected the predictive accuracy of models in different studies.

2.4.2.2 Objective vs. subjective variables and issues surrounding data collection methodology

As previously discussed, the variables included in risk stratification tools may be classified as objective (e.g. biochemistry and haematology assays), subjective (e.g. interpretation of chest radiographs) and patient reported (e.g. smoking history). In some clinical settings, the reliability of non-objective data may be questionable. For example, previous reports have demonstrated significant inter-rater variation in the interpretation of both chest radiographs [301] and ECGs. [302] Patients may also under- or over-estimate various elements of their clinical or social history when questioned in the hospital setting. Despite these concerns, the discrimination of predictors incorporating patient-reported and subjective variables was high in the studies included. This may be due to publication bias; it may also be explained by the fact that in all of these studies, data were collected prospectively by trained staff. Previous work has demonstrated an association between inter-observer variability in the recording of risk and outcome measures, and the level of training that data collection staff have received.[194] These caveats are important when considering the generalizability of this study's findings to the everyday clinical setting, where data-reporting and interpretation may be conducted by different types and grades of clinical staff. Finally, concerns have also been raised over the clinical accuracy of administrative data used for case-mix

adjustment purposes.[221;303] However, one large study included in this review [266] showed high discriminant performance when using ICD-9 and 10 administrative coding data to define the Charlson Index variables.

2.4.3 Limitations of this study

This study is limited by a number of factors.

First, the focus was on studies that measured the discrimination and/or calibration of risk stratification tools in cohorts that were heterogeneous in terms of surgical specialities; therefore, a large number of single-speciality cohort studies identified in the search were excluded from the analysis.

Second, while the inclusion criteria for the review ensured that a standard measure of discrimination was reported (AUROC or c-statistic), many studies did not report measures of calibration. However, in a systematic review such as this, calibration may be seen to be a less important measure of goodness-of-fit than discrimination for a number of reasons. Calibration can only be used as a measure of performance for models that generate an individualized predicted percentage risk of an outcome (e.g. the POSSUM systems) as opposed to summative scores, which use an ordinal scale to indicate increasing risk (e.g. the ASA-PSS). Calibration drift is likely to occur over time and will be affected by changes in healthcare delivery; good calibration in a study over 30 years ago may be unlikely to correspond to good calibration today.[304;305] While such calibration drift may affect the usefulness of a model for predicting an individual patient's risk of outcome, poorly calibrated but highly discriminant models will still be

of value for risk adjustment in comparative audit. Finally, the probability of the Hosmer-Lemeshow statistic being significant (thereby indicating poor calibration) increases with the size of the population being studied.[306] This may explain why many of the large high-quality studies which were evaluated did not report calibration, or reported that calibration was poor.

Third, by using the area under the receiver-operator-characteristic curve as the sole measure of discrimination, a number of studies were excluded, particularly earlier papers which used correlation coefficients between risk scores and post-operative outcomes. This was felt to be necessary, as a uniform outcome measure provides clarity to the reader.

Fourth, publication bias, where studies are preferentially submitted and accepted for publication if the results are 'positive' is likely to be a particular problem in cohort studies.

Finally, despite an extensive literature search, it is possible that some studies which would have been eligible for inclusion may have been missed. Multiple strategies have been used to prevent this; however, in a review of this size, it is possible that a small number of appropriate articles may have been omitted.

2.4.4 Future directions

Undertaking clinical risk prediction should be a key tenet of safe high quality patient care. It facilitates informed consent, and enables the perioperative team to plan their clinical management appropriately. Equally, accurate risk adjustment is required to enable meaningful comparative audit between teams and institutions, to facilitate quality improvement for patients and providers. While dozens of scores and models have been identified, which have been used to predict or adjust for risk, very few of these achieved the aspiration of being derived from entirely preoperative data, and of being accurate, parsimonious and simple to implement. The Surgical Risk Scale is the system that comes closest to achieving these goals; the P-POSSUM score is more accurate but its value is limited by the fact that some of the variables are only available after surgery has been completed. Future work which might be of value would include further comparison of the Surgical Risk Scale, P-POSSUM and objective models such as the BHOM in international multi-centre cohorts and further investigation of models which combine novel variables such as measures of functional capacity, nutritional status and frailty.

There is another possible approach. The American College of Surgeons' National Surgical Quality Improvement Programme (ACS-NSQIP) was created in the 1990s to facilitate risk-adjusted surgical outcomes reporting in Veterans' Affairs hospitals, and now also includes a number of private sector institutions. Risk adjustment models are produced annually, and observed: expected (O:E) ratios of surgical outcomes are reported back to institutions and surgical teams to facilitate quality improvement. A

number of risk calculators have been published by the ACS-NSQIP to help clinicians to provide informed consent and plan perioperative care. However, none of these calculators have been included in this review, as they have all been developed and validated for use in either specific types of surgery (e.g. pancreatectomy,[307] bariatric[308;309]or colorectal[309] surgery) or for specific outcomes (e.g. cardiac morbidity and mortality).[310] A parsimonious, entirely preoperative NSQIP model for predicting mortality in heterogeneous cohorts would be of value in the US; its validation in international multi-centre studies would also be a worthwhile endeavour.

Finally, while there are multiple studies aimed at developing and validating risk stratification tools, how widely such tools are used is not clear. Use of mobile technology, such as apps to enable risk calculation using complex equations at the bedside, might increase the use of accurate risk stratification tools in day-to-day practice. Importantly, in surgical outcomes research, there is an absence of impact studies, measuring the effect of using risk stratification tools on clinician behaviour, patient outcome and resource utilization. Randomized controlled trials to evaluate impact, further validation of existing models across healthcare systems, and establishing the infrastructure required to facilitate such work, including the routine data collection of risk and outcome data, should be of the highest priority in health services research into surgical outcome.[311]

2.5 Conclusions

1. The P-POSSUM predictor and the Surgical Risk Scale (SRS) were demonstrated to be the most consistently accurate tools that have been validated in multiple studies; however, both have limitations.

2. Future work should focus on further evaluation of these and other parsimonious risk predictors, including validation in international cohorts. There is also a need for studies examining the impact that the use of these tools has on clinical decision-making and patient outcome.

Appendix 1: Search strategy

Medline:

risk adjustment.mp. or exp Health Care Reform/ or exp Risk Adjustment/ or exp "Outcome Assessment (Health Care)"/ or exp Models, Statistical/ or exp Risk/ OR exp Risk Assessment/ or risk prediction.mp. or exp Risk/ or exp Risk Factors/ OR predictive value of tests.mp. or exp "Predictive Value of Tests"/ OR exp Prognosis/ or risk stratification.mp. OR case mix adjustment.mp. or exp Risk Adjustment/ OR severity of illness index.mp. or exp "Severity of Illness Index"/ OR scoring system.mp.

combined with:

Surgical Procedures, Operative/ OR surgery.mp. or General Surgery/ OR operation.mp. or exp Postoperative Complications/

combined with:

mortality.mp. or exp Hospital Mortality/ or exp Mortality/ OR morbidity.mp. or exp Morbidity/ OR outcome.mp. or exp Fatal Outcome/ or exp "Outcome Assessment (Health Care)"/ or exp "Outcome and Process Assessment (Health Care)"/ or exp Treatment Outcome/ OR postoperative complications.mp. or exp Postoperative Complications/ OR intraoperative complications.mp. or exp Intraoperative Complications/ OR exp Perioperative Care/ or perioperative complications.mp. OR prognosis.mp. or exp Prognosis/

Embase:

Risk Factor/ or risk adjust$.mp. OR cardiovascular risk/ or high risk patient/ or high risk population/ or risk assessment/ or risk factor OR risk stratification.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR *"Scoring System"/ OR "Severity of Illness Index"/ OR Multivariate Logistic Regression Analysis/ or Logistic Regression Analysis OR logistic models/ or risk assessment/ or risk factors/ OR exp Scoring System OR Prediction/ or possum.mp. or Scoring System/ OR exp Risk Assessment/ or risk stratification.mp. OR predict$.mp. OR exp Quality Indicators, Health Care/ OR Risk Adjustment/

combined with:

exp Surgery/ OR exp Surgical Procedures, Operative/ OR specialties, surgical/ or surgery/ OR surg$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR peri-operative period.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR perioperative.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR postoperative.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR perioperative care/ or intraoperative care/ or postoperative care/ or preoperative care

combined with:

complicat$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR adverse outcome/ or prediction/ or prognosis/ OR exp Postoperative Complication/co, di, ep, su, th [Complication, Diagnosis, Epidemiology, Surgery, Therapy] OR exp Perioperative Complication/ or exp Perioperative Period/ OR exp Mortality/ or exp Surgical Mortality/ OR exp Morbidity/ OR outcome.mp. or "Outcome Assessment (Health Care)"/ or "Outcome and Process Assessment (Health Care)" OR treatment outcome/

Limits:

1980-31 August 2011

Exclusions:

("all infant (birth to 23 months)" or "all child (0 to 18 years)" or "newborn infant (birth to 1 month)" or "infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)" or "adolescent (13 to 18 years)") or (cats or cattle or chick embryo or dogs or goats or guinea pigs or hamsters or horses or mice or rabbits or rats or sheep or swine) or (communication disorders journals or dentistry journals or "history of medicine journals" or "history of medicine journals non index medicus" or "national aeronautics and space administration (nasa) journals" or reproduction journals) or Angioplasty, Balloon/ or Angioplasty, Laser/ or Angioplasty/ or Angioplasty, Balloon, Laser-Assisted/ or Angioplasty, Transluminal, Percutaneous Coronary/ or ANGIOPLASTY.mp. OR Eye/ or Ophthalmology/ or Eye Diseases/ or

OPTHALMOLOGY.mp. or Hearing Loss OR CARDIAC SURGERY.mp. or HEART SURGERY.mp. or Myocardial Revascularization/ or Coronary Artery Bypass/ or CORONARY SURGERY.mp. or Coronary Artery Bypass, Off-Pump/

Hand-searching of reference lists:

The following keywords were searched separately on Medline, Embase and ISI Web of Science:

POSSUM + surgery

NSQIP

E-PASS

ACE-27

APACHE

In addition, the original development studies for all risk prediction models identified in the initial search were then snowballed by hand searching for citations on Medline, Embase and ISI Web of Science

Inclusion / Exclusion criteria

Studies were eligible if they fulfilled the following criteria:

- Studies in adult humans undergoing non-cardiac non-neurological surgery
- Study cohorts which included at least two different surgical sub-specialities

- Studies which described the predictive precision of risk models using analysis of Receiver Operator Characteristic curves

Studies were excluded on the basis of these criteria:

- Cohorts including children (under the age of 14 years)
- Cohorts including patients undergoing cardiac surgery
- Cohorts including patients who did not undergo surgery
- Single speciality cohort studies (e.g. vascular, orthopaedic)
- Studies of ambulatory (day-case) surgery
- Studies describing the development of a risk prediction model without subsequent validation in a separate cohort (either in the original study or subsequent cohorts), with the exception of studies of data from the American College of Surgeons' National Surgical Quality Improvement Programme
- Studies in which the items comprising the risk stratification tool were not disclosed in the study report or available from other sources (such as references)
- Studies using outcomes other than morbidity or mortality as their sole outcome measures (e.g. discharge destination, length of stay)

Studies using only a single pathological outcome measure (e.g. re-operation, cardiac morbidity, infectious complications, renal failure)

Appendix 2: Morbidity Definitions

Author	Model(s) validated	Morbidity definition
Daley, 1997	NSQIP	Cardiovascular: cardiac arrest requiring cardiopulmonary resuscitation, myocardial infarction, pulmonary oedema Respiratory: pneumonia, unplanned intubation, pulmonary embolism, failure to wean from the ventilator 48 hours after operation, Renal: progressive renal insufficiency, renal failure requiring dialysis, urinary tract infection, Neurological: cerebrovascular accident (stroke), coma persisting > 24 hours postoperatively, other neurologic deficits (eg, peripheral neuropathy), Infectious: systemic sepsis. Wound: superficial wound infection; deep wound infection; wound dehiscence Other: prolonged ileus, bleeding requiring > 4 U of transfused blood, graft or prosthesis failure, deep vein thrombophlebitis,
DasGupta, 2009	Detsky Index Edmonton Frail Scale	Cardiac: ischemia, congestive heart failure, new arrhythmia or sudden death: Respiratory: pneumonia, significant bronchospasm, deep venous thrombosis or pulmonary embolism (DVT or PE), or the excessive need for respiratory support Delirium: required the acute onset and fluctuating course of at least one of the following symptoms, as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Revised third edition (DSMIIIR), occurring anytime on or after postoperative day 1: Disorganized thinking or inattention, altered level of consciousness, psychomotor agitation, disorientation or memory impairment, new perceptual disturbances, or new sleep disturbances (e.g., agitation at night or excessive drowsiness during the day). If patients had a known diagnosis of dementia or were on cholinesterase inhibitors, the occurrence of delirium required more than just disorientation or memory impairment.
Davenport, 2006	NSQIP	One or more of 21 specific NSQIP defined complications: not listed
Gawande, 2007	Surgical APGAR score	According to NSQIP's established definitions: Cardiovascular: cardiac arrest requiring CPR, MI,

		Respiratory: Respiratory: pneumonia, unplanned intubation, pulmonary embolism, failure to wean from the ventilator 48 hours after operation Renal: acute renal failure Neurological: coma for 24 hours or longer, stroke, Infectious: septic shock, sepsis, systemic inflammatory response syndrome Wound: wound disruption, deep or organ space surgical site infection Other: bleeding requiring >4 U red cell transfusion within 72 hours after operation, deep venous thrombosis, and vascular graft failure
Goffi, 1999	ASA-PSS APACHE II	**Major** Cardiac failure Abdominal sepsis Haemoperitoneum Respiratory failure Intestinal obstruction Renal failure **Minor** Urinary infection Respiratory infection Wound infection
Haynes, 2011	Surgical APGAR	NSQIP defined (see Gawande 2007)
Jones, 1992	POSSUM APACHE II	Cardiovascular: Myocardial infarct; Cardiac failure; Hypotension (<90mmHg for 2 h); Respiratory failure; Renal: Impaired renal function (urea increase > ***5*** mmol/l from preoperative level) Infection: Chest; Wound; Urinary tract; Deep; Septicaemia; Pyrexia of unknown origin; Other Wound dehiscence: Superficial; Deep; Anastomotic leak Haemorrhage: Wound; Deep; Other Thrombo-embolic: Deep vein thrombosis; Pulmonary embolus; Cerebrovascular accident; Other; Other: Any other complication
Kuzu, 2006	Nutritional Risk Index Maastricht Index Subjective Global	Cardiovascular: Myocardial Infarct; Cardiac failure; Hypotension Respiratory: Atelectasis; Bronchopleural fistula; Chest infection; Empyema; Persistent air leak; Respiratory failure; Pulmonary embolus

	Assessment Mini Nutritional Assessment	GI/liver: Gastrointestinal hemorrhage; Hepatic dysfunction Renal: Impaired renal function; Urinary extravasation/ ureterohydronephrosis; Urinary infection Infectious: Pyrexia of unknown origin; Septicemia and bacteremia; Septic shock Neurological: Cerebrovascular accident Wound: Abscess (intraperitoneal/extraperitoneal); Anastomotic leakage; Deep haemorrhage; Superficial and deep surgical site infection; Wound dehiscence; Wound haemorrhage Thrombosis: Deep venous thrombosis and/or graft thrombosis:
Liebman, 2010	IRIS (Identification of Risk in Surgical Patients)	Cardiovascular: Myocardial infarction Respiratory: Pneumonia GI: Intra-abdominal abscess; Anastomotic leak Renal: urinary tract infection Neurological: CVA Infectious: sepsis Wound: Deep wound infection; Re-bleeding or significant wound haematoma Thrombosis and / or pulmonary emboli Pressure ulcers Other: Miscellaneous; Multiple Organ Failure
Makary, 2010	ASA, Lee RCRI and Eagle Scores alone combined with Frailty Index	NSQIP defined
Pillai, 1999	Otago Surgical Audit Score	Complications classed according to severity: 0: no complication; technical complications (some): e.g. anaesthetic compications; Non-operative complications: e.g. no lesion found, PUO 1: Minor: Patient discomfort e.g. postoperative atelectasis; urinary retention 2: Intermediate: significant compromise: e.g. prolonged ileus; DVT 3: Severe: Major threat to life: e.g. DIC; MI; renal failure
Story, 2009	Perioperative mortality risk score	Unplanned ICU admission: decision made to admit to ICU, Coronary Care unit or High dependency unit made during or after surgery Systemic inflammation: New finding of at least two of the following: Temp>38.3 or <36; WCC>12,000c/ml; RR>20 breaths/min; HR >90bpm; or a positive blood culture alone Acute renal impairment: Creatinine increase >20% pre-op value or admission to ICU for RRT

Appendix 3: Risk stratification tools validated in single studies

Author	Model	Outcome	No of variables	Age	Sex	Race	Smoking	Surgery type	Surgery urgency	ASA	Haem	Biochem	IHD or arrhythmia	CCF	COPD	Neuro	Renal	Diabetes	Cancer	Other pre-op factors	Intra-op factors	Post-op factors
Beattie, 2009[312]	New Anaemia Model	90d mortality	10	X							Hb[&]			X			RD			Periop beta blockers, CCBs ACEI Inpatient stay>5d pre-operatively		Any post-op NSAIDs RBC transfusion at any stage
DasGupta, 2009[276]	Detsky	Morbidity to hospital discharge	9	X				X					X	X						General poor functional status		
DasGupta, 2009[276]	Edmon-ton Frail Scale	Morbidity to hospital discharge																				
Hadjuanastassiou, 2004[278]	Surgical Mortality Score	Mortality to hospital discharge	6	X	X			X	X											Onset time of surgery Duration of surgery		
Haga, 2011[313]	E-PASS	Mortality to hospital discharge and 30d	10	X						X			X*	X*	X			X		Body weight Performanc e status	Blood loss Duration of surgery Incision type	
Haga,	mE-PASS	Mortality	7	X				X		X			X*	X*	X			X		Performanc		

2011[313]		to hospital discharge and 30d																		e status		
Kuzu, 2006[281]	Nutrit-ional Risk Index	Mortality and morbidity at 30d or Hospital discharge	3								Alb									Normal weight Current weight		
Kuzu, 2006[281]	Mini Nutri-tional Assess-ment	Mortality and morbidity at 30d or Hospital discharge	18																	Height Weight BMI Nutritional history Subjective assessments of general well-being and comorbidities		
Kuzu, 2006[281]	Maas-tricht Index	Mortality and morbidity at 30d or Hospital discharge	4								Lym pho cytes	Alb Pre-albumin								Ideal weight		
Liebman, 2010[282]	IRIS	Mortality & Morbidity to hospital discharge	4	X				X	X											Hospital admission status (acute vs non-acute)		

Makary, 2010[275]	Eagle score	Morbidity to hospital discharge	5	X									X[$]	X				DM				
Makary, 2010[275]	Frailty index	Morbidity to hospital discharge	5																	Shrinking Decreased Grip strength Exhaustion Low physical activity Slow walking speed		
Nathanson, 2009[283]	MPM0-III	Mortality to hospital discharge																				
Neary, 2007[269]	BHOM	30d and 1y mortality	8	X	X			X			Hb WC C	Ur Na K										
Neary, 2007[269]	RCRI	30d and 1y mortality	6					X					X	X		CVAR	RD	ID				
Osler, 1998[263]	ICISS	Mortality to hospital discharge																		PRODUCT OF SURVIVAL RISK RATIOS OF ALL ICD-9 CLASSIFICATION		

																				CODES		
Pillai, 1999[284]	Otago	Morbidity to hospital discharge	12	X	X	X		X	X											Admission type Number of operations Pre-operative LOS Day case vs In-patient surgery	Duration of surgery Operator grade Wound category	
Stachon, 2008[285]	SAPS II	Mortality to hospital discharge	15																			
Stachon, 2008[285]	APACHEN	Mortality to hospital discharge	15																	Nucleated red cell assay added to APACHE II score as an independent variable		
Stachon, 2008[285]	SAPSN	Mortality to hospital discharge	16																	Nucleated red cell assay added to SAPS II score as an independent variable		
Stachon, 2008[286]	DELAWARE	Mortality to	9	X							Plts WC	ALT CK										

		hospital discharge									C	Chol K TGC CRP										
Story, 2009[287]	Peri-operative risk score	30d Mortality	6	X						X		Alb										Acute renal impairment Unplanned ICU admission Inflammation

* : Cardiac co-morbidity classed as single variable
+: Described as Pulmonary co-morbidities
++: Described as CNS comorbidities
&: WHO classification of anaemia
** Davenport: Cardiac disease defined as previous PTCA or previous cardiac operation
*** Hall: Angina, MI, previous PTCA and previous CABG all classed separately in model
$: Eagle criteria: score separately for history of angina vs history of MI

AC: Any cancer; Alb: Serum Albumin; ALP: Serum Alkaline Phosphatase; ARF: Acute renal failure; Bili: Serum Bilirubin; BUN: Blood Urea Nitrogen; Chemo: Chemotherapy; Co: Coma ; CNST: CNST tumour; Creat: Serum creatinine; CVA: Cerebrovascular accident without residual deficit; CVAD: Cerebrovascular accident with residual deficit; DC: Disseminated cancer; Dia: Dialysis dependent renal failure; DM: Any definition of Diabetes Mellitus; Hb: Haemoglobin; HP: Hemiplegia; ID: Insulin dependent; IS: Impaired sensorium; NID: Non-insulin dependent diabetes; Plt: Platelet count; RD: Other definition of renal dysfunction; RT: Radiotherapy; SGOT: Serum glutamic oxaloacetic transaminase
Ur: Serum urea; WCC: White cell count; Work RVU: Work relative value units

Chapter Three: Validation of six perioperative risk stratification tools in a UK surgical population

3.1 Introduction

3.1.1 Background

In Chapter 2, a qualitative systematic review found that the Portsmouth Physiological and Operative Severity Score for the enumeration of Mortality and morbidity (P-POSSUM) and Surgical Risk Scale (SRS) were the two most consistently validated risk stratification tools in heterogeneous populations of patients undergoing non-cardiac, non-neurological surgery. In addition, the ASA-PSS had been evaluated in several studies.

It was also observed that there is a lack of risk prediction tools which use solely preoperative data and which are multiply validated and accurate.

3.1.2 Aims

1. To compare the discrimination and calibration of P-POSSUM and SRS and also the original POSSUM and American Society of Anesthesiologists' Physical Status Score (ASA-PSS), for a variety of surgical outcomes in a sample of patients who underwent major non-cardiac surgery and were recruited into studies by UCL/UCLH Surgical Outcomes Research Centre (SOuRCe).
2. To test the predictive accuracy of the POSSUM physiology score, which in its entirety is available preoperatively, and to compare its performance in this cohort with that of existing validated risk prediction tools.

3.1.3 Objectives

1. To describe the general characteristics of the study population used throughout the remainder of this thesis
2. To calculate the population and sub-population mortalities according to risk categories derived from the various risk stratification tools
3. To determine and compare the predictive accuracy (discrimination) and calibration for each risk stratification tool for a variety of postoperative outcomes
4. To discuss the strengths and weaknesses of the different risk stratification tools for this population and to make recommendations for use of these rules in clinical practice.

3.2 Methods

3.2.1 Background and ethical considerations

This is a prospectively conducted observational cohort study with long-term follow up of vital status (alive or dead). It studies two separate cohorts of patients who underwent elective major non-cardiac, non-neurological surgery at the Middlesex Hospital, London UK, between 2001 and 2005.

Cohort One was recruited into a prospective, observational cohort study conducted within the UCL/UCLH Surgical Outcomes Research Centre, (SOuRCe), which validated the Post Operative Morbidity Survey[183], for the first time in the UK.[56] This study underwent full ethics review and received approval from the Joint UCLH/UCL Committee on the Ethics of Human Research (reference number 01/0116). All patients who were over the age of 18 years and admitted for major elective surgery between July 2001 and September 2003 were eligible for inclusion. All individuals fulfilling these criteria were approached for consent and consenting patients were enrolled into the study.

Cohort Two was monitored as part of a service evaluation of the Departments of Surgery and Anaesthesia at the Middlesex Hospital between March 2004 and April 2005. This service evaluation was also conducted by SOuRCe. This was approved by the local Research Ethics Committee and patient consent was deemed not to be required. This study terminated when SOuRCe at the Middlesex Hospital closed in April 2005.

Approval for disclosure of data held on the Central Register for flagging of patients to provide fact of death or exit from the NHS was granted by the Ethics and Confidentiality Committee of the National Information Governance Board in June 2009.

3.2.2 Inclusion criteria

Cohort One

Major elective surgery was defined as procedures expected to last more than two hours or with an anticipated blood loss greater than 500 ml. Patients in whom the planned surgery included any of the following procedures were eligible for recruitment:

- Orthopaedic surgery: revision hip arthroplasty, total hip replacement, total knee replacement, fusion/instrumentation of multiple lumbar or thoracic vertebrae;
- General abdominal surgery: laparotomy including partial hepatectomy, pancreatic surgery, re-operative colon surgery, abdomino-perineal resections, anterior resections, pan-proctocolectomies, hepatobiliary bypass procedures;
- Urological surgery: radical prostatectomy, radical cystectomy, radical nephrectomy.

Cohort Two

In the second study, the definition of major elective surgery was broadened to include patients undergoing the following vascular surgical procedures: aortic aneurysm repair; carotid endarterectomy; arterial reconstruction; and amputation for vascular disease.

3.2.3 Dataset

3.2.3.1 Perioperative data (SOuRCe dataset)

Irrespective of cohort, the following data were collected on all patients: patient demographics (age, date of birth); measures of perioperative risk (ASA-PSS score and the elements of POSSUM scores); name of planned and actual operation; date of surgery; postoperative destination; length of stay on the critical care unit and in the hospital; vital status at hospital discharge; and post-hospital discharge destination. The Post Operative Morbidity Survey (POMS) was administered on days 3, 5, 8 and 15 after surgery. All data were collected prospectively by one of two study nurses who had received specific training in how to identify and interpret both all of the preoperative risk and postoperative outcome variables within the dataset. Required data were collected through review of clinical notes and charts, direct patient questioning and examination, retrieval of data from the hospital information technology system and/or consultation with the patients' clinical teams.

3.2.3.2 Post-discharge follow up

Data linkage between the SOuRCe dataset and the Medical Research Information Service (MRIS) at the NHS Information Centre provided mortality data for the cohort. Patient records were linked using patient name, date of birth, postcode (at time of surgery) and NHS number. All patients who could be traced and had not exited the NHS were 'flagged' for mortality status. The MRIS provided updated records on the cohort every three months, commencing in January 2010. The cohort was right-censored on 1 March 2012.

3.2.4 Risk stratification tools

POSSUM mortality and morbidity models, P-POSSUM, SRS and ASA-PSS were all evaluated. In addition, the preoperative POSSUM (POSSUM physiology) score and the additive POSSUM (calculated as the sum of the POSSUM score, without entering the POSSUM variables into regression equations) were evaluated.

3.2.5 Analysis

3.2.5.1 General description of study population

Comparisons were made between Cohorts One and Two, and between the four different surgical specialities for the whole cohort.

For between-cohort analyses, group means were compared using two sample t-tests. Pearson's Chi-squared test was used for comparing group proportions, except for the comparison of ethnicity for which the Fisher's Exact Test was used (because of small numbers in the 'non-white' categories).

For between-speciality analyses, group proportions were compared using Chi-squared tests (three degrees of freedom); Bonferroni's correction was used when undertaking multiple paired comparisons using Chi squared testing to isolate which group(s) differed from the other(s). Between-speciality comparison of means was undertaken using one-way Analysis of Variance (ANOVA) with Bonferroni's correction for multiple testing.

All p-values were two-sided and $p<0.05$ was considered to be statistically significant. Where Bonferroni corrected p values are quoted, these are denoted p'.

3.2.5.2 Risk stratification tools: general description

The distribution of scores in the study population was calculated for each of the risk stratification tools being evaluated. Outcomes for patients in the overall cohort were described in terms of four 'risk categories' for each score. For the ASA-PSS, risk categories were defined according to the established classification system (I-IV). For the POSSUM, P-POSSUM, additive POSSUM and POSSUM physiology predictors, four categories of risk were defined on the basis of dividing the study population into evenly distributed quartiles. These groups were labelled Category One to Four with Category One being the lowest risk and Category Four being the highest risk. For the SRS, it was not possible to divide the population into evenly distributed quartiles; therefore, the categories were determined by ensuring as near to equal spread as possible between three groups, and one further smaller group at the highest end of predicted risk. As with the various POSSUM risk stratification tools, these groups were labelled Category One to Four with Category One being the lowest risk, and Category Four being the highest risk. In addition, the population mortalities for the highest risk 10% of patients according to POSSUM (total score and physiology score) were separately calculated.

3.2.5.2 Risk stratification tools: discrimination

The primary measure of discrimination was the area under the Receiver-Operator-Characteristic (ROC) curve (AUROC). ROC curves were plotted for all models for each of

the outcomes being studied, and the area under the ROC curve for each risk stratification tool was calculated using a trapezoidal method. Standard error and 95% confidence intervals were also calculated.

The accuracy of the POSSUM (mortality equation), P-POSSUM, SRS, ASA-PSS, POSSUM Physiology score and additive POSSUM scores were each assessed for prediction of inpatient, thirty day and one-year mortality. In addition, the accuracy of the POSSUM (morbidity equation), POSSUM Physiology, additive POSSUM, SRS and ASA-PSS were assessed for prediction of postoperative morbidity (defined by at least one positive Post Operative Morbidity Survey (POMS) at any stage postoperatively).

Next, the AUROC for each of the risk stratification tools was compared with the 'gold standard' for each outcome, which was defined as the tool with the highest AUROC on the previous analysis. This comparison was conducted using Chi squared tests, and both raw and adjusted p values (using Bonferroni's correction, denoted p') were calculated and reported.

Predictive precision was classed as poor if AUROC <0.7; moderate if 0.70 – 0.9 and high if ≥0.90.[249] Again, p values were considered significant if <0.05.

3.2.5.3 Risk stratification tools: Calibration

Calibration was measured by calculation of the Pearson Chi squared statistic for the POSSUM and P-POSSUM models. On Chi squared testing, if $p<0.05$, then this was considered significant and the risk stratification tool was interpreted to be poorly calibrated for that outcome. Calibration was not measured for the other risk

stratification tools, as they are scoring systems rather than regression models producing estimates of percentage risk.

Stata InterCooled (Release 12.1) software (StataCorp LP, College Station, Texas, USA) was used for all analyses

3.3 Results

3.3.1 Characteristics of the study population

In total, data were analysed for 1362 patients.

450 patients were enrolled into Cohort One. Of these, nine had their surgery cancelled, and three were withdrawn from the study (one withdrew their consent, one was found to be participating in an interventional study, and one was withdrawn by their attending consultant). Therefore data were analysed for 438 patients.

Data were collected on 996 patients in Cohort Two. Of these, 72 patients were excluded, as follows: no surgery (n=7); patient previously enrolled in Cohort One (n=21); duplicate patient (n=29); no trace records available for long-term follow up (n=13) and lost to follow up (n=2). Therefore data were analysed for 924 patients.

The cohorts were similar in terms of age, gender and surgical speciality, except that Cohort One did not include any vascular surgical patients (see Table 13).

3.3.1.2 Estimates of perioperative risk according to different risk stratification tools

There were small, but statistically significant, differences in the POSSUM and p-POSSUM predicted mortality, POSSUM predicted morbidity and distribution of Surgical Risk Scale categories between the two cohorts, with Cohort One predicted to be higher risk. Overall, predicted mortality and morbidity was high: mean POSSUM predicted morbidity was 28.04% (standard deviation, s.d. =19.07) and mortality 6.33% (s.d. = 7.81) was for the entire cohort of 1362 patients. P-POSSUM predicted mortality was somewhat lower (mean =1.93%; s.d. =3.81). There was no difference in ASA-PSS scores between the two cohorts.

		Overall population	Cohort 1	Cohort 2	p value
Number		1362	438	924	
Mean age (s.d.)		63.53 (15.26)	62.95 (15.65)	63.80 (15.08)	NS
Female (%)		773 (56.75)	259 (59.13)	514 (55.63)	NS
Ethnicity: n (%) *					
Black		69 (5.1)	20 (4.6)	49 (5.3)	NS
White		1198 (88.0)	404 (92.2)	794 (85.9)	p<0.001
Asian		44 (3.2)	7 (1.6)	37 (4.0)	p<0.03
Other		49 (3.6)	5 (1.1)	44 (4.8)	p<0.001
Missing		2 (0.2)	2 (0.5)	0	p<0.0001
POSSUM predicted mortality risk (%) :					
Mean (s.d.)		6.3 (7.8)	7.9 (10.3)	5.6 (6.2)	p<0.0001
Median (IQR)		3.7 (2.6-6.5)	4.2 (2.9- 8.3)	3.47 (2.5-6.0)	
p-POSSUM predicted mortality risk (%):					
Mean (s.d.)		1.93 (3.8)	2.5 (5.4)	1.7 (2.8)	p<0.001
Median (IQR)		1.0 (0.6-1.8)	1.1 (0.7- 2.2)	0.82 (0.6-1.6)	
POSSUM predicted morbidity risk (%):					
Mean (s.d.)		28.0 (19.1)	31.9 (21.3)	26.2 (17.7)	p<0.0001
Median (IQR)		21.1 (14.6-34.8)	23.9 (16.3-42.3)	19.7 (14.2-32.7)	
Speciality: n (%)					
	Orthopaedic	855 (62.8)	288 (65.8)	567 (61.4)	NS
	General	296 (21.7)	101 (23.1)	195 (21.1)	NS
	Urology	147 (10.8)	49 (11.2)	98 (10.6)	NS
	Vascular	64 (4.7)	0 (0)	64 (6.9)	p<0.0001
ASA-PSS Score: n (%)					
	I	223 (16.4)	79 (18.0)	144 (15.6)	NS
	II	808 (59.3)	253 (57.8)	555 (60.1)	
	III	299 (22.0)	99 (22.6)	200 (21.7)	
	IV	13 (1.0)	2 (0.5)	11 (1.2)	
	Missing	19 (1.4)	5 (1.1)	14 (1.5)	
Surgical Risk Scale Category: n (%)					
	One (score 3-5)	302 (22.5)	81 (18.7)	221 (24.3)	p<0.005
	Two (score 6)	669 (49.8)	213 (49.2)	456 (50.1)	
	Three (score7-8)	316 (23.5)	125 (28.9)	191 (21.0)	
	Four (score ≥9)	56 (4.2)	14 (3.2)	42 (4.6)	

Table 13: Baseline patient characteristics

3.3.2 Comparison of surgical specialities

Across the overall population of 1362 patients, 855 (62.8%) patients underwent orthopaedic surgery, 296 (21.7%) general surgery, 147 (10.8%) urological surgery and 64 (4.7%) vascular surgery. Comparisons of baseline characteristics according to surgical speciality are reported below, and in Table 14.

Comparison of means between speciality groups was undertaken using ANOVA. When interpreting these results, it should be noted that Bartlett's test of equal variances was statistically significant in all analyses. This is likely to be because both the variances and the group sizes were substantially different.

There was a difference in gender distribution between surgical specialities ($p<0.001$). On ANOVA, the F test was positive for all variables where this analysis was undertaken: age ($F [3, 1358] = 43.84$; $P<0.0001$), POSSUM predicted morbidity ($F [3, 1358] = 78.13$; $p<0.0001$), POSSUM predicted mortality ($F [3, 1358] = 55.60$; $p<0.0001$), p-POSSUM predicted mortality ($F [3, 1358] = 25.95$; $p<0.0001$) and postoperative length of stay ($F [3, 1358] = 15.70$; $p<0.0001$. Therefore, the means of these variables were compared between surgical speciality groups. Including adjustment with Bonferroni's correction for multiple analyses the following comparisons were statistically significant:

- Age: all comparisons ($p'<0.01$) except for vascular versus orthopaedic surgery (NS);
- POSSUM predicted morbidity: orthopaedics compared with each other speciality individually ($p'<0.001$); all other comparisons NS

- POSSUM predicted mortality risk: orthopaedics versus general; orthopaedics versus urology (both p'<0.001); orthopaedics versus vascular (p'<0.01); all other comparisons NS
- p -POSSUM predicted mortality: orthopaedics versus general; orthopaedics versus urology (both p'<0.001); orthopaedics versus vascular p'<0.01; all other comparisons NS
- Postoperative length of stay: orthopaedics versus general (p'<0.001); all other comparisons: NS

Table 14: Comparison of baseline characteristics for the overall population according to surgical speciality

	Overall population (n=1362)	Orthopaedic (n=855)	General (n=296)	Urology (n=147)	Vascular (n-64)
Age [years]: mean(sd)	63.52 (15.26)	65.92 (14.75)	60.28 (14.52)	52.93 (15.40)	70.89 (9.95)
Female (%)	56.75	61.75	55.41	46.26	20.31
POSSUM predicted mortality (%)	6.3 (7.8)	4.4 (4.3)	10.1 (11.1)	9.6 (11.5)	7.5 (6.7)
p-POSSUM predicted mortality (%)	1.9 (3.8)	1.3 (2.4)	3.1 (5.1)	3.2 (6.3)	2.5 (3.1)
POSSUM predicted morbidity (%)	28.0 (19.1)	22.5 (13.1)	38.8 (23.5)	36.2 (24.5)	33.9 (19.9)
Postoperative length of stay (days): mean (s.d.)	13.0 (15.1)	11.0 (12.2)	17.6 (18.4)	14.4 (19.7)	16.1 (16.8)
Postoperative length of stay (days): median	9	8	13	9	9

3.3.3 Overall population mortality according to risk prediction categories

3.3.3.1. ASA-PSS grade

The distribution of ASA-PSS scores in the overall population is presented in Figure 5 and population mortality by ASA-PSS score is summarised in Table 15. ASA-PSS scores were missing for 19 patients. There were more inpatient deaths in the ASA-PSS Class III group when paired comparisons were made with ASA-PSS I and II patients (ASA-PSS I versus III p'<0.05; ASA-PSS II versus III p'<0.001). However, here was no difference in 30-day mortality between ASA-PSS grades.

One-year mortality was similar in ASA-PSS I and II patients and in ASA-PSS III and IV patients. One-year mortality was significantly lower in ASA-PSS I patients when compared with ASA-PSS III (p'<0.001) and with IV (p'<0.001). ASA-PSS Class II one-year mortality was also lower than Class III (p'<0.001) and IV (p'<0.05).

Five-year mortality remained similar in ASA-PSS III and ASA-PSS IV patients. Five-year mortality for patients graded as ASA-PSS I was lower than that for patients categorised as ASA-PSS II (p'<0.05), ASA-PSS III (p'<0.001) and ASA-PSS IV (p'<0.001).

Mortality in ASA-PSS II patients was lower at five years than ASA-PSS III (p'<0.001) and ASA-PSS IV patients (p'<0.05).

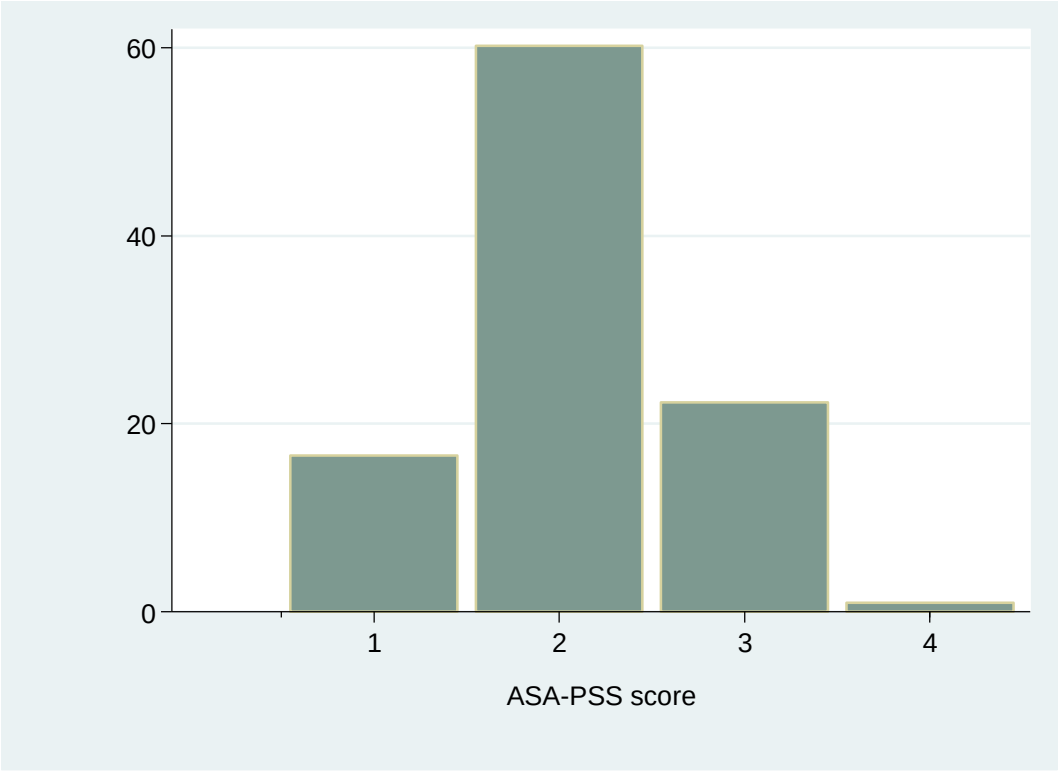

Figure 5: Distribution of ASA-PSS scores in overall population (n=1343)

	Hospital n=1343*	30-day n=1343*	One-year n=1329*	Five-year n=1321*
All patients	1.5	1.1	6.8	21.0
I	0.5	0.9	2.3	8.6
II	0.7	0.7	5.3	17.8
III	4.7	2.3	14.2	37.8
IV	0	0	23.1	46.2

**Note smaller sample size, as ASA-PSS scores were missing for 19 patients*

Table 15: Percentage mortality of overall population at four endpoints according to ASA-PSS category

3.3.3.2 Surgical Risk Scale

The distribution of the Surgical Risk Scale (SRS) in the overall population is presented in Figure 6. Figure 7 presents the distribution of the SRS into four categories of increasing risk as follows: Category One (SRS≤5; n=302), Category Two (SRS=6; n=669), Category Three (SRS=7; n=316) and Category Four (SRS ≥8; n=75). Mortality according to SRS categories is summarised in Table 16. Hospital mortality was lower in Category One patients than in Category Three (p'<0.05) and Category Four patients (p'<0.001). Similarly, Category Two patients had a lower hospital mortality than Category Three (p'<0.005) and Category Four (p'<0.001) patients. Similar patterns were seen with one-year mortality: Category One versus Three (p'<0.001); Category One versus Four (p'<0.001); Category Two versus Three (p'<0.001) and Four (p'<0.001).

At five years, observed mortality in Category Three and Four patients was similar; however, with this one exception, there were significant differences in mortality between all categories (Category One versus Two: p'<0.02; all other comparisons p'<0.001).

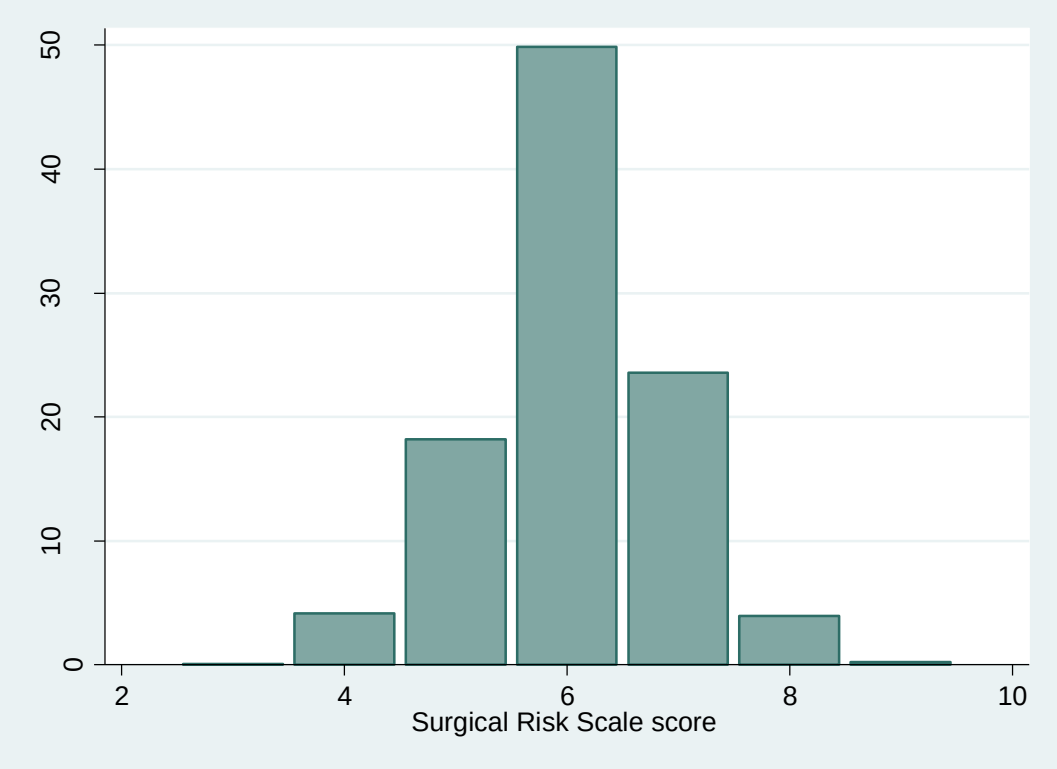

Figure 6: Distribution of SRS scores in overall population (n=1343)

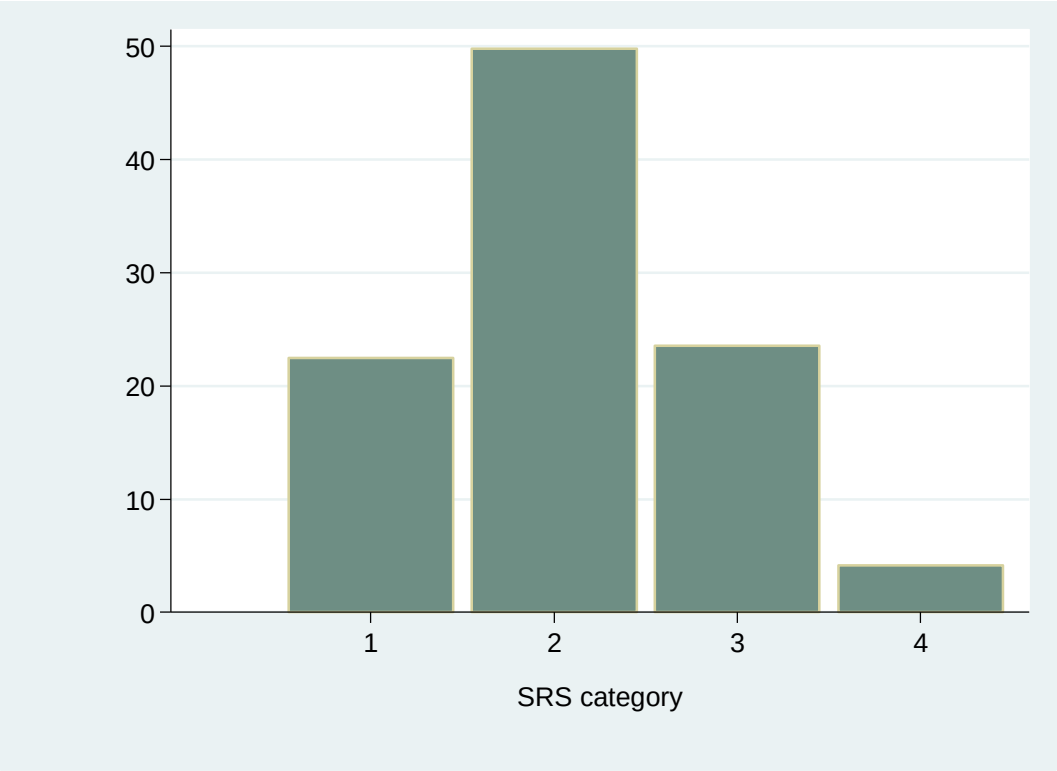

Figure 7: Distribution of SRS Categories (n=1343)

SRS score (Category)	Hospital n=1343	30-day n=1343	One-year n=1329	Five-year n=1321
All patients (n=1343)*	1.5	1.1	6.8	21.0
3-5 (Category One)	0.3	0.7	2.3	9.4
6 (Category Two)	0.6	0.6	4.2	16.6
7 (Category Three)	3.5	2.2	13.8	35.5
≥8 (Category Four)	8.8	3.5	25.0	53.6

*SRS scores missing for 19 patients

Table 16: Percentage population mortality by SRS category

3.3.3.3 Additive POSSUM score

The distribution of the additive POSSUM score is presented in Figure 8.

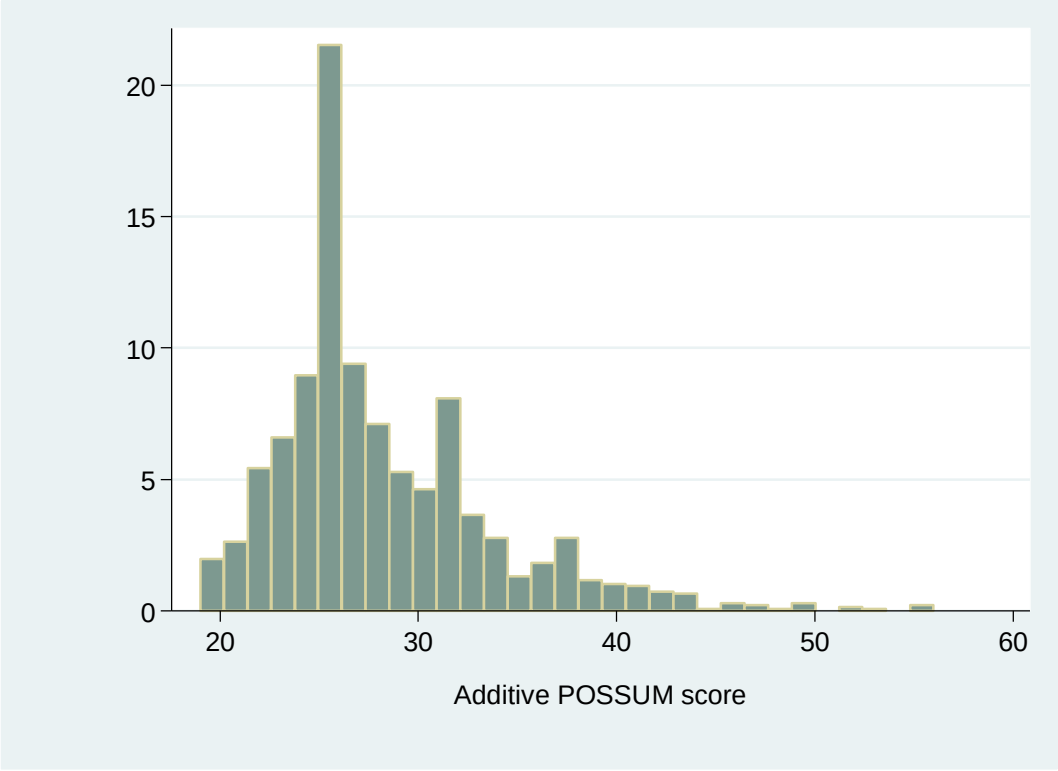

Figure 8: Distribution of Additive POSSUM scores (n=1362)

Patients were subdivided into quartiles of predicted risk of mortality according to the additive POSSUM score as follows: Category One (score ≤ 24, n=349); Category Two (score 25-27, n=421); Category Three (score 27-31, n=294); and Category Four (score ≥32, n = 298). Mortality according to these categories of predicted risk is summarised in Table 17.

Additive POSSUM score (risk category)	Inpatient n=1362; (%)	30 day n=1362; (%)	1 year n=1347; (%)	5 year n=1339; (%)
All patients	1.5	1.1	6.8	21.0
≤24 (Category 1)	0.00	0.0	0.86	5.78
25-27 (Category 2)	0.24	0.71	2.89	14.49
28-31 (Category 3)	2.04	1.70	6.48	24.48
≥32 (Category 4)	4.70	2.35	20.00	43.60

Table 17: Percentage population mortality by Additive POSSUM risk category

Thirty-day mortality was lower in Category One patients than Category Three (p'<0.03); all other comparisons were non-significant. Hospital mortality was lower in Category One versus Category Three (p'<0.05) and Category Four (p'<0.001) patients and Category Two versus Category Four patients (p'<0.001).

At one year, there was no difference in mortality between Category One and Two patients or between Category Three and Four patients; however there were significant differences between all other risk categories (all comparisons p'<0.001). At five years, mortality was significantly different for all category comparisons (p'<0.001 for all comparisons with the exception of Category Two versus Category Three (p'<0.005).

The observed mortality of the estimated highest risk 10% of this cohort (Additive POSSUM ≥36; n=144) was as follows: 6.3% before hospital discharge; 3.5% at thirty days; 16.0% at six months and 25.7% at one year. By five years, 52.8% of this highest risk category had died.

3.3.3.4 POSSUM and P-POSSUM

The distribution of POSSUM and P-POSSUM predicted mortality risks are shown in Figures 9 and 10 respectively. The predicted and observed (at each of the four endpoints) mortality by the four quartile categories of P-POSSUM predicted mortality are shown in Table 18.

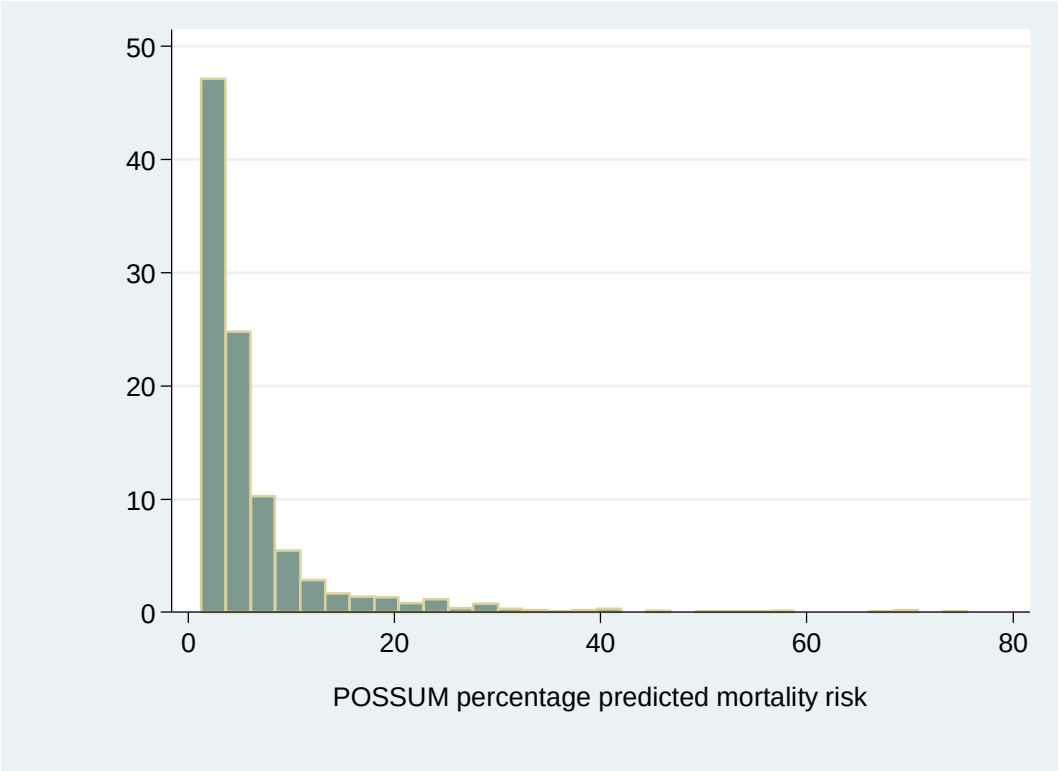

Figure 9: Distribution of POSSUM predicted mortality risk (n=1362)

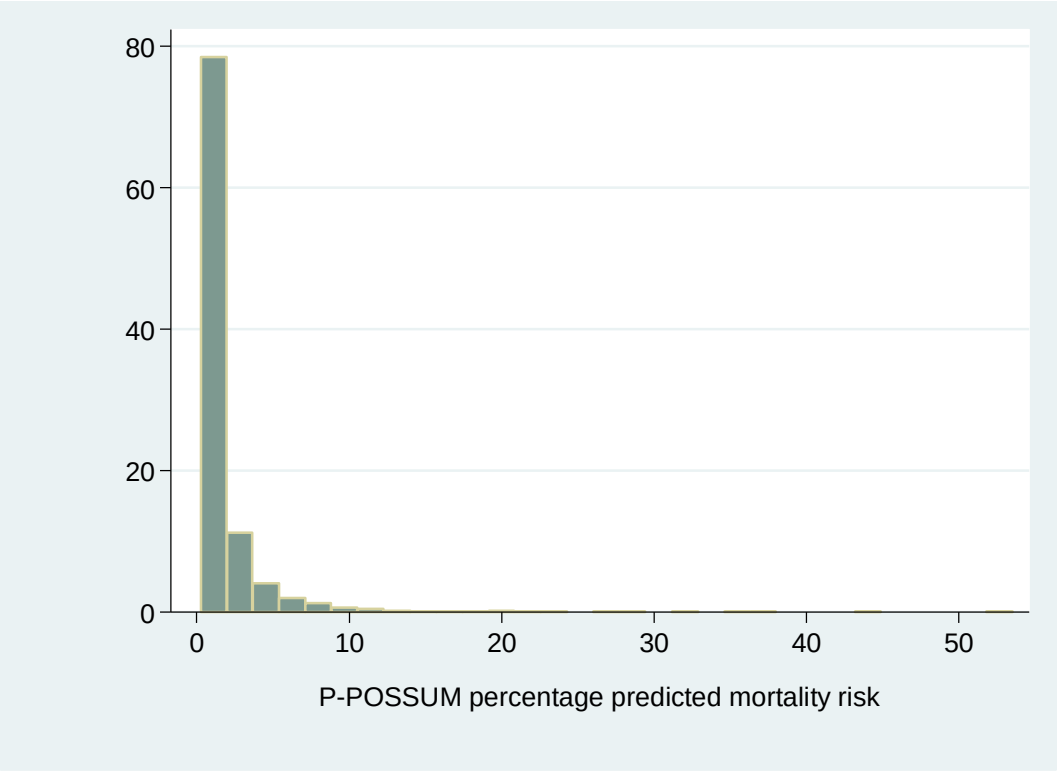

Figure 10: Distribution of P-POSSUM predicted mortality risk (n=1362)

Risk category	Mean percentage P-POSSUM predicted mortality (standard error)	95% C.I.	Inpatient n=1362; (%)	30 day n=1362; (%)	1 year n=1347; (%)	5 year n=1339; (%)
All patients	1.9 (0.10)	1.7-2.1	1.5	1.1	6.8	21.0
Category One	0.5 (0.01)	0.47-0.49	0.00	0.00	0.9	5.4
Category Two	0.8 (0.00)	0.76-0.78	0.3	0.9	3.2	14.5
Category Three	1.2 (0.01)	1.21-1.26	0.9	0.9	4.1	20.0
Category Four	5.2 (0.4)	4.54-5.94	5.0	2.6	19.2	43.0

Table 18: Predicted versus observed mortality for categories of risk according to P-POSSUM prediction at four end-points

3.3.3.5 POSSUM Physiology score

The distributions of POSSUM physiology score are shown in Figure 11. Again, four risk categories were defined based on quartiles of the study population: Category One (score ≤14; n=334), Category Two (score 15-16; n=364), Category Three (score 17-18; n=337) and Category Four (score ≥19; n=327). Population mortalities for each of these categories at the four measured end-points are detailed in Table 19. The patients who were classified as being in the highest risk 10% according to POSSUM physiology score (score ≥21; n=167) had a mortality of 5.4% while in hospital, 3.0% at thirty days, 10.2% at 6 months, 17.4% at one year and 40.7% at 5 years.

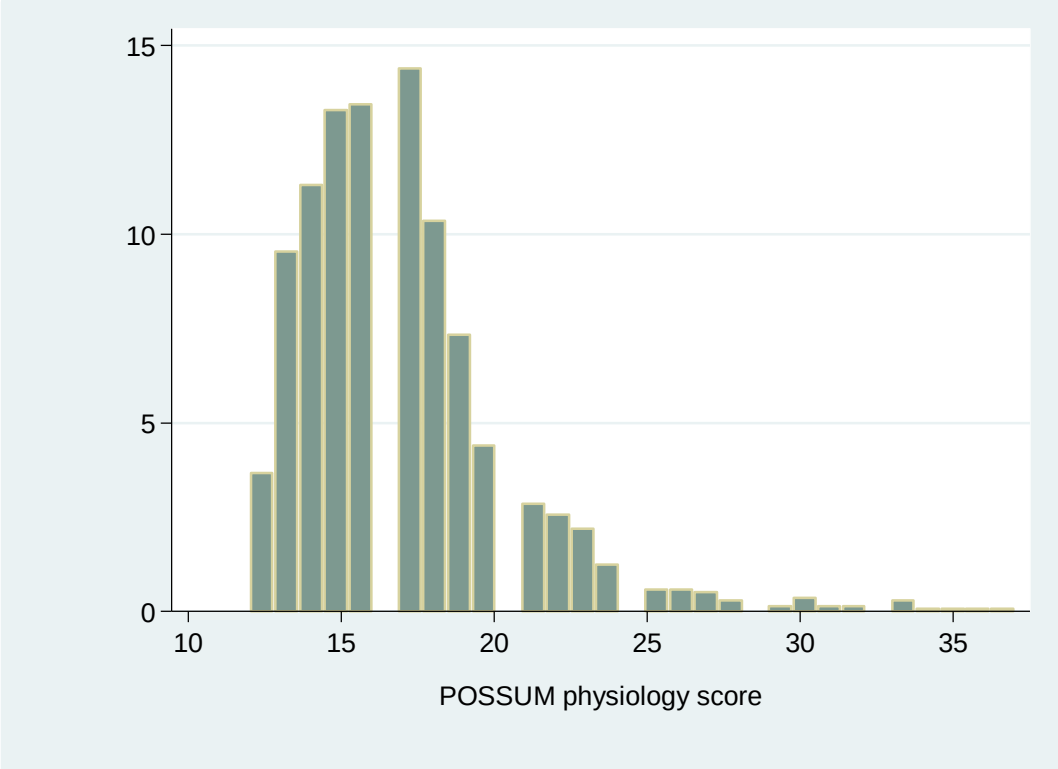

Figure 11: Distribution of POSSUM physiology scores

POSSUM physiology score (risk category)	Inpatient n=1362; (%)	30 day n=1362; (%)	1 year n=1347; (%)	5 year n=1339; (%)
All patients	1.5	1.1	6.8	21.0
≤14 (Category 1)	0.00	0.00	3.9	13.0
15-16 (Category 2)	1.1	0.6	4.4	15.1
17-18 (Category 3)	1.2	1.5	5.2	20.6
≥19 (Category 4)	4.0	2.5	14.3	35.0
p value	**p<0.001**	**p=0.014**	**p<0.001**	**p<0.001**

Table 19: Percentage population mortality by POSSUM physiology category

3.3.4 Accuracy of risk stratification tools for predicting postoperative outcomes

3.3.4.1 Inpatient mortality

Data were analysed for 1343 patients: there were 1362 in the overall cohort, and ASA-PSS data were missing for 19 patients. Hospital mortality was 1.54% (21 deaths). ROC curve analyses are presented in Table 20 and Figure 12. All models demonstrated moderate discrimination for the prediction of hospital mortality. Only the ASA-PSS was significantly less accurate than the gold standard (P-POSSUM) (see table 21). Neither POSSUM model was well calibrated for this population; on examination of the observed: expected (O: E) ratios for deciles of risk, both POSSUM mortality models significantly over-predicted hospital death across the range of risks.

Risk stratification tool	AUROC	Standard Error	95% C.I.	Pearson Chi^2	p value (for probability >Chi^2)
P-POSSUM	0.85	0.03	0.78- 0.91	498.87	**<0.0001**
POSSUM mortality	0.84	0.04	0.77- 0.91	421.74	**<0.0001**
Additive POSSUM	0.84	0.04	0.78-0.91		
POSSUM physiology	0.77	0.05	0.67- 0.87		
SRS	0.77	0.05	0.68- 0.87		
ASA-PSS	0.73	0.05	0.62-0.83		

Table 20: Discrimination and calibration of risk stratification tools for inpatient mortality

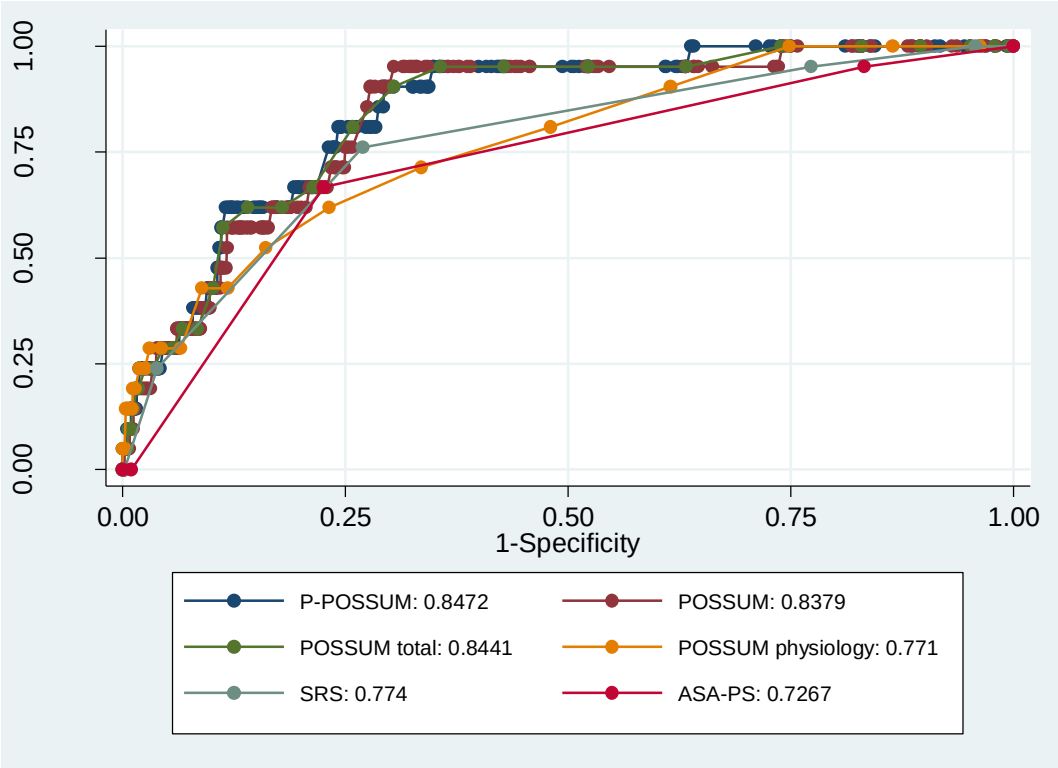

Figure 12: ROC curves for inpatient mortality (AUROC shown in legend)

Risk stratification tool	AUROC	Standard Error	Chi^2	d.f.	Probability $>Chi^2$	Bonferroni Probability $>Chi^2$
p-POSSUM	0.85	0.03				
POSSUM mortality	0.84	0.04	1.60	1	0.21	1.00
Additive POSSUM	0.84	0.03	1.10	1	0.29	1.00
POSSUM physiology	0.77	0.05	2.42	1	0.12	0.60
SRS	0.77	0.05	3.31	1	0.07	0.35
ASA-PSS	0.73	0.05	6.52	1	**0.01**	**0.05**

Table 21: Comparison of ROC curves for prediction of inpatient mortality

3.3.4.2 30-day mortality

Data were analysed for 1343 patients of the overall cohort: ASA-PSS scores were missing in 19 patients. Thirty-day mortality was 1.10% (15 deaths). ROC curve analyses are detailed in Table 22 and Figure 13. The POSSUM mortality equation, p-POSSUM, POSSUM physiology and additive POSSUM scores were all moderately accurate predictors of thirty-day mortality. The AUROCs for the SRS and ASA-PSS indicated poor accuracy, although neither AUROC value was statistically different from the 'gold standard' p-POSSUM model when Bonferroni's correction was applied (see Table 23). Neither POSSUM mortality model were well calibrated for this population.

Risk stratification tool	AUROC	Standard Error	95% C.I	Pearson Chi²	p value (for probability >Chi²)
p-POSSUM	0.76	0.05	0.65 - 0.86	425.66	**<0.0001**
POSSUM mortality	0.75	0.06	0.63 - 0.86	389.20	**<0.0001**
Additive POSSUM	0.75	0.06	0.64 - 0.86		
POSSUM physiology	0.74	0.06	0.63 - 0.86		
SRS	0.67	0.07	0.52 - 0.81		
ASA-PSS	0.61	0.08	0.46 - 0.76		

Table 22: Discrimination and calibration of risk stratification tools for 30 day mortality

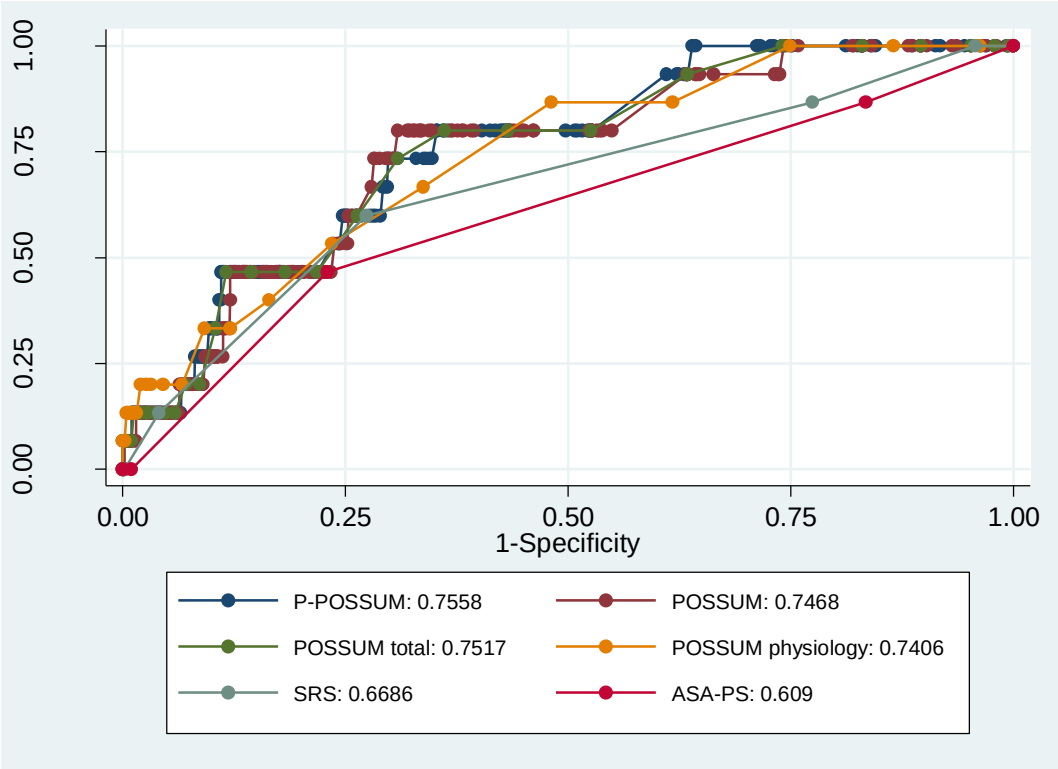

Figure 13: ROC curves for 30-day mortality (AUROC shown in legend)

Risk stratification tool	AUROC	Standard Error	Chi^2	d.f	Probability>Chi^2	Bonferroni Probability >Chi^2
p-POSSUM mortality	0.76	0.05				
POSSUM mortality	0.75	0.06	0.93	1	0.34	1.00
Additive POSSUM	0.75	0.06	1.08	1	0.30	1.00
POSSUM physiology	0.74	0.06	0.09	1	0.76	1.00
SRS	0.67	0.07	2.33	1	0.13	0.63
ASA-PSS	0.61	0.08	4.68	1	**0.03**	0.15

Table 23: Comparison of ROC curves for prediction of 30-day mortality

3.3.4.3. Morbidity

Data were analysed for 1343 patients. Morbidity (defined as any positive POMS during postoperative stay) occurred in 73.0% of patients (n=994). ROC curve analyses are detailed in Table 24 and Figure 14. The POSSUM morbidity predictor and additive POSSUM score were both moderately precise predictors of inpatient morbidity, and although the AUROC look similar, statistical comparison revealed a significant difference in discrimination; this is most likely to be due to the very high event rate in this analysis (see Table 25). The SRS, ASA-PSS and POSSUM physiology score were all poor predictors of morbidity. The POSSUM morbidity model was poorly calibrated for this population.

Risk stratification tool	AUROC	Standard Error	95% C.I.	Pearson Chi²	p value (for probability >Chi²)
POSSUM morbidity	0.73	0.01	0.71 - 0.76	673.39	**<0.0001**
Additive POSSUM	0.73	0.01	0.70 - 0.75		
SRS	0.67	0.01	0.64 - 0.70		
ASA-PSS	0.59	0.01	0.57 - 0.62		
POSSUM physiology	0.58	0.02	0.55 - 0.61		

Table 24: Discrimination and calibration of ROC curves for Day 5 POMS defined morbidity

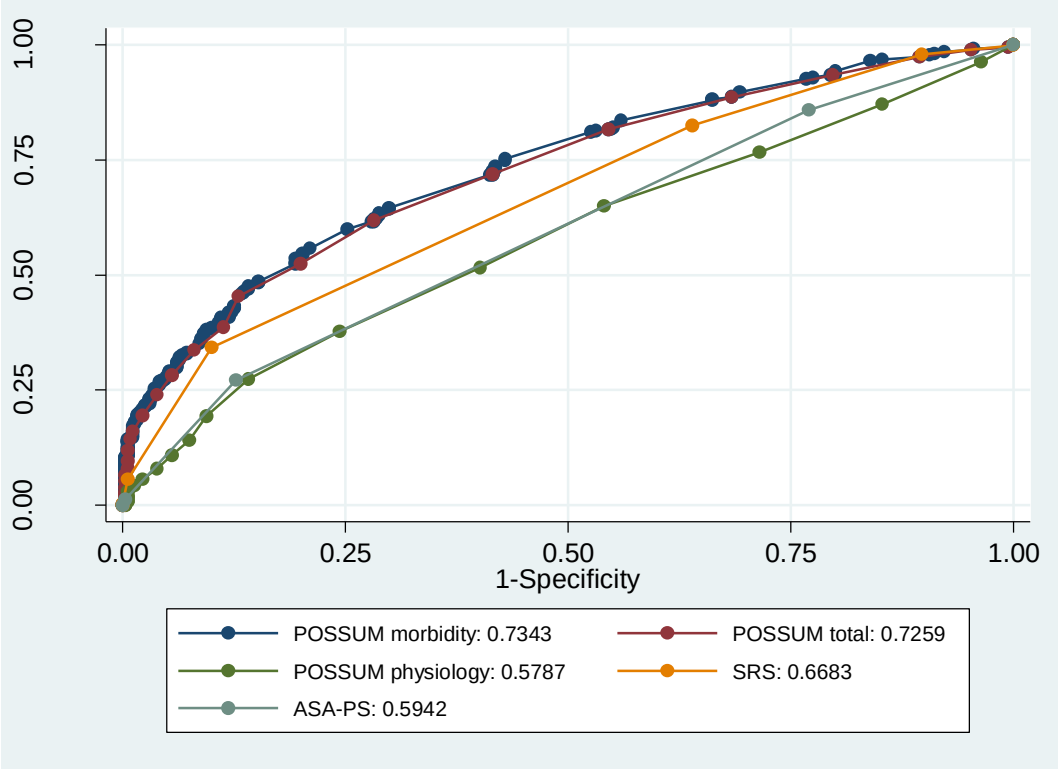

Figure 14: ROC curves for inpatient morbidity (AUROC shown in legend)

Risk stratification tool	AUROC	Standard Error	Chi^2	d.f.	Probability $>Chi^2$	Bonferroni Probability $>Chi^2$
POSSUM morbidity	0.73	0.01				
Additive POSSUM	0.73	0.01	47.42	1	**<0.0001**	**<0.0001**
SRS	0.67	0.01	16.93	1	**<0.0001**	**<0.001**
ASA-PSS	0.60	0.01	61.22	1	**<0.0001**	**<0.0001**
POSSUM physiology	0.58	0.02	155.10	1	**<0.0001**	**<0.0001**

Table 25: Comparison of ROC curves for prediction of inpatient morbidity

3.3.4.4 One - year mortality

Data were analysed for 1329 patients: 19 patients did not have an ASA-PSS score and 14 patients either exited the NHS before one year follow-up could be completed or were censored with no known date of exit. One-year mortality was 6.8% (92 deaths). The P-POSSUM, POSSUM mortality and total POSSUM predictors all demonstrated high-moderate accuracy and were not significantly different from each other. The SRS was moderately precise, but the ASA-PSS and POSSUM physiology scores were poor predictors of one-year mortality; these three scores all demonstrated significantly different accuracy from the gold standard (P-POSSUM). However, despite high precision, the calibration of the P-POSSUM was poor for one year mortality prediction in this population; analysis of deciles of O:E ratios for mean observed versus predicted risk showed that P-POSSUM significantly under-predicted one year mortality.

Risk stratification tool	AUROC	Standard Error	95% confidence intervals	Pearson Chi^2	p value (for probability >Chi^2)
P-POSSUM	0.80	0.02	0.75 - 0.84	538.25	**<0.0001**
POSSUM mortality	0.80	0.02	0.75 - 0.84	446.28	**<0.0001**
Additive POSSUM	0.80	0.02	0.75 - 0.85		
SRS	0.71	0.03	0.66 - 0.76		
POSSUM physiology	0.67	0.03	0.61 - 0.73		
ASA-PSS	0.66	0.03	0.61 - 0.71		

Table 26: Discrimination and calibration of risk stratification tools for one-year mortality

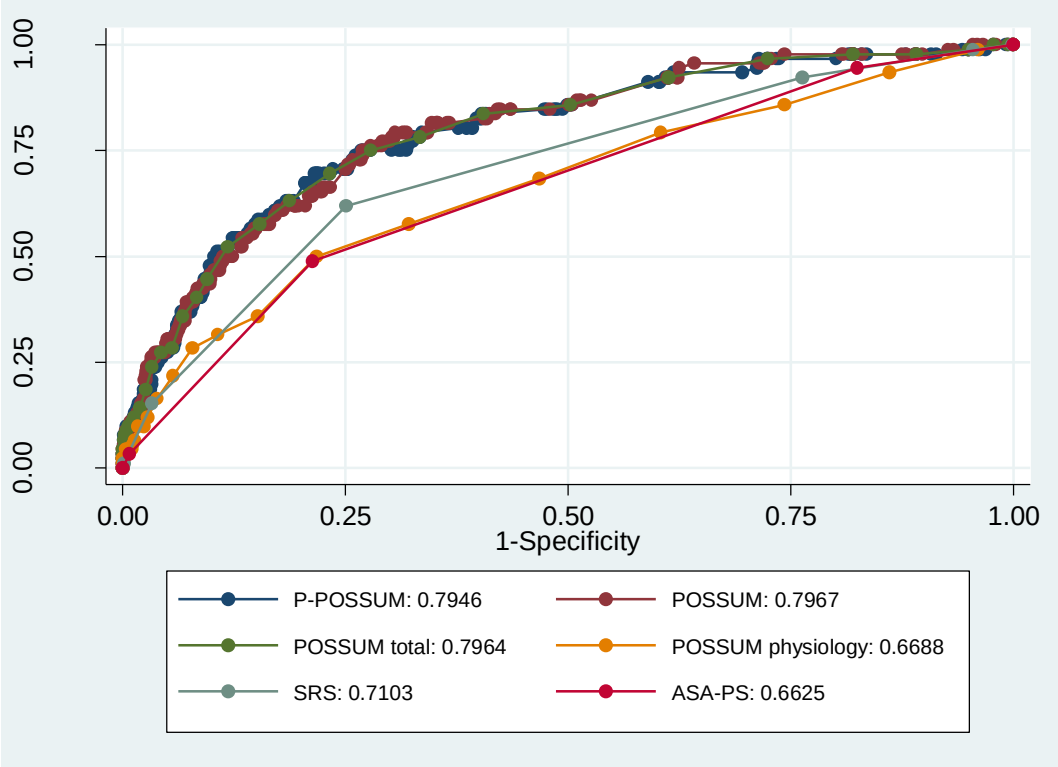

Figure 15: ROC curves for one-year mortality (AUROC shown in legend)

Risk stratification tool	AUROC	Standard Error	Chi^2	d.f.	Probability>Chi^2	Bonferroni Probability >Chi^2
POSSUM mortality	0.80	0.02				
p-POSSUM	0.80	0.02	0.27	1	0.61	1.00
Additive POSSUM	0.80	0.02	1.12	1	0.29	1.00
SRS	0.71	0.03	21.65	1	**<0.003**	**<0.02**
POSSUM physiology	0.67	0.03	18.69	1	**<0.0001**	**<0.0001**
ASA-PSS	0.66	0.03	17.65	1	**<0.0001**	**<0.002**

Table 27: Comparison of ROC curves for prediction of one-year mortality

3.4 Discussion

3.4.1 Summary of findings

The POSSUM risk stratification tools (both the original model and the P-POSSUM variant as well as the additive POSSUM) are able to discriminate between patients at different levels of risk for the prediction of mortality with high precision for death while in hospital and with moderate precision for death within thirty days. In addition, they have good precision for predicting mortality within one year of surgery. While the POSSUM physiology score had good precision for inpatient and thirty day mortality and fitted the data well (adequate calibration), it was a poor predictor of one year mortality. All the evaluated risk stratification tools were more precise predictors of inpatient mortality than 30 day mortality. They were also all better predictors of mortality at any of the

measured end-points, than morbidity as defined by a positive POMS result at any stage during the inpatient episode.

Although by the standards defined in the analysis plan, the P-POSSUM, total POSSUM and POSSUM mortality predictors reached high-moderate precision for inpatient mortality precision (all AUROC ≥0.84), these systems were not significantly better than the Surgical Risk Scale and POSSUM physiology score (both AUROC 0.77) on statistical analysis. However, this lack of statistical difference may be due to a Type 1 error, as the event rate was very low (inpatient mortality 1.5%).

Both POSSUM and P-POSSUM were poorly calibrated for this population.

3.4.2 Results in context

These results are consistent with the findings of the systematic review in Chapter 2: the POSSUM and P-POSSUM rules are moderately accurate tools for the prediction of adverse perioperative outcome. The SRS was less accurate; the difference was not statistically significant but this may have been due to inadequate study power. As the SRS can be calculated preoperatively, it may be a useful predictor for use in clinical practice; similarly the POSSUM physiology score was a moderate discriminator of 30-day and inpatient outcome.

Mortality rates for this cohort in the immediate perioperative period (inpatient and 30-day) are comparable with those quoted in the literature. The inpatient mortality in a UK database (CHKS) analysis of over four million non-cardiac non-neurosurgical

admissions was 1.9%,[7] while the inpatient death rate for SOuRCe cohort in this study was 1.54%; however, the cohort characteristics were different (SOuRCe cohort – major elective surgery, CHKS database – all inpatient surgery, including minor surgery and emergency surgery). The European Surgical Outcomes Study (EuSOS) reported an overall in-hospital mortality of 4.0%; however, 24.7% of patients in the EUSOS study had urgent or emergency operations.[2] In the systematic review in Chapter two of this thesis, inpatient mortality (where it was reported) in unselected cohorts varied between 0.8 [276] and 4.3%.[281]

The accuracy and calibration of all the risk stratification tools evaluated for predicting postoperative morbidity was poor. This is consistent with the results of the systematic review which found that all but one of the studies where both mortality and morbidity were measured, predictive precision was higher for mortality. The main reason for these findings is likely to be variation in the way in which complications are described in different clinical and academic settings. Although the POMS is a standardised validated outcome measure, the definitions of morbidity which are used in the POMS mean that morbidity estimates may be higher than when other methods are used to define complications.[183] If a different POMS related outcome measure (for example Day15 POMS status) had been used, the results might have been quite different. From these data, and those in the systematic review, it may be concluded that the lack of standardisation in the definition and measurement of postoperative morbidity obviate the potential usefulness of morbidity prediction tools.

In interpreting these results, consideration must be given to the purpose of using these tools. Broadly speaking, there are three reasons why clinicians would choose to use a clinical prediction rule: to guide perioperative management, to facilitate and improve the quality of informed patient consent, and to enable comparative audit of outcomes between individual clinicians, teams or institutions. Each of these will now be discussed in turn.

3.4.2.1 Using risk stratification tools to guide perioperative management

A number of recent reports have highlighted failings in preoperative risk assessment for surgical patients, leading to high risk patients being managed in low acuity ward areas, even though they may have benefitted from a higher level of care (Level two or three critical care or Post Anaesthetic Care Unit).[7;237;241] Two of these papers define the high risk category as those patients with a predicted mortality of >5%, and suggest that these patients should be cared for in critical care postoperatively. [7;237]

The recent NCEPOD report showed that anaesthetists considered 20% of the patients who had surgery during the time of the audit to be high risk. It is unlikely, that at least in the near future, that 20% of patients in the UK might be offered critical care or PACU admission in the postoperative period. Clinical experience shows that despite scoring systems such as the POSSUM predictors being widely known about, and calculators being available on the internet and through smart phone apps, they are not widely used. There may be a number of reasons for this, including uncertainty over the accuracy of these tools, and lack of willingness to sit at a computer / calculator and enter the

numbers into the required equations. There are also concerns over the calibration of these tools over time, which are borne out by the results from this study: the original POSSUM model was poorly calibrated for inpatient mortality prediction, whereas the P-POSSUM and additive POSSUM scores fitted the data better. As surgical and anaesthesia techniques improve, calibration drift will continue to occur; however, the systematic review data and this study demonstrate that the additive POSSUM score itself has remained an accurate discriminator of risk. Therefore there may be merit in simply using the total score as a marker of perioperative risk.

This study shows that the predictive accuracy of the POSSUM physiology score was not significantly different to the gold standard P-POSSUM score for the prediction of inpatient mortality. This lack of difference may be due to lack of statistical power. Nevertheless, by the criteria set out at the beginning, POSSUM physiology was moderately precise (AUROC 0.77). The population of patients with a POSSUM physiology score of ≥21 (10% of the study population) had an inpatient mortality rate of 5.39%. Therefore, based on these data, it would be reasonable to consider postoperative critical care for all patients in whom major elective surgery is planned, and who scored a POSSUM physiology total of ≥21. This would enable planned referral to critical care or PACU for patients from the preoperative assessment clinic. While the tool is not sufficiently precise to enable it to be the sole determinant of perioperative management strategy, it would be a reasonable adjunct to clinical judgement, and other risk stratification tools such as cardiopulmonary exercise testing.

The POSSUM total score was an accurate discriminator of inpatient mortality. Patients in the top quartile of risk as predicted by a total POSSUM score of ≥31 were part of a population with an inpatient mortality of 4.70%; the highest risk 10% of the cohort (POSSUM total ≥36) had an inpatient mortality of 6.25%. Therefore, for patients who had not been previously booked critical care beds on the basis of POSSUM physiology score alone, if at the end of surgery the estimated POSSUM total score is ≥36, then critical care admission should be strongly recommended; it should also be considered for patients with a score of ≥31, although this may encompass a large number of perioperative patients, and therefore may not be practicable. However, ensuring a higher level of surveillance on a regular ward, for example by "outreach" or "patient at risk" teams, would potentially be of benefit in ensuring early detection of complications, and appropriate escalation of management. This sort of approach, of re-evaluating risk at the end of surgery as part of a 'bundle' of care, is advocated in recent guidance.[237]

3.4.2.2 Using risk stratification tools to improve informed consent

The General Medical Council's guidance 'Consent: patients and doctors making decisions together', [314] has this to say about the communication of risks to patients:

"You must tell patients if an investigation or treatment might result in a serious adverse outcome, even if the likelihood is very small. You should also tell patients about less serious side effects or complications if they occur frequently, and explain what the patient should do if they experience any of them."

Elsewhere in the guidance it states:

"Your discussions with patients should focus on their individual situation and the risk to them."

Therefore, the communication of individualised estimations of mortality and morbidity are both required as part of the informed consent process. As the POSSUM physiology score demonstrated reasonable precision for the prediction of inpatient death, to use this as a tool to guide the preoperative consent process in patients having major elective surgery is valid. While it is not an ideal predictor (as would be defined by an AUROC >0.9) it is certainly more accurate than quoting the population risk for the procedure overall, which may be default standard practice at present. Therefore, as with the use of these tools for guiding perioperative management strategy, it would be reasonable to use the POSSUM physiology score as part of a multi-modal assessment which may personalise and improve the quality of informed consent.

3.4.2.3 Using risk stratification tools for risk adjustment for comparative audit

The calibration of the prediction rules is particularly important when considering their use for comparative audit. If calibration is imprecise, it is possible that problems may go unnoticed, if observed: expected ratios are satisfactory. In Copeland's original POSSUM paper, the cohort mortality was 4.0%;[38] and in Prytherch's first external validation of the P-POSSUM in 1998, mortality was 2.9%.[69] The mortality in this study cohort was 1.1% at 30 days and 1.6% at hospital discharge. These differences in mortality may be due to improvements in standards of care which have occurred over time, or differences in the characteristics of the patient populations which were studied; either way, the

difference in calibration of the POSSUM and P-POSSUM models between previously published studies and this one can be clearly seen.

3.4.3 Limitations of this study

The main limitation of this study is that it was conducted in an elective population of patients; therefore its generalizability to non-elective patients is unknown. Second, these are data from 2001-5; changes in practice may mean that the discrimination and calibration of the prediction rules which have been established in this study are no longer representative even of elective surgical cohorts in 2013. Nevertheless, as previously stated, the mortality rates of this population are consistent with contemporary estimates and changes in practice are less likely to affect the discrimination of such measures; it is predominantly calibration that would be affected by changes (improvements) in clinical care; these data suggest that both the POSSUM and P-POSSUM tools are poorly calibrated for this cohort.

3. 5 Conclusions

1. The additive POSSUM score is a moderately accurate predictor of short term outcome (inpatient and 30 day mortality) in patients undergoing elective major non-cardiac surgery, and comparable with the POSSUM and P-POSSUM regression models.

2. The POSSUM physiology score has moderate precision for the prediction of postoperative mortality, and therefore may be used as part of a multi-modal assessment process to guide perioperative patient management and improve the quality of informed consent.

3. The POSSUM and P-POSSUM models were poorly calibrated for this population of patients.

4. The SRS and ASA-PSS were not accurate predictors of 30-day mortality in this cohort of patients.

Chapter Four: Epidemiology of outcome after major non-cardiac surgery

4.1 Aim and Objectives

This chapter reports the postoperative outcomes of the SOuRCe cohort of patients who underwent major non-cardiac surgery at the Middlesex Hospital between 2001 and 2005.

4.1.1 Aim

1. To describe the epidemiology of postoperative morbidity in the Middlesex Hospital SOuRCe cohort, and to analyse its relationship with other outcomes, both in the shorter term (inpatient and 30-day) and longer term (years).

2. To use these data to determine whether Day 5 POMS has face validity as a measure of postoperative morbidity

4.1.2 Objectives:

1. Description of the epidemiology of postoperative morbidity, as measured by the Post Operative Morbidity Survey
2. Analysis of the univariate relationships between risk (stratified by the P-POSSUM, additive POSSUM, POSSUM physiology, Surgical Risk Scale and ASA-PS predictors) and shorter and longer-term mortality

3. Analysis of the univariate relationships between postoperative morbidity (POMS defined: overall, by physiological domain and by postoperative day) and shorter and longer-term mortality (dichotomous outcome at different end-points)
4. Analysis of the univariate relationship between duration of morbidity and longer-term survival.
5. To measure the relationship between Day 5 POMS and other postoperative outcomes in order to make an assessment of the face validity of Day 5 POMS as a point-prevalence measure of postoperative morbidity.

4.2 Methods

4.2.1 Background

This was a prospective observational longitudinal cohort study. The cohort studied is the same as in Chapter 3 (known from here forward as the SOuRCe cohort). Recruitment, ethics, data collection methods and dataset have all been described in detail in Chapter 3; however, a brief summary is as follows:

The SOuRCe cohort comprised 1362 patients who had major inpatient surgery (orthopaedic, urological, general or vascular) between 2001 and 2005 at the Middlesex Hospital, London. Data collected on all of these patients included the variables of POSSUM, ASA-PSS grade, and surgical type and severity. Post operative morbidity was measured on days 3, 5, 8 and 15 postoperatively by administration of the Post Operative Morbidity Survey. Local ethics committee approval was granted for the local collection

of perioperative data; the National Information Governance Board's Ethics and Confidentiality Committee granted approval (Section 251 exemption) for the linkage of these data to long-term mortality statistics

4.2.2 Analysis Plan

4.2.2.1 General description of study population and postoperative morbidity

The demographic characteristics of the study population have been described in Chapter 3. Mortality is reported as a dichotomous outcome at four end-points: one location based (inpatient mortality) and three time-based (30-day, one and five years). Morbidity is reported as a dichotomous outcome under the following definitions: any inpatient morbidity (POMS positive at any stage); POMS domain defined (separate dichotomous outcomes for each of the nine domains); POMS day defined (separate dichotomous outcomes for each of the four days - (3, 5, 8 and 15) - on which POMS was measured)

4.2.2.2 Long-term survival

Univariate analysis of long term survival using Kaplan Meier curves compared patients by risk category for the ASA-PS, additive POSSUM score, POSSUM physiology score and Surgical Risk Scale, using the category definitions described in Chapter 3.

Kaplan Meier curves were also used for univariate comparison of patients with and without any POMS defined morbidity, and for patients categorised by POMS defined morbidity status on each of Days 3, 5, 8 and 15.

4.2.2.3 Severity / duration of morbidity

The POMS has not been validated as a cumulative score to reflect severity of morbidity.[56] However, there are known associations between patient risk and postoperative length of stay,[165] and between complications and length of stay;[165;183] furthermore, previous studies have shown that there is a strong correlation between severity of complications and length of stay. [195] It is therefore proposed that the duration of morbidity, defined as the 'final morbidity day' (FMD) is a reasonable surrogate for severity of complications: that is, that patients who remain in hospital with postoperative morbidity on Day 15 and therefore have a longer duration of postoperative morbidity, have more significant complication(s) than patients whose last morbidity day was 8, 5 or 3; similarly, patients who are in the 'FMD 8' category have more significant morbidity than FMD 3 or 5, and so on.

Kaplan-Meier survival curves were constructed for univariate comparison of long-term survival based on morbidity severity, as defined by 'Final Morbidity Day'.

4.2.2.4 Statistical analysis

Bi-variate analyses of dichotomous outcomes (mortality, morbidity) were conducted using Chi squared testing with Bonferroni's correction for multiple comparisons when multiple end-points were evaluated. The Student's t-test (2-sided) was used for comparison of means (length of stay). Survival times were compared using Log-rank statistics. A p value of <0.05 was considered significant. Bonferroni corrected p values are denoted p'.

Stata InterCooled (Release 12.1) software (StataCorp LP, College Station, Texas, USA) was used for all analyses and calculations

4.3 Results

4.3.1 General description

Long-term follow up data were available for 1347 patients; (lost to follow up n=1; no trace n=14). The mortality rates by cohort at hospital discharge, 30 days, one year and five years is summarised in Table 28.

		Overall	Cohort One	Cohort Two
Inpatient	N	1362	438	924
	Mortality: n (%)	21 (1.5)	5 (1.1)	16 (1.7)
30 day	n	1362	438	924
	Mortality: n (%)	15 (1.1)	3 (0.7)	12 (1.3)
1 year	n	1347	432	915
	Mortality: n (%)	92 (6.8)	33 (7.5)	59 (6.6)
5 year	n	1339	424	915
	Mortality: n (%)	277 (20.7)	86 (20.3)	191 (20.9)

Table 28 : Comparison of mortality at different endpoints between Cohort One and Two

4.3.2 Mortality by surgical speciality – univariate analyses

There were statistically significant differences between surgical specialities for both 30 day ($p<0.01$) and inpatient ($p<0.001$) mortality (see Table 29).

Using Chi squared testing to compare thirty day mortality between the four specialities a significant difference was found between the groups; paired comparisons (using Bonferroni's correction, and p' values) showed that this difference was due to lower mortality in urology (p'<0.05) and orthopaedics (p'<0.005) when compared with vascular surgery. Inpatient mortality was lower in orthopaedics than general (p'<0.001) and vascular surgery (p'<0.001). All other paired comparisons demonstrated similar inpatient mortality rates between specialities.

At one year, orthopaedic patient mortality was significantly lower than that for general and vascular patients (orthopaedic versus general p'<0.001; orthopaedic versus vascular p'<0.001) There was no significant difference between urology and vascular, or urology and orthopaedic mortality at one year. However, mortality was significantly lower at one year after urological surgery when compared with the general surgery cohort (p'<0.05).

At five years, mortality was significantly lower in orthopaedic patients when compared to general surgery (p'<0.001) and vascular (p'<0.001). Five year mortality after urological surgery was significantly lower than after general surgery (p'<0.001) but not significantly different to the vascular and orthopaedic cohorts.

	Inpatient mortality n=1362 (%)	30 day mortality n=1362 (%)	1 year mortality n=1347 (%)	5 year mortality n=1339 (%)
All patients	1.5	1.1	6.8	20.7
Orthopaedics	0.5	0.6	2.9	13.4
Urology	1.4	0.0	6.2	18.3
General	3.3	2.4	16.4	40.4
Vascular	7.8	4.7	15.9	32.8
p value	**p<0.001**	**p<0.01**	**p<0.001**	**p<0.001**

Table 29: Postoperative mortality according to surgical speciality

4.3.3 Inpatient morbidity: overview

Figure 16 shows the prevalence of postoperative morbidity (as defined by POMS on any of days 3, 5 8 or 15) in the whole population of patients studied.

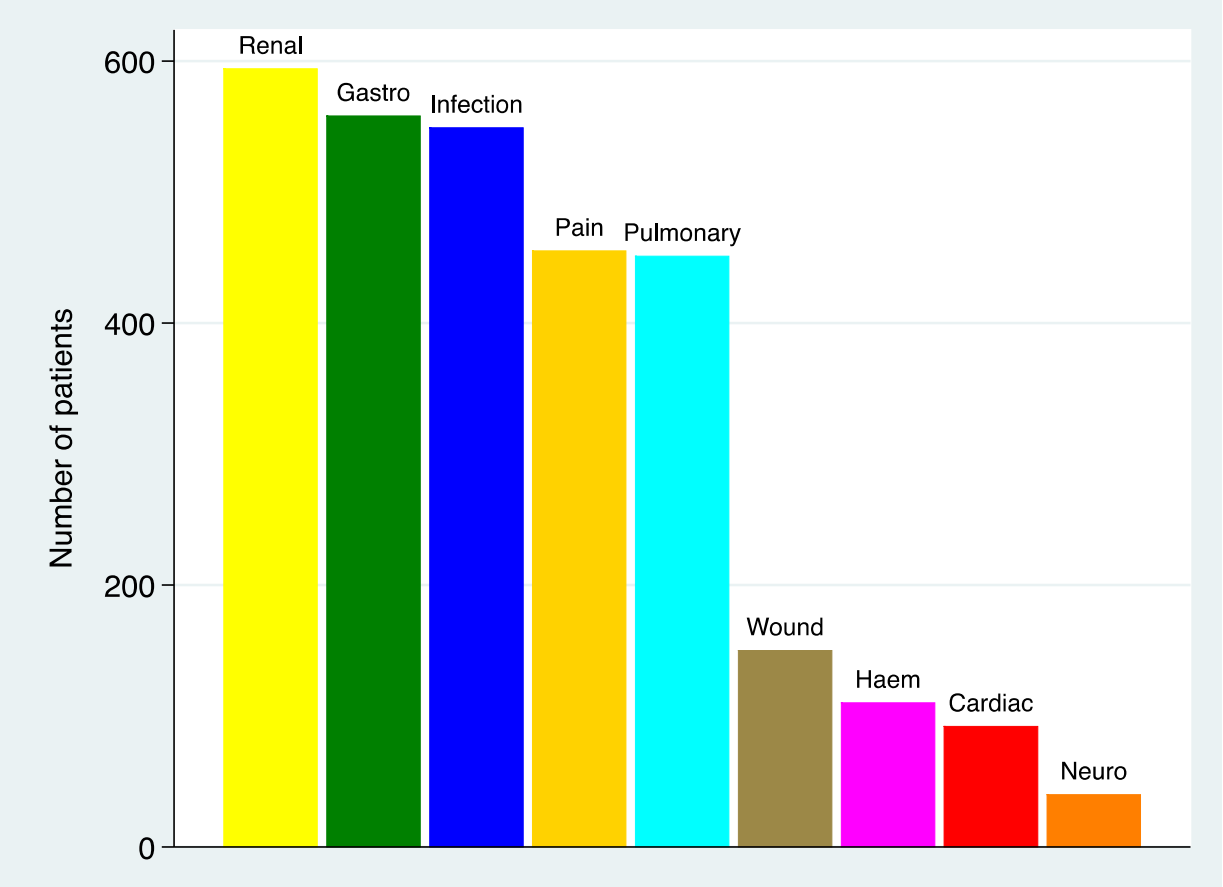

Figure 16: Prevalence of POMS defined morbidity occurring on any of Days 3, 5, 8 or 15 postoperatively (n=1362)

Table 30 summarises POMS defined morbidity across surgical specialities. There were significant differences in postoperative morbidity between specialities, both overall (occurrence of any morbidity as defined by a positive POMS), and on comparing the proportion of patients who scored on the POMS on each postoperative day on which it was measured. There were also significant differences in POMS defined morbidity in seven out of nine domains, with the exceptions of the wound and haematology categories. In the pulmonary, infection, renal, gastrointestinal and pain domains these

differences were accounted for by a lower incidence of morbidity in orthopaedic patients when compared to each of the other surgical specialities. Vascular patients were significantly more likely to develop cardiac or neurological morbidity.

	Total (n=1362)	Orthopaedic (n=855)	General (n=296)	Urology (n=147)	Vascular (n-64)	p' value
Any Morbidity	72.9	62.9	90.9	91.2	82.8	**<0.001**
Pulmonary	33.1	22.6	52.4	50.3	45.3	**<0.001**
Infection	40.3	32.4	47.3	64.6	57.8	**<0.001**
Renal	43.6	31.1	59.5	76.2	62.5	**<0.001**
Gastrointestinal	40.9	23.9	83.8	55.1	39.1	**<0.001**
Cardiac	6.8	5.6	7.4	4.8	23.4	**<0.001**
Neurological	2.9	2.5	2.0	4.8	9.4	**<0.01**
Wound	11.0	11.1	11.5	9.5	10.9	NS
Haematology	8.1	7.8	6.8	10.9	10.9	NS
Pain	33.4	20.1	60.8	53.7	37.5	**<0.001**
Day 3 POMS positive	67.8	55.6	90.2	90.5	75.0	**<0.001**
Day 5 POMS positive	50.0	35.8	77.4	70.1	64.1	**<0.001**
Day 8 POMS positive	31.5	21.0	53.0	44.9	42.2	**<0.001**
Day 15 POMS positive	15.6	8.2	29.4	26.5	25.0	**<0.001**

Table 30: POMS comparisons between surgical specialities (percentages)

4.3.4 Patterns of morbidity after different types of surgery

The patterns of morbidity in the four different surgical speciality cohorts are graphically displayed in Figures 17 to 24. There was a significant difference between the prevalence of the top five domains of morbidity (pulmonary, infection, gastrointestinal, renal or pain) and the bottom four (cardiac, wound, neurological or haematological morbidity) (947 versus 313 patients; p<0.0001).

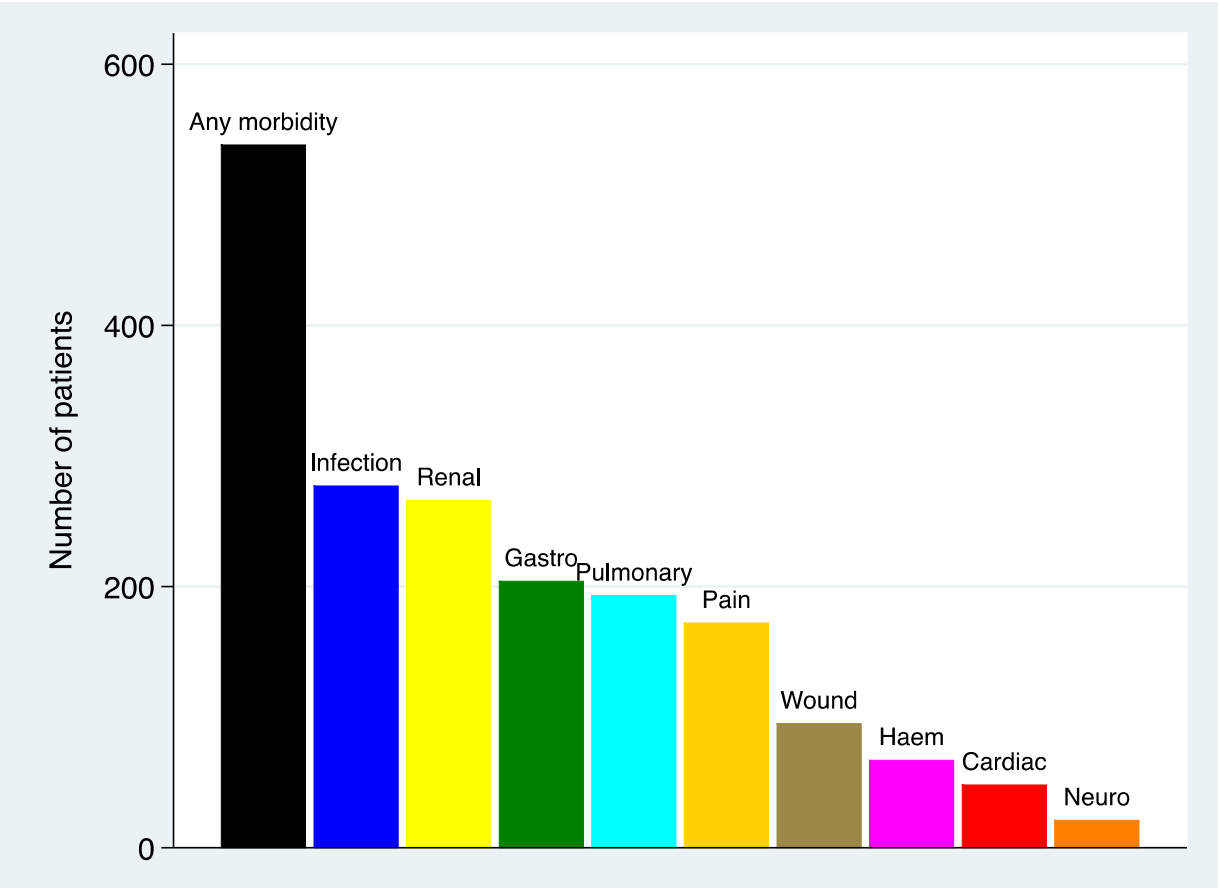

Figure 17: Domains of POMS defined morbidity after orthopaedic surgery (n=855)

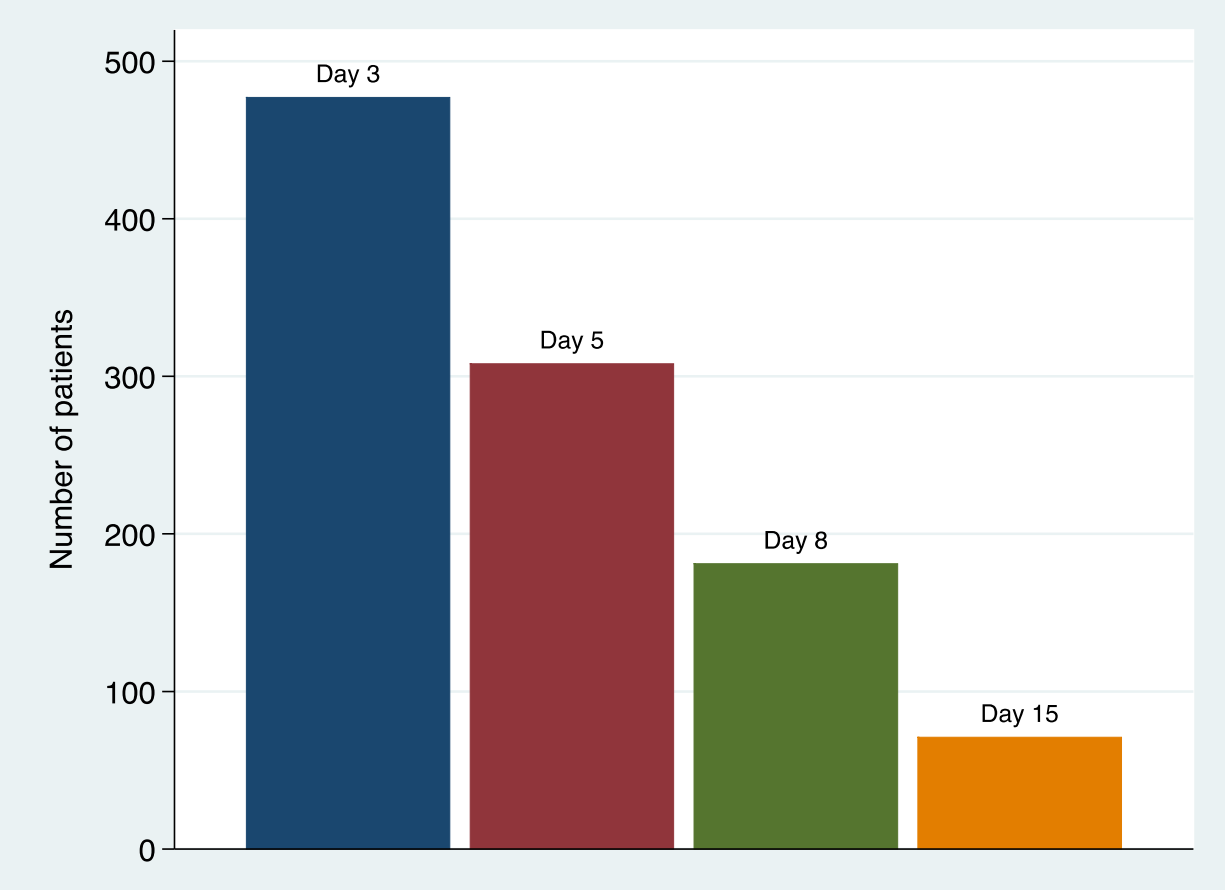

Figure 18: Prevalence of POMS defined morbidity after orthopaedic surgery (n=855)

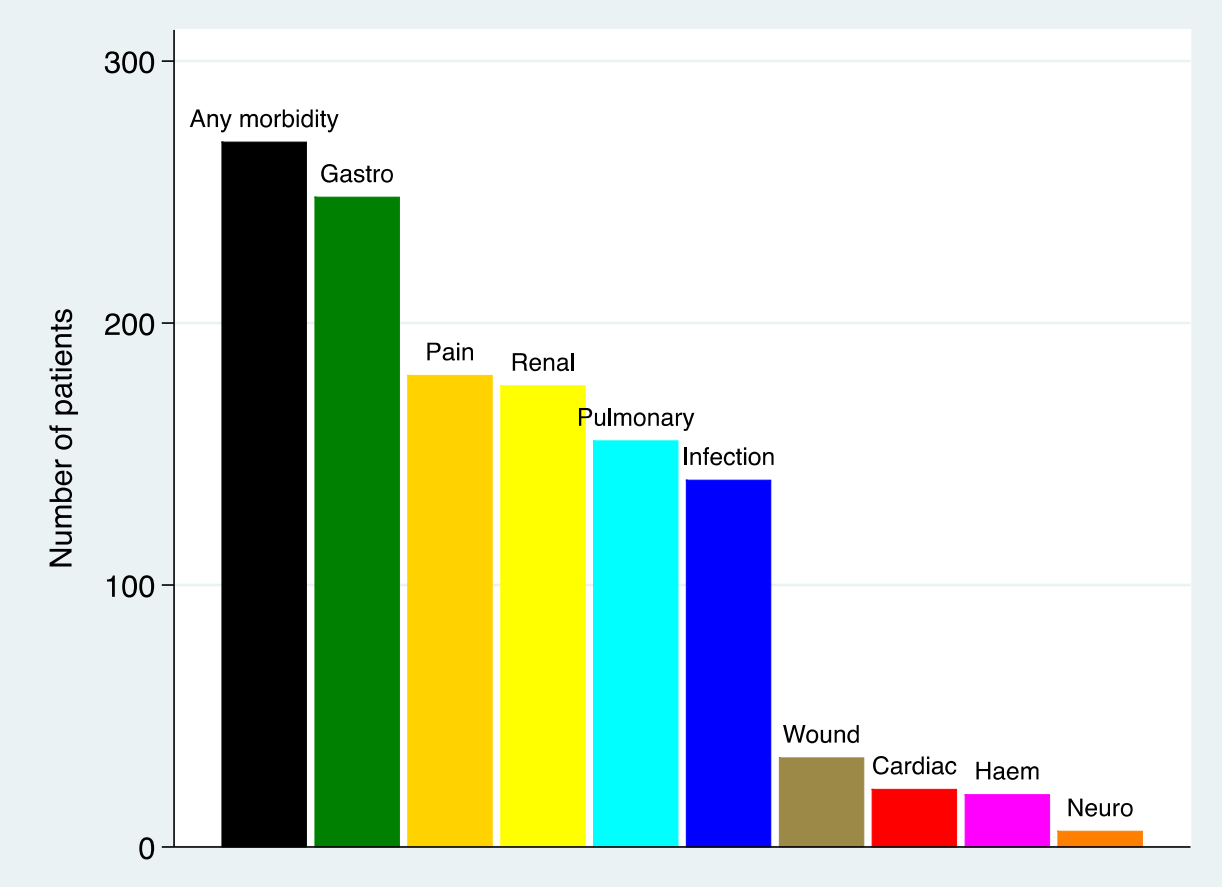

Figure 19: Domains of POMS defined morbidity after general surgery (n=296)

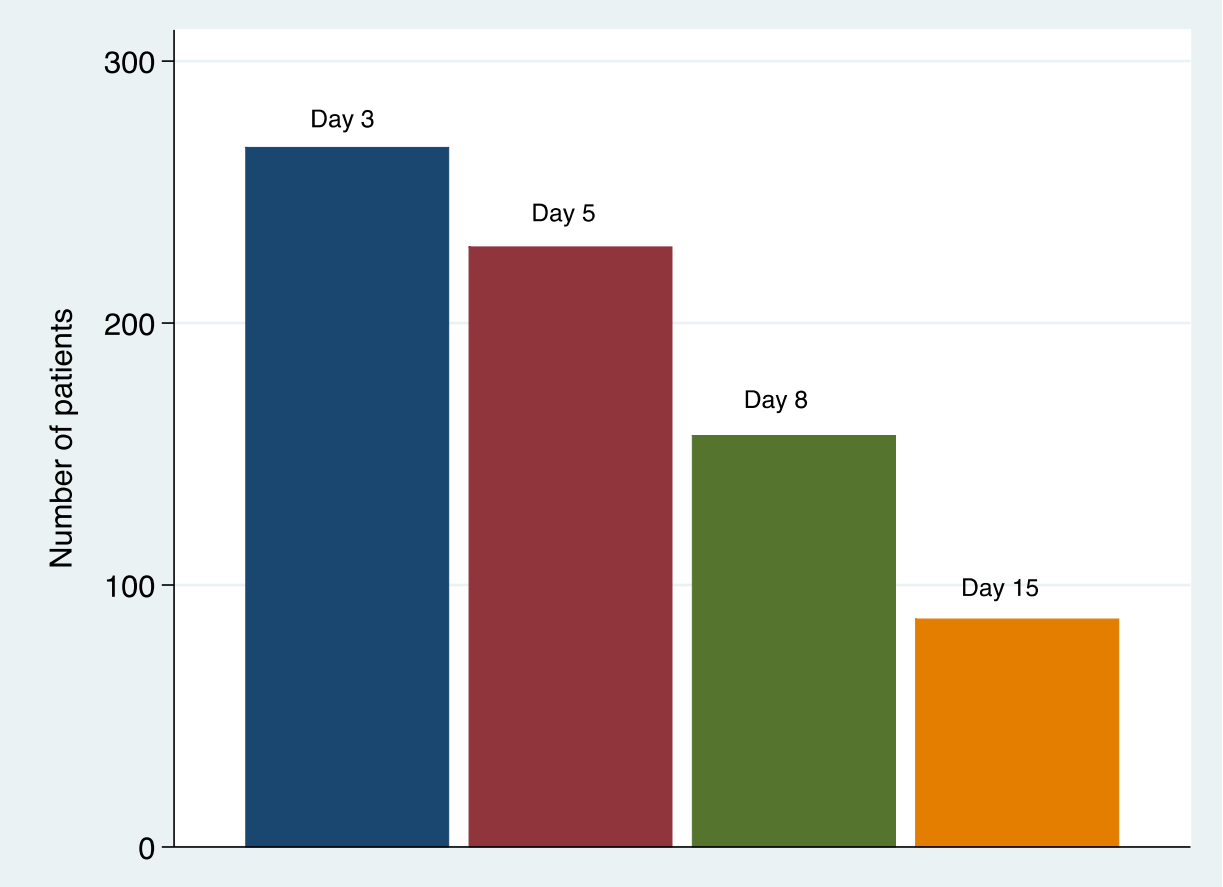

Figure 20: Prevalence of POMS defined morbidity after general surgery (n=296)

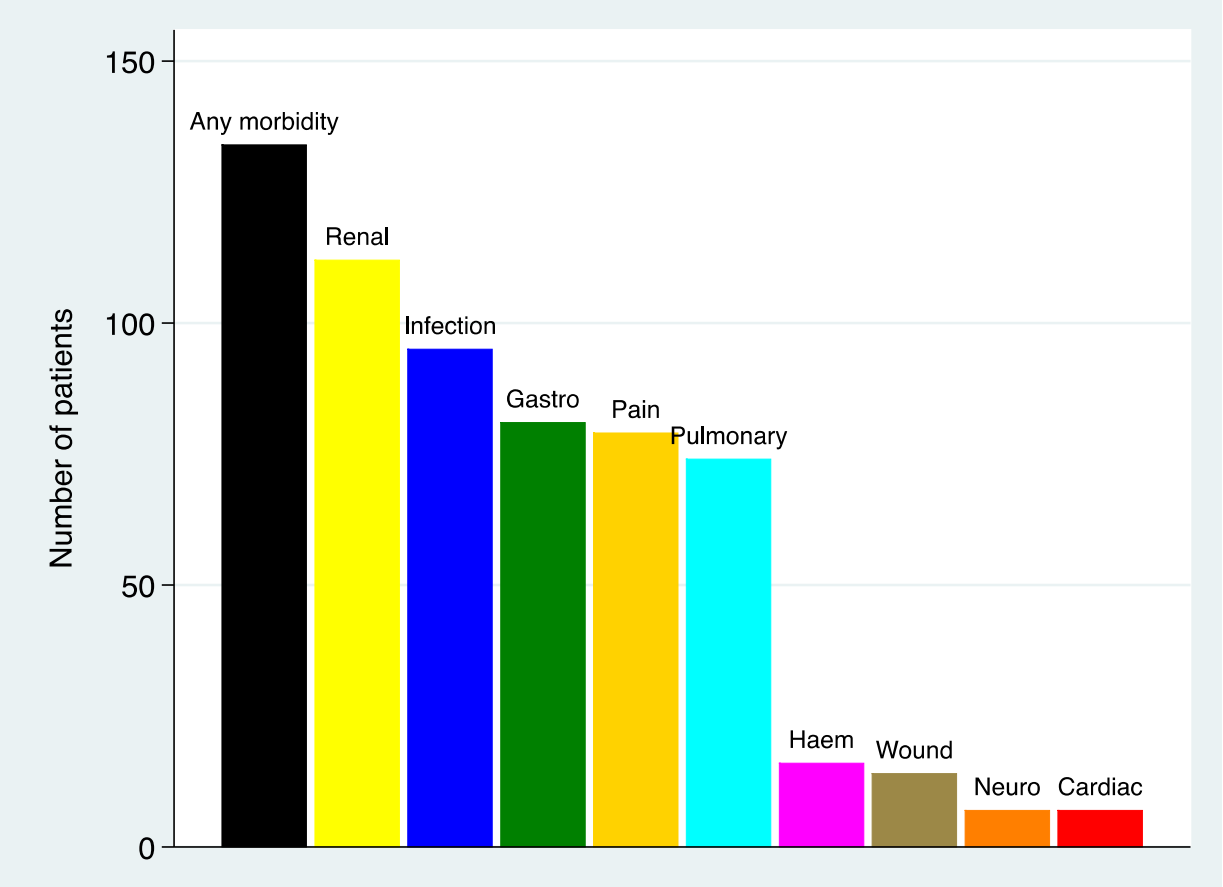

Figure 21: Domains of POMS defined morbidity after urological surgery (n=147)

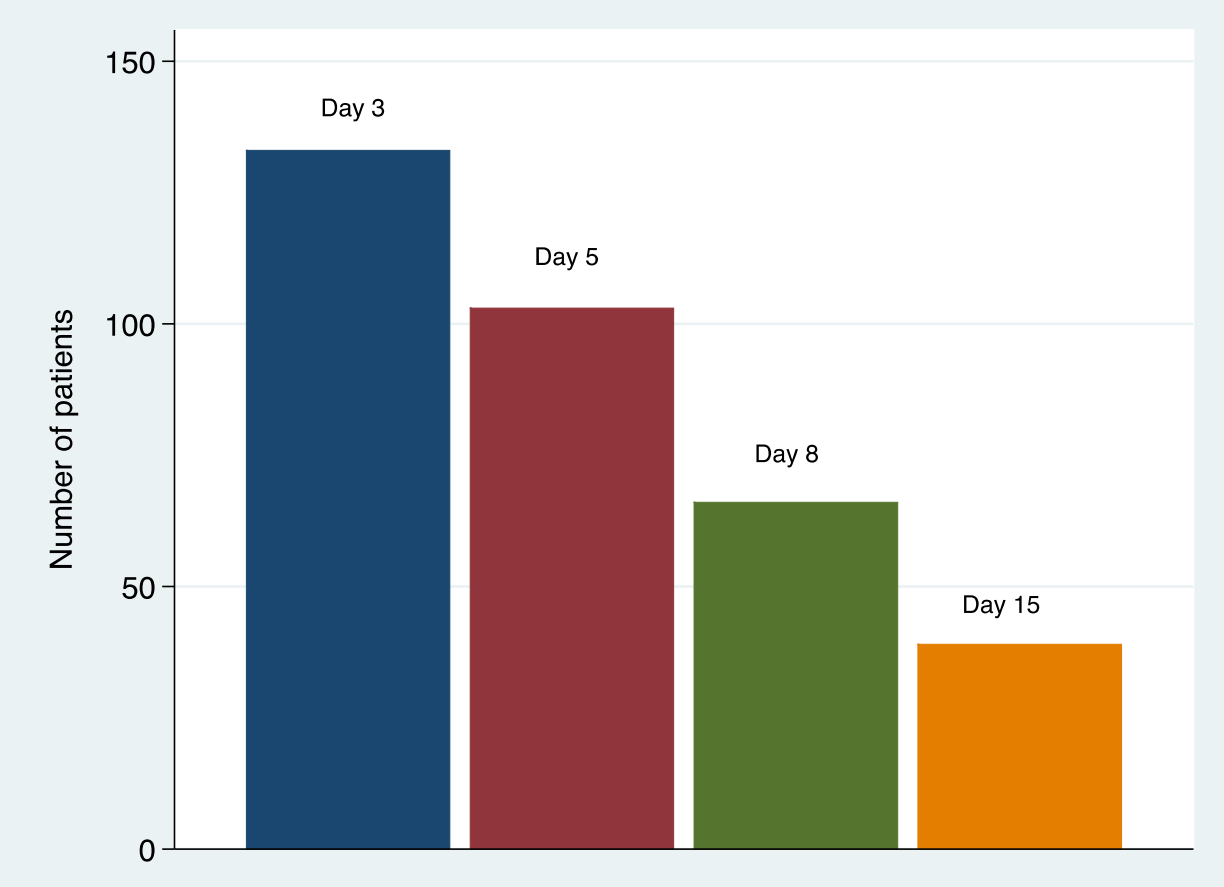

Figure 22: Prevalence of POMS defined morbidity after urological surgery (n=147)

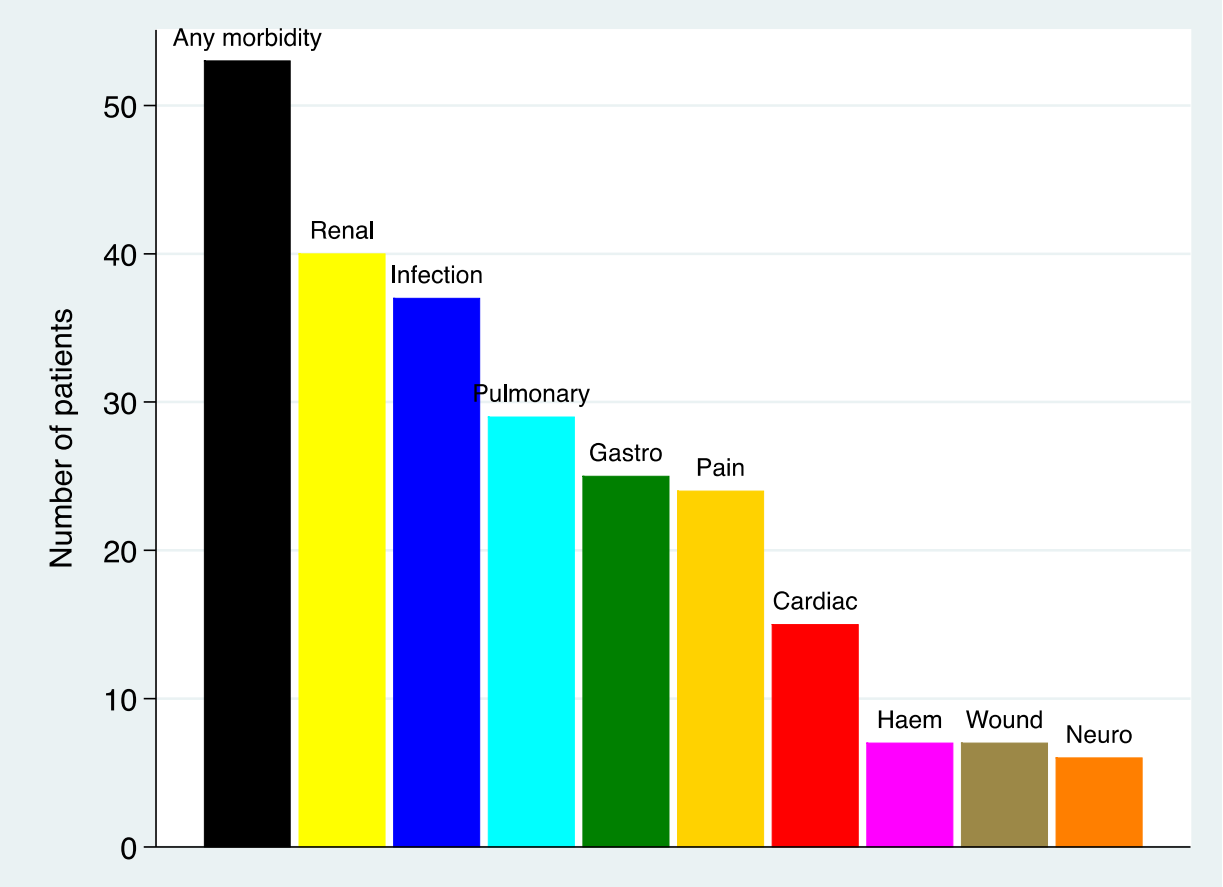

Figure 23: Domains of POMS defined morbidity after vascular surgery (n=64)

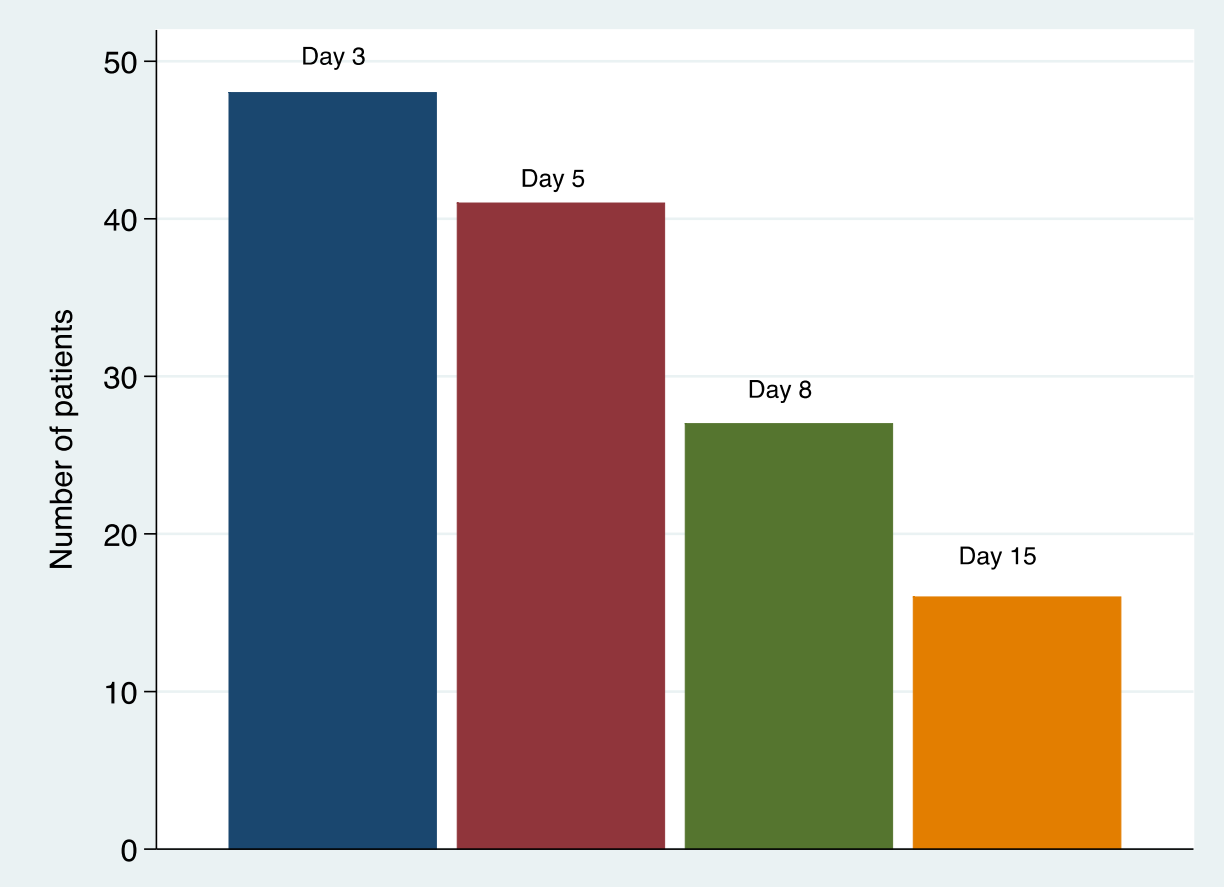

Figure 24: Prevalence of POMS defined morbidity after vascular surgery (n=64)

4.3.5 Patterns of morbidity according to postoperative day

The patterns of morbidity changed over the postoperative period. For seven out of the nine domains, morbidity prevalence was highest on Day 3 postoperatively and reduced over time. However, cardiac morbidity peaked on Day 5 (see Figure 30) and wound morbidity, while uncommon in the immediate postoperative period, became more common as time progressed, peaking on Day 8 (see Figure 31)

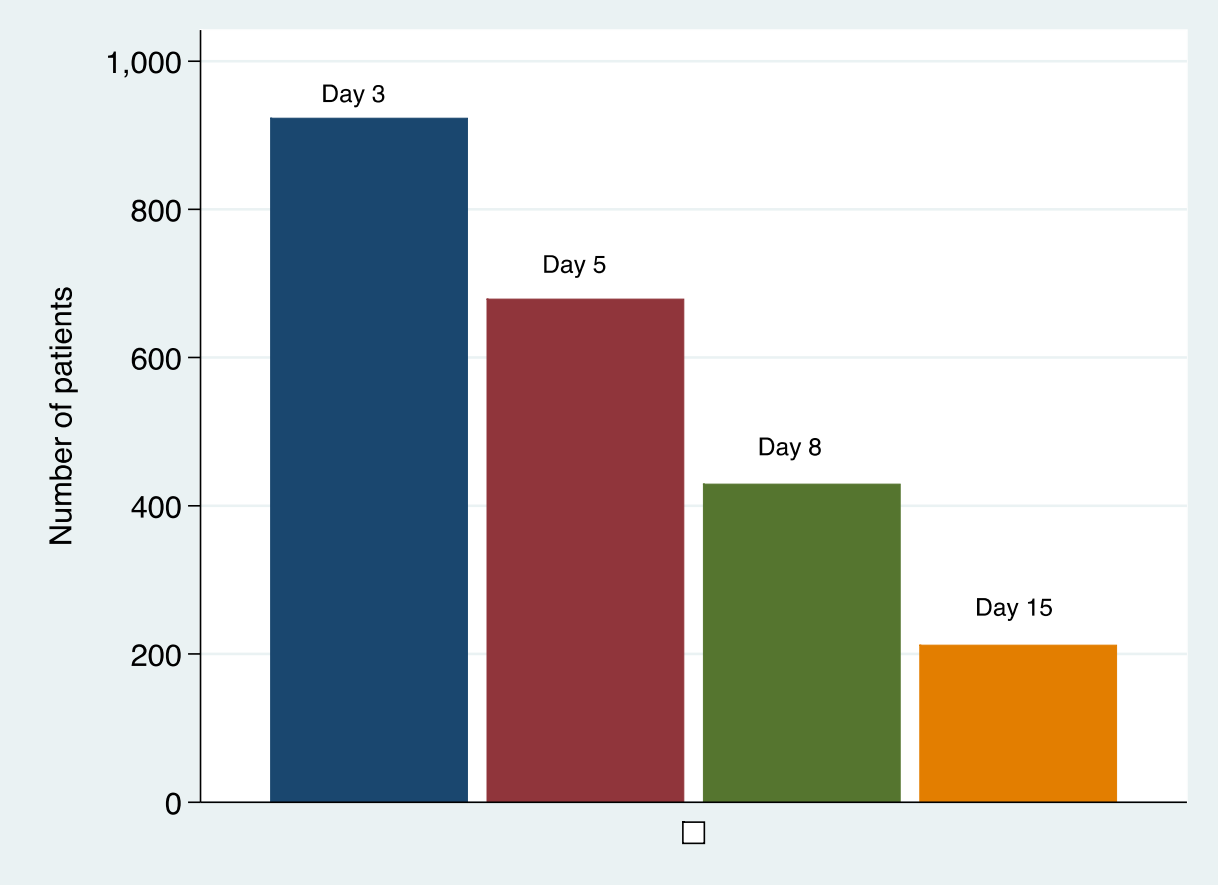

Figure 25: Prevalence of POMS defined morbidity in overall cohort (n=1362)

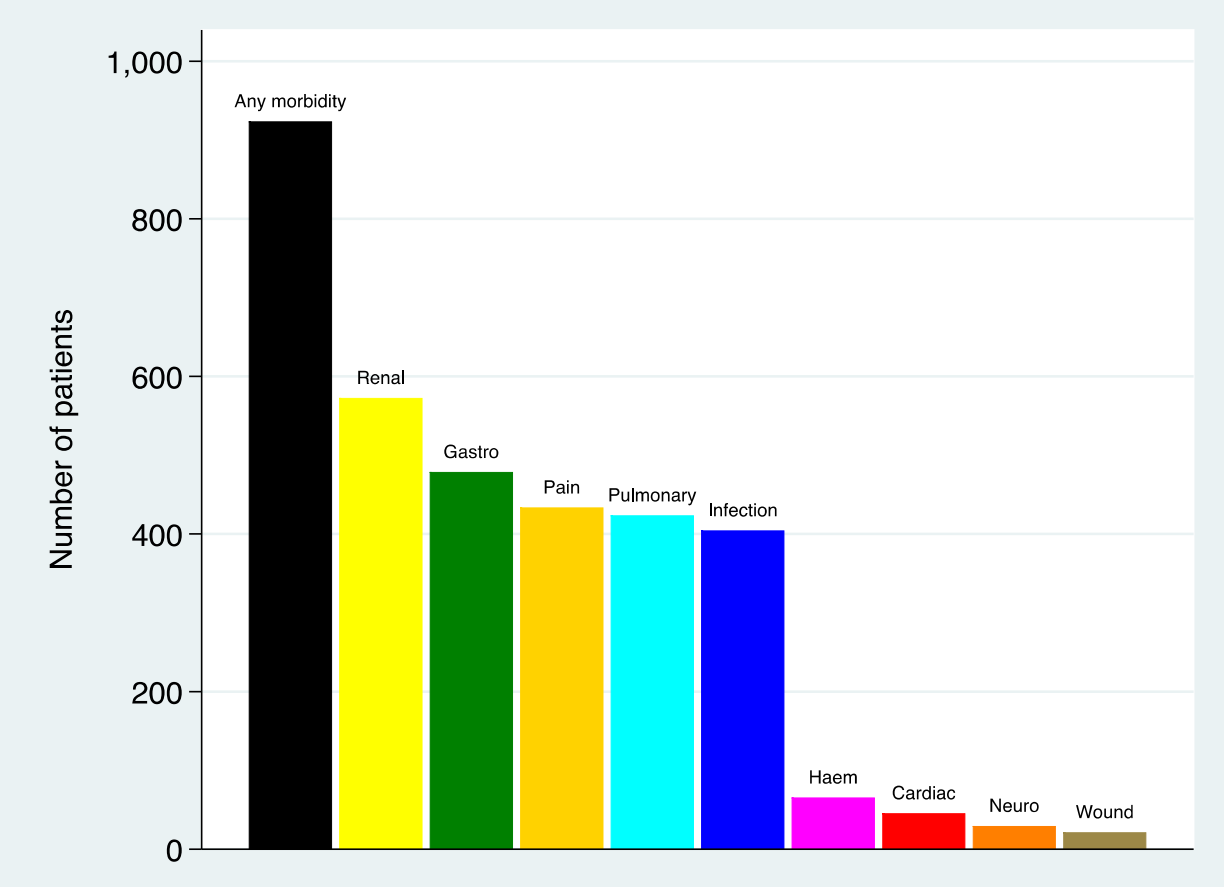

Figure 26: Prevalence of day 3 morbidity in overall cohort (n=1362)

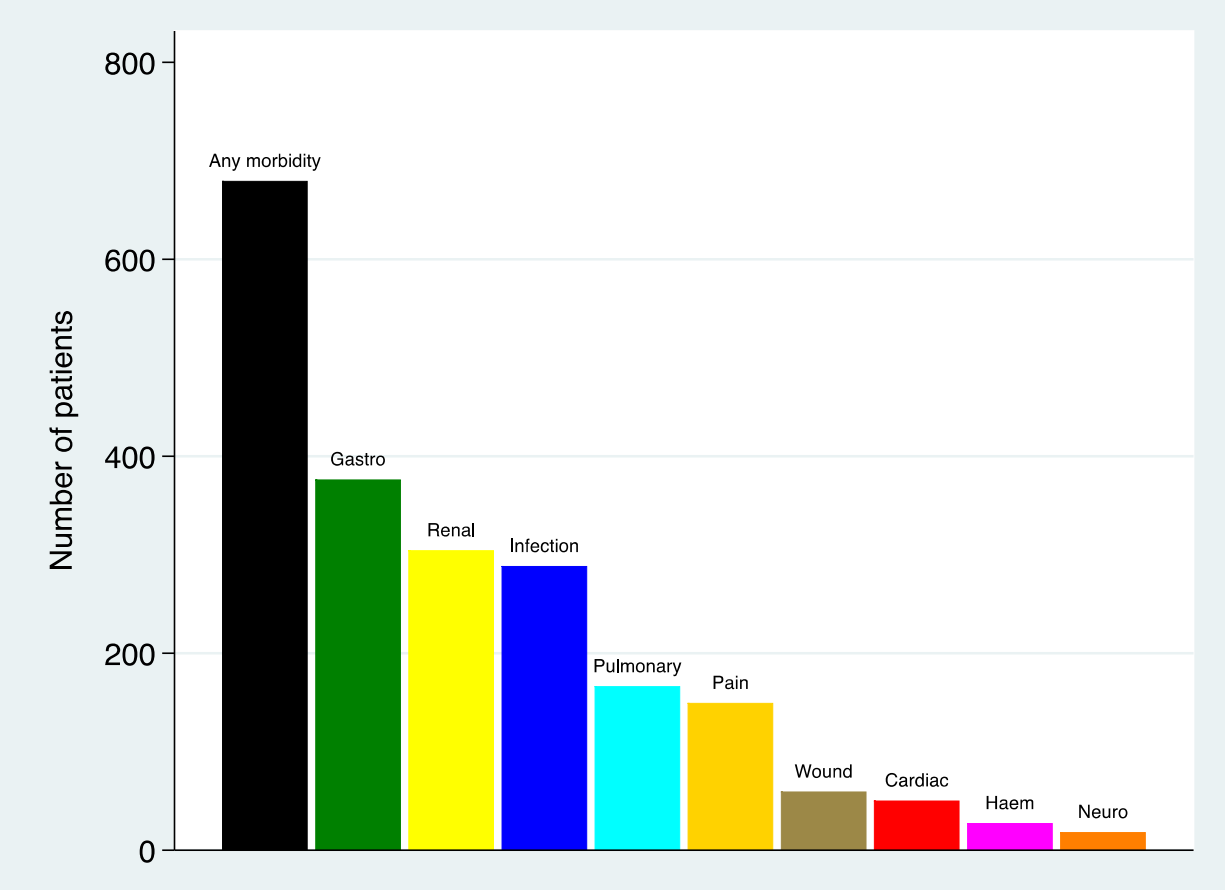

Figure 27: Prevalence of day 5 morbidity in overall cohort (n=1362)

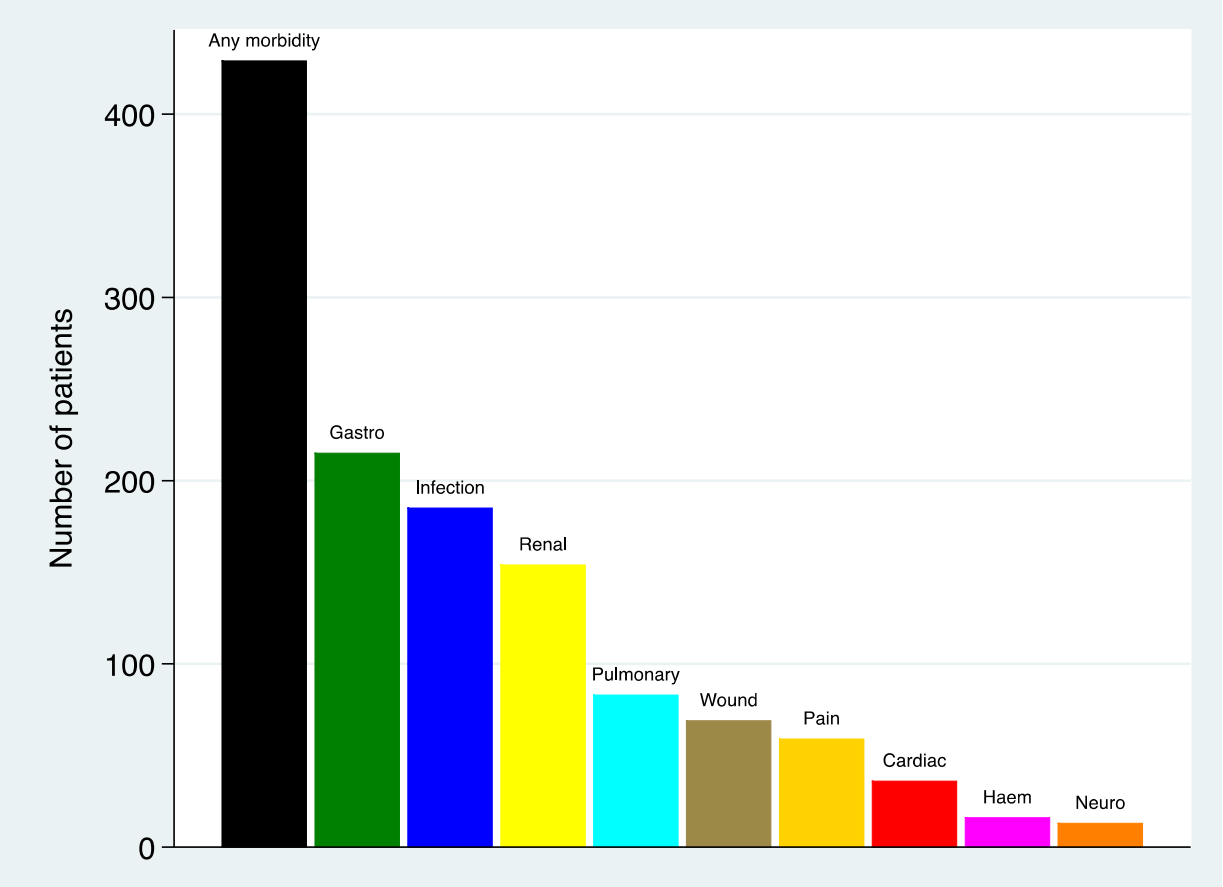

Figure 28: Prevalence of day 8 morbidity in overall cohort (n=1362)

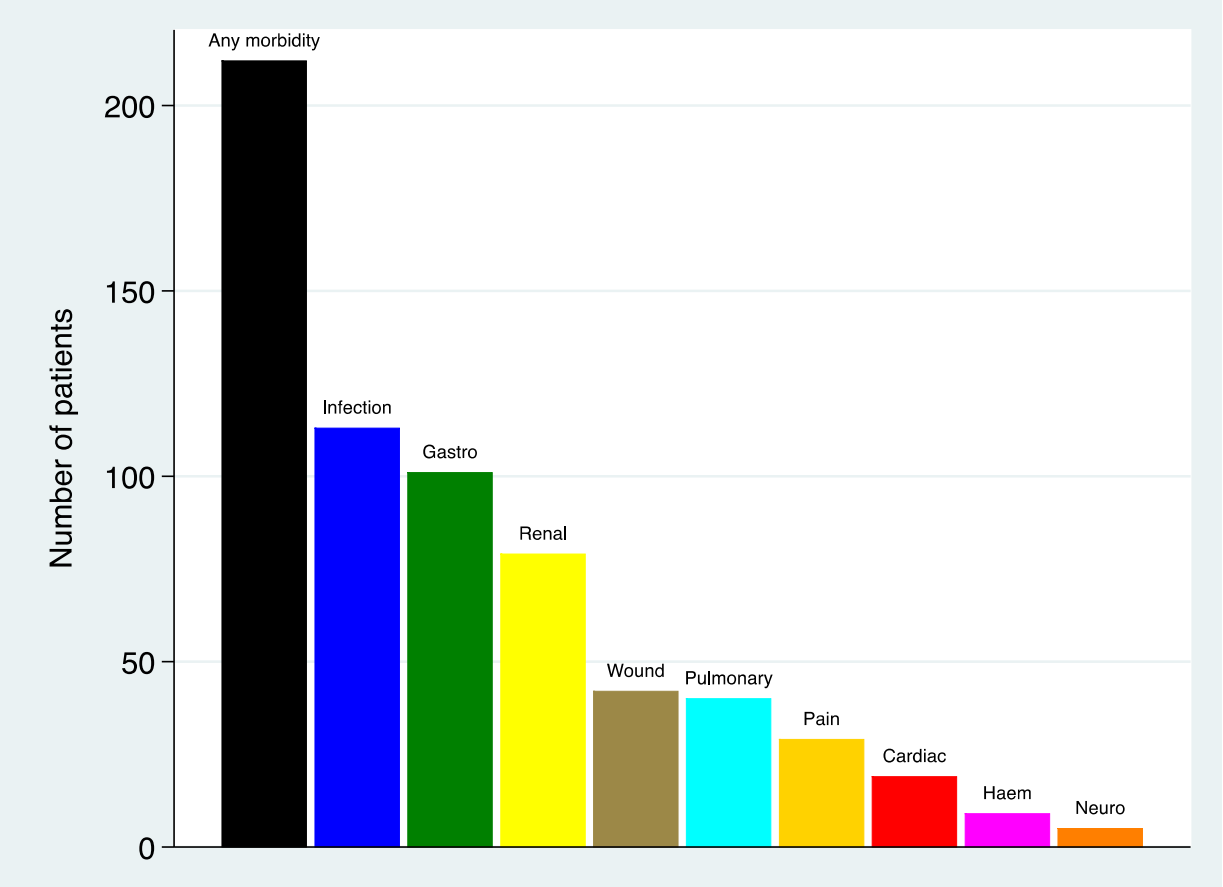

Figure 29: Prevalence of day 15 morbidity in overall cohort

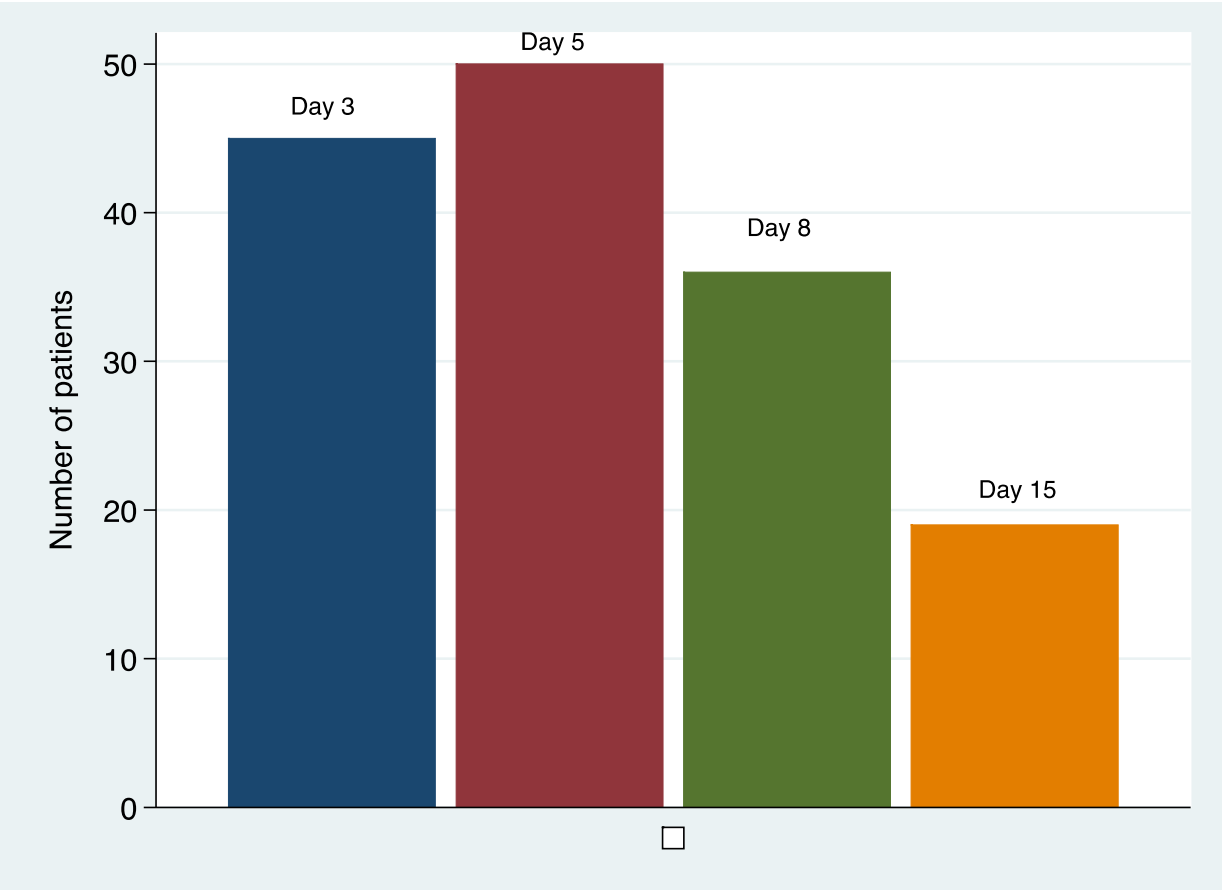

Figure 30: Prevalence of cardiac morbidity in overall cohort (n=1362)

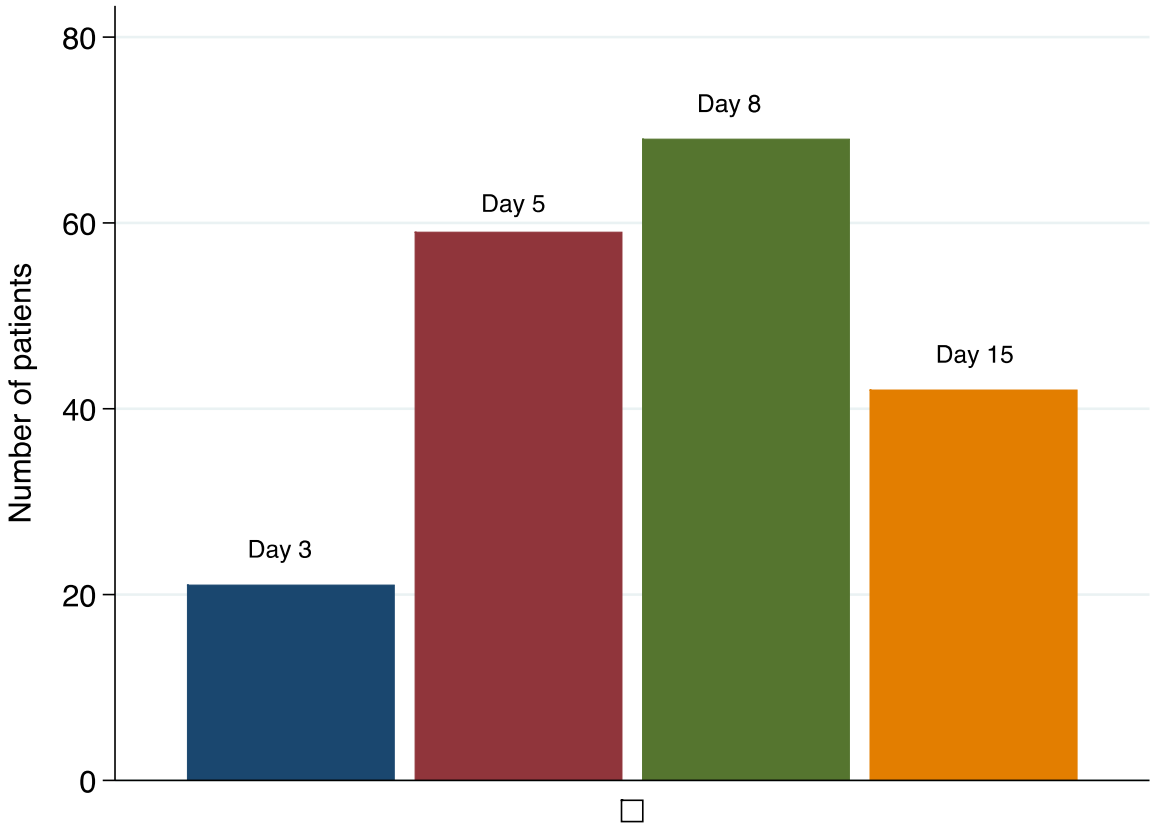

Figure 31: Prevalence of wound morbidity in overall cohort (n=1362)

4.3.6 Postoperative survival by perioperative risk category

Long term survival according to different methods of categorising perioperative risk (additive POSSUM, POSSUM physiology, P-POSSUM, Surgical Risk Scale and ASA-PSS) are depicted using Kaplan-Meier curves in Figures 32 to 36. Log-rank statistics testing for differences in survival were significant for all methods of classifying risk. However, visual examination of the Kaplan-Meier curves shows that there is overlap in the confidence intervals between some risk categories. With the additive POSSUM score,

there is overlap between Categories One, Two and Three both in the early and later analysis stages (figure 32). However, pair-wise comparisons using log-rank statistics show a significant difference in survival between all of the individual classes (Category Two versus Three, p<0.001; all other comparisons, p<0.0001). Similar overlap between Categories One, Two and Three is seen with the POSSUM physiology score (figure 33). With the SRS, overlap is seen between Categories One and Two, and Categories Three and Four (figure 35). Finally, with the ASA-PSS, there is overlap in confidence intervals between all four categories (figure 36).

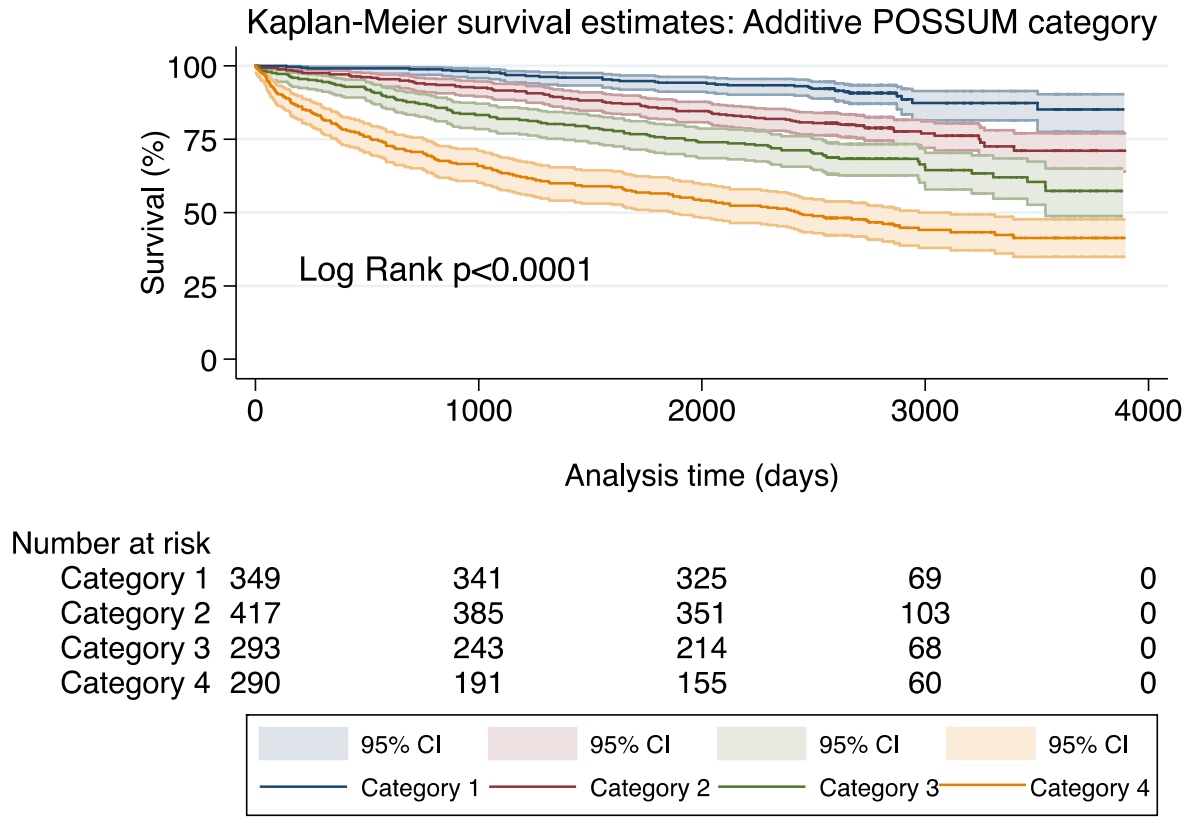

Figure 32: Univariate analysis of long-term survival by Additive POSSUM category

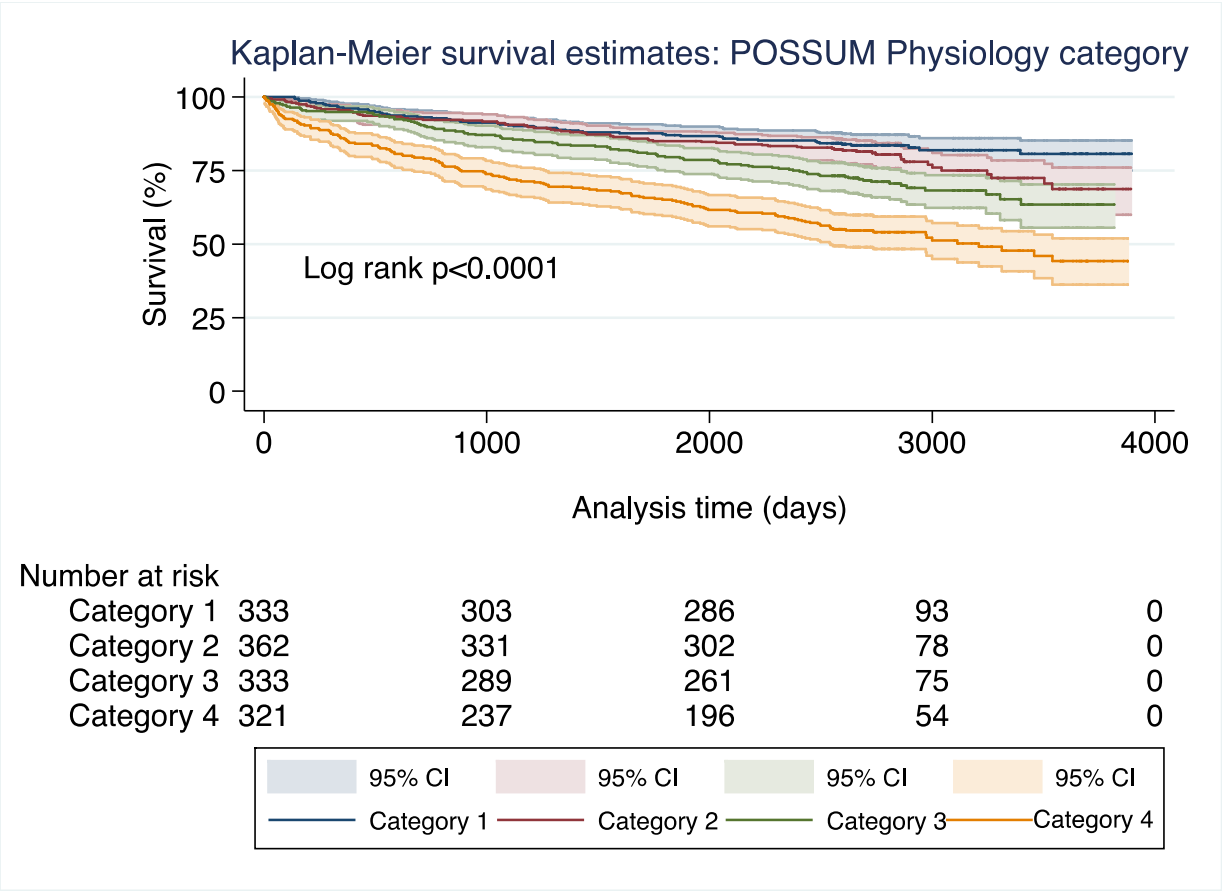

Figure 33: Univariate analysis of long-term survival by POSSUM physiology category

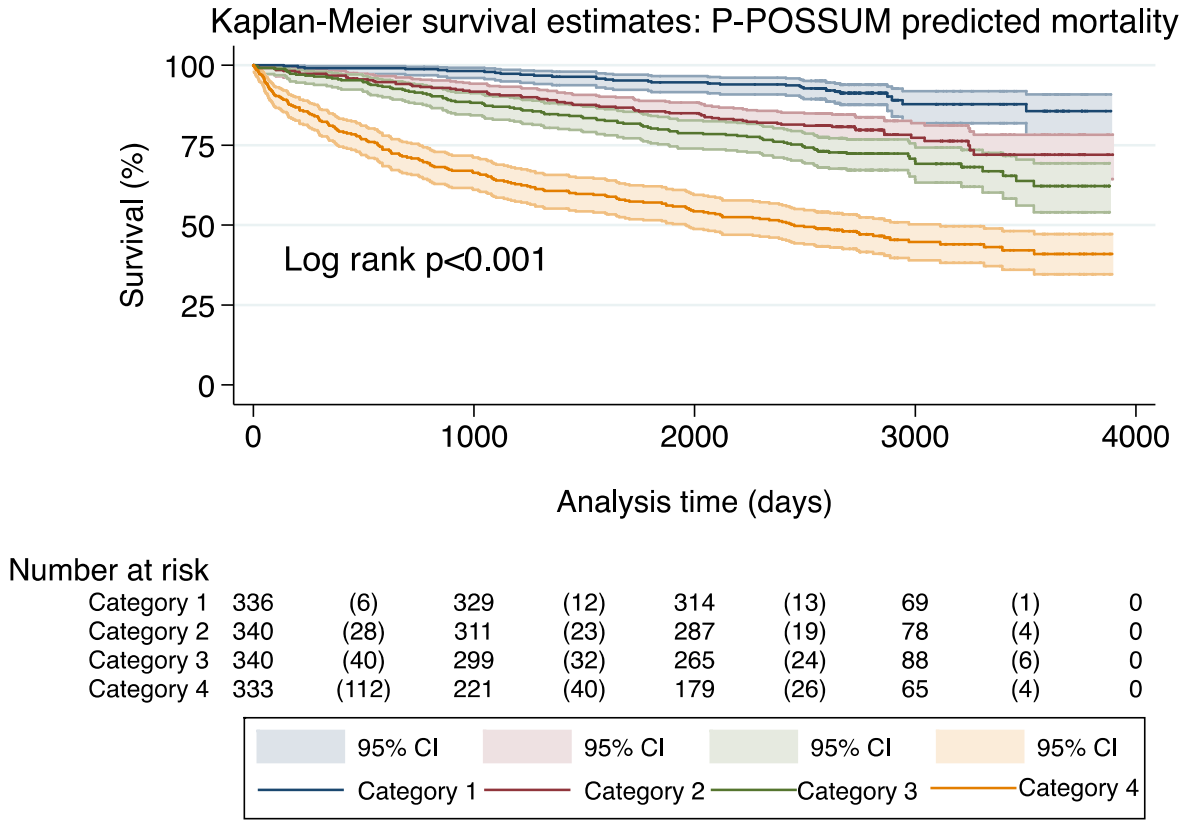

Figure 34: Univariate analysis of long-term survival by P-POSSUM predicted mortality category

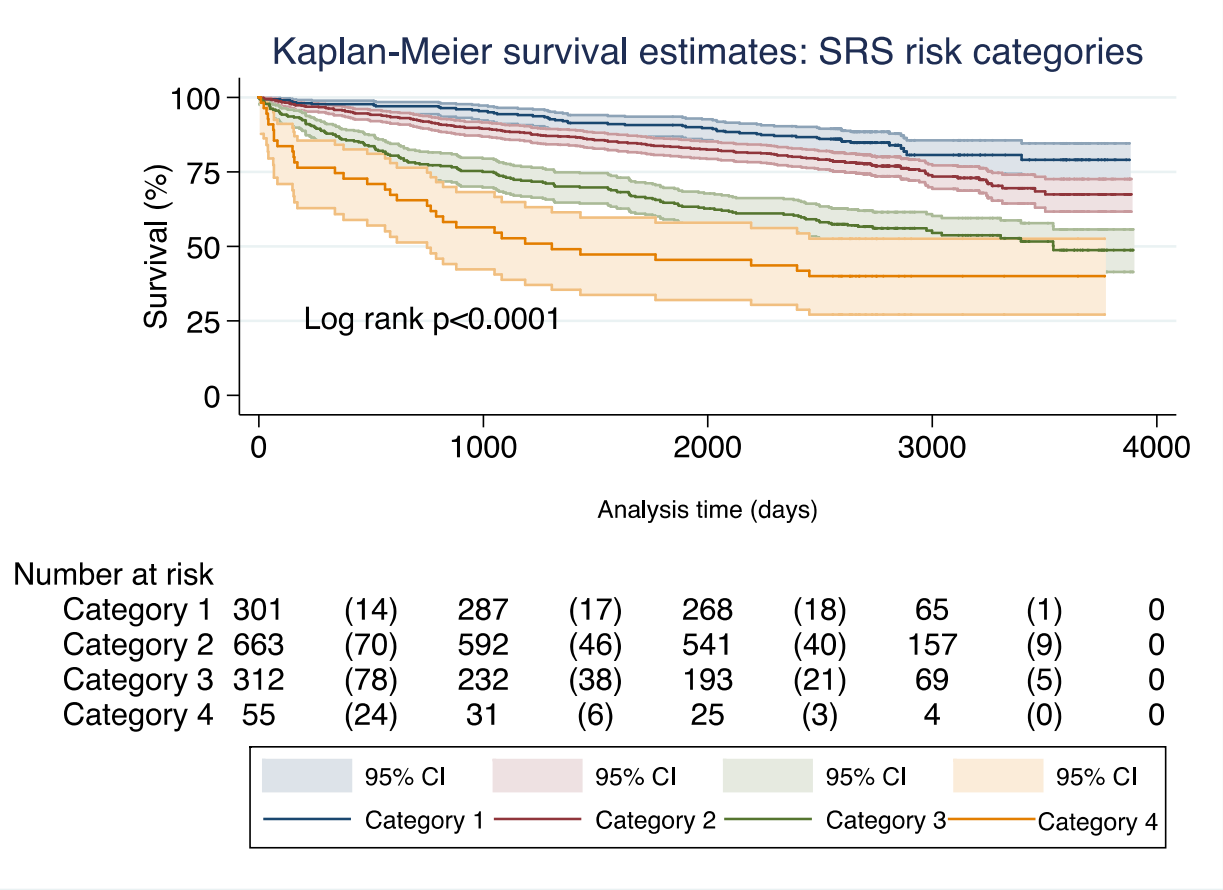

Figure 35: Univariate analysis of long-term survival by Surgical Risk Scale category

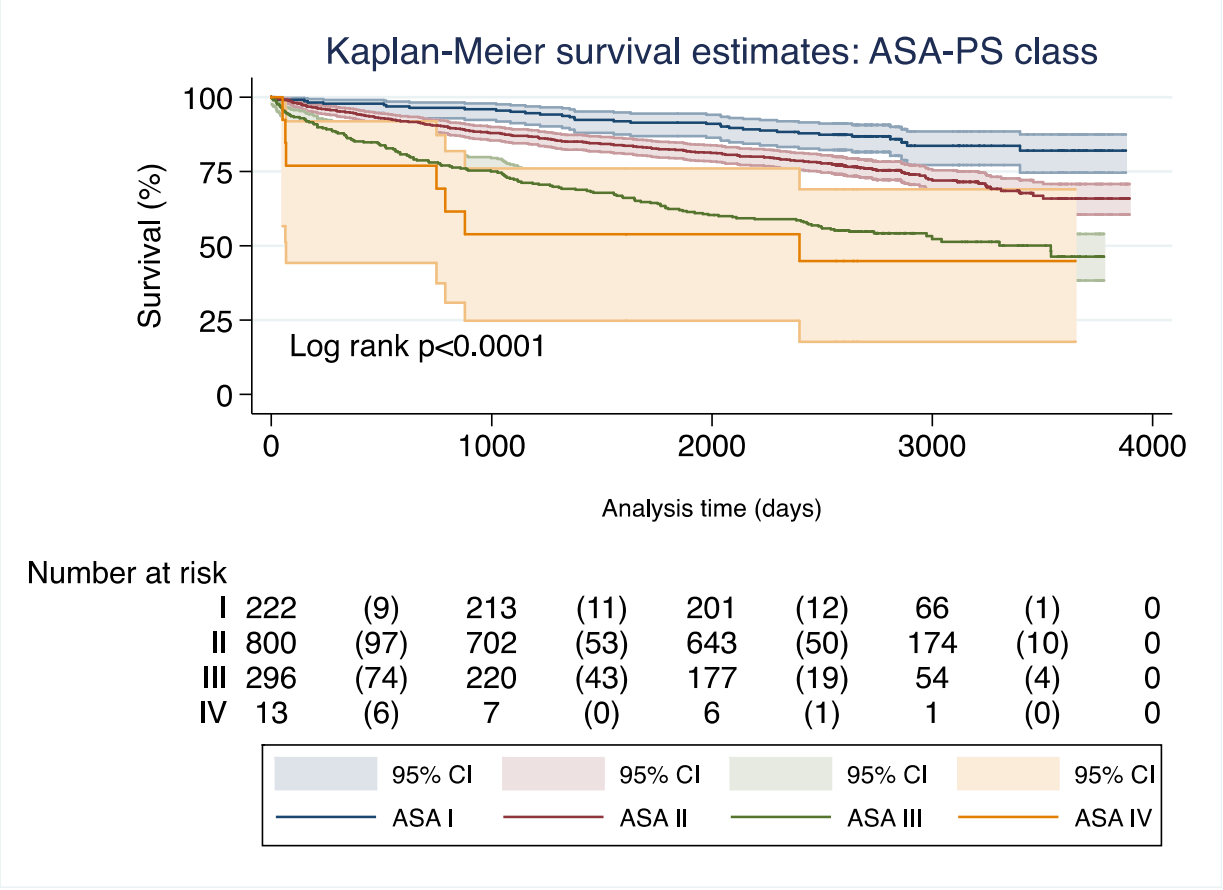

Figure 36 Univariate analysis of long-term survival by ASA-PS class

4.3.7 Relationship between POMS defined morbidity and postoperative mortality

Comparisons of patients with and without POMS defined morbidity overall, and at different time points are tabulated in Tables 31 to 35. On univariate analysis, the occurrence of postoperative morbidity (either any POMS defined morbidity, or the occurrence of POMS defined morbidity on any of days 3, 5, 8 or 15 postoperatively) is associated with significantly higher mortality both at one year and five years (p'<0.01

for both one-year and five-year mortality with Bonferroni correction for five comparisons, as five different measures of morbidity were used).

	30 day (n=1362)	Inpatient (n=1362)	1 year (n=1347)	5 year (n=1339)
Any morbidity (n=994; 73.0%)	1.3	1.9	8.6	24.5
No morbidity (n=368; 27.0%)	0.5	0.5	1.9	9.0
p value	NS	NS	**<0.001**	**<0.001**
p' value	NS	NS	**<0.01**	**<0.01**

Table 31: Mortality according to development of any POMS defined morbidity

	30 day (n=1362)	Inpatient (n=1362)	1 year (n=1347)	5 year (n=1339)
Day 3 POMS + (n=923; 67.8%)	1.4	2.1	9.2	26.4
Day 3 POMS – (n=439; 32.2%)	0.5	0.5	1.8	8.7
p value	NS	**<0.03**	**<0.001**	**<0.001**
p' value	NS	NS	**<0.01**	**<0.01**

Table 32: Mortality according to POMS status on Postoperative Day 3

	30 day (n=1362)	Inpatient (n=1362)	1 year (n=1347)	5 year (n=1339)
Day 5 POMS+ (n=679; 49.9%)	1.6	2.4	10.5	29.2
Day 5 POMS – (n=683; 50.2%)	0.6	0.7	3.2	12.3
p value	NS	**<0.02**	**<0.001**	**<0.001**
p' value	NS	NS	**<0.01**	**<0.01**

Table 33: Mortality according to POMS status on Postoperative Day 5

	30 day (n=1362)	Inpatient (n=1362)	1 year (n=1347)	5 year (n=1339)
Day 8 POMS + (n=429; 31.5%)	2.1	4.0	13.2	34.5
Day 8 POMS – (n=933; 68.5%)	0.6	0.4	3.9	14.3
p value	**<0.02**	**<0.001**	**<0.001**	**<0.001**
p' value	**NS**	**<0.01**	**<0.01**	**<0.01**

Table 34: Mortality according to POMS status on Postoperative Day 8

	30 day (n=1362)	Inpatient (n=1362)	1 year (n=1347)	5 year (n=1339)
Day 15 POMS+ (n=212; 15.6%)	2.8	7.1	19.6	43.5
Day 15 POMS- (n=1150; 84.4%)	0.8	0.5	4.5	16.5
p value	**p<0.01**	**<0.001**	**<0.001**	**<0.001**
p' value	**NS**	**<0.01**	**<0.01**	**<0.01**

Table 35: Mortality according to POMS status on Postoperative Day 15

4.3.8 Long term survival by postoperative morbidity status

Kaplan Meier survival curves comparing patients with and without postoperative morbidity are shown in figures 37-41.

Log-rank statistics show a significant difference in long-term survival between patients with and without morbidity overall, and on each of Days 3, 5, 8 and 15.

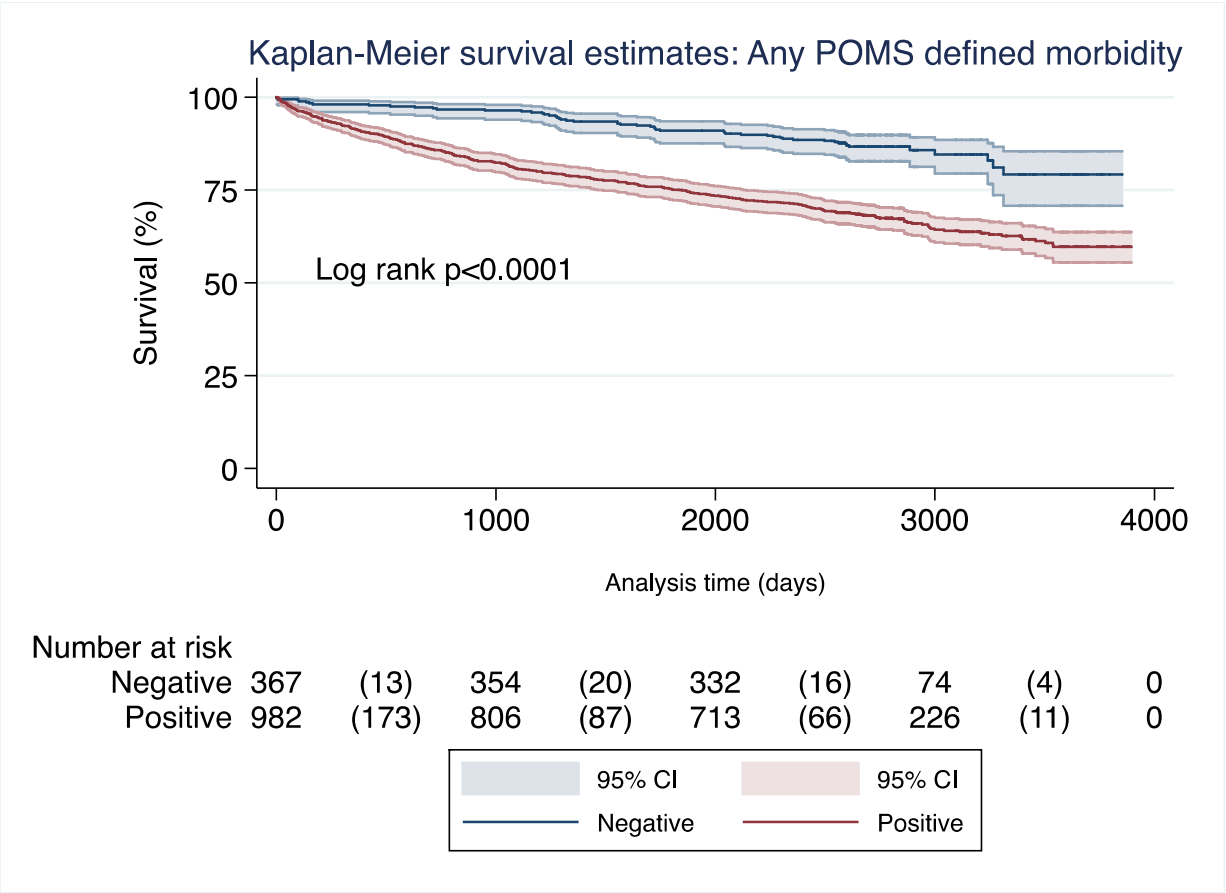

Figure 37: Univariate analysis of survival: any POMS defined inpatient morbidity

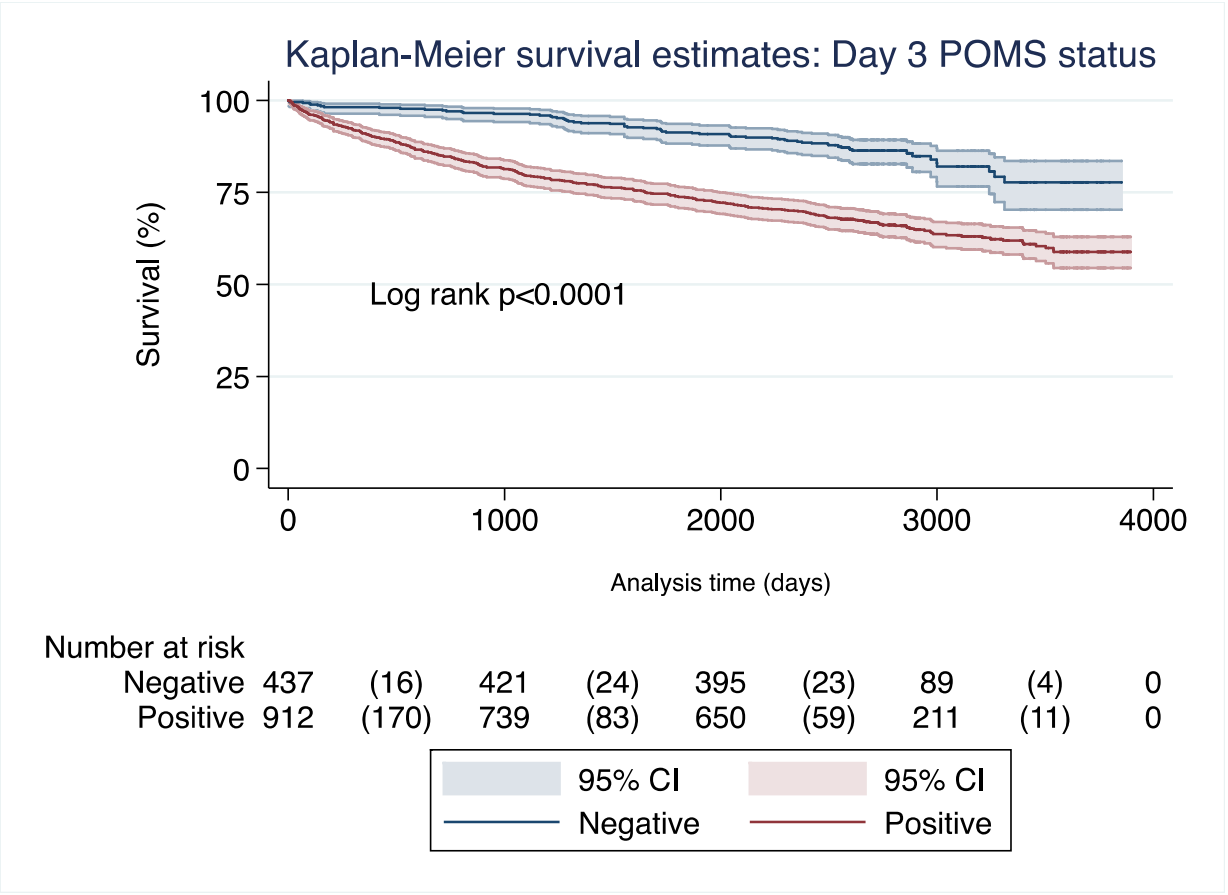

Figure 38: Univariate analysis of survival: Day 3 POMS status

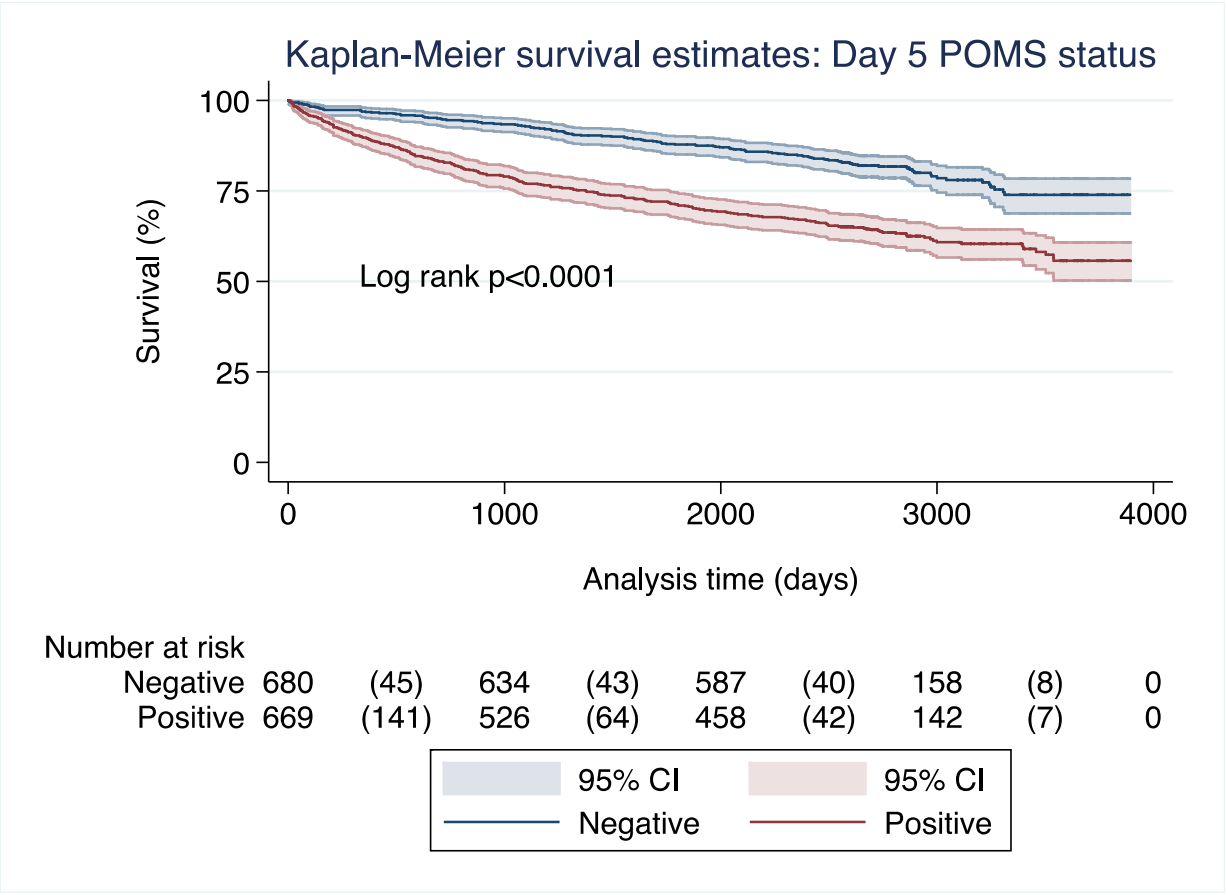

Figure 39: Univariate analysis of survival: Day 5 POMS status

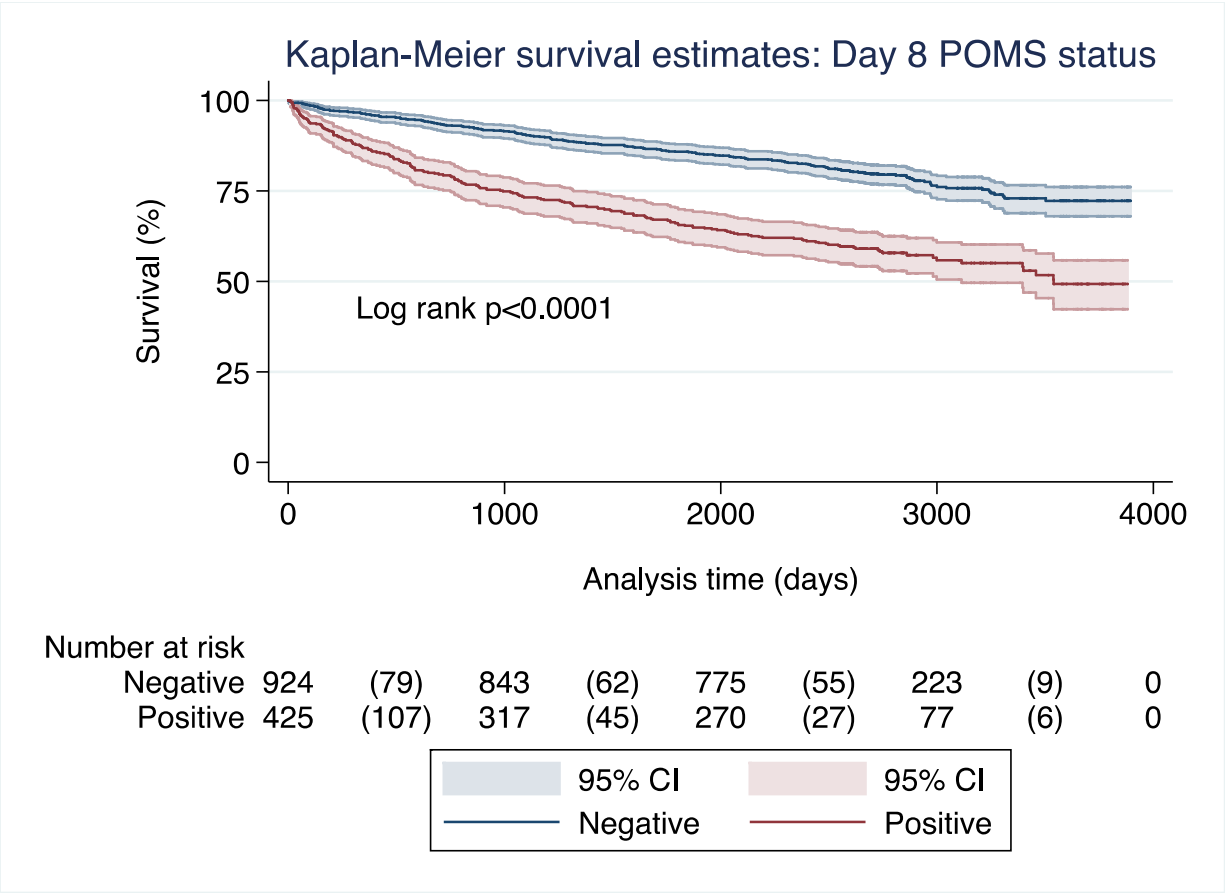

Figure 40: Univariate analysis of survival: Day 8 POMS status

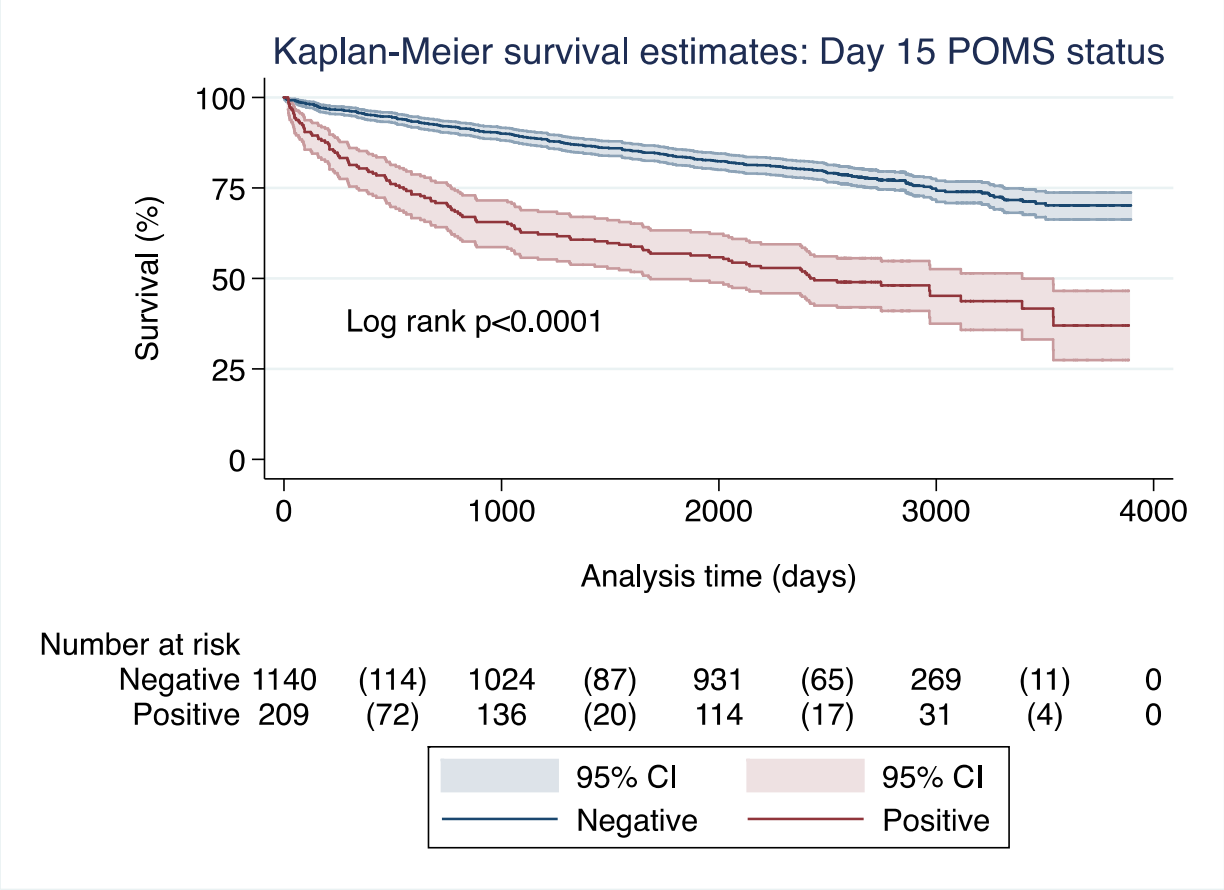

Figure 41: Univariate analysis of survival: Day 15 POMS status

4.3.9 Long term survival and postoperative morbidity domains

The relationship between each domain and long term survival was plotted using Kaplan Meier survival curves and log-rank testing the difference. The survival curve for pulmonary morbidity is shown in figure 42. Log rank comparisons of long term survival between patients who did and did not develop pulmonary, renal, infectious, gastrointestinal, neurological, cardiac and pain morbidity all showed statistically significant differences ($p<0.0001$). There was also a significant difference in survival

between patients who were positive and negative for haematological morbidity, with a p value of 0.0023. There was no difference in long term outcome based on the development of wound morbidity (p=0.4187)

Kaplan Meier curves were also plotted (not shown) with the observations deleted for patients who died either within thirty days or while still in hospital after surgery (n=23; 1.7%); these also showed a significant difference in long-term survival based on patients overall POMS morbidity status, and their status on days 3, 5, 8 and 15 (log rank $p<0.0001$ for any POMS morbidity and for pulmonary, renal, infection, gastrointestinal, cardiac, neurological and pain domains; p=0.038 for haematology and p=0.5004 for wound).

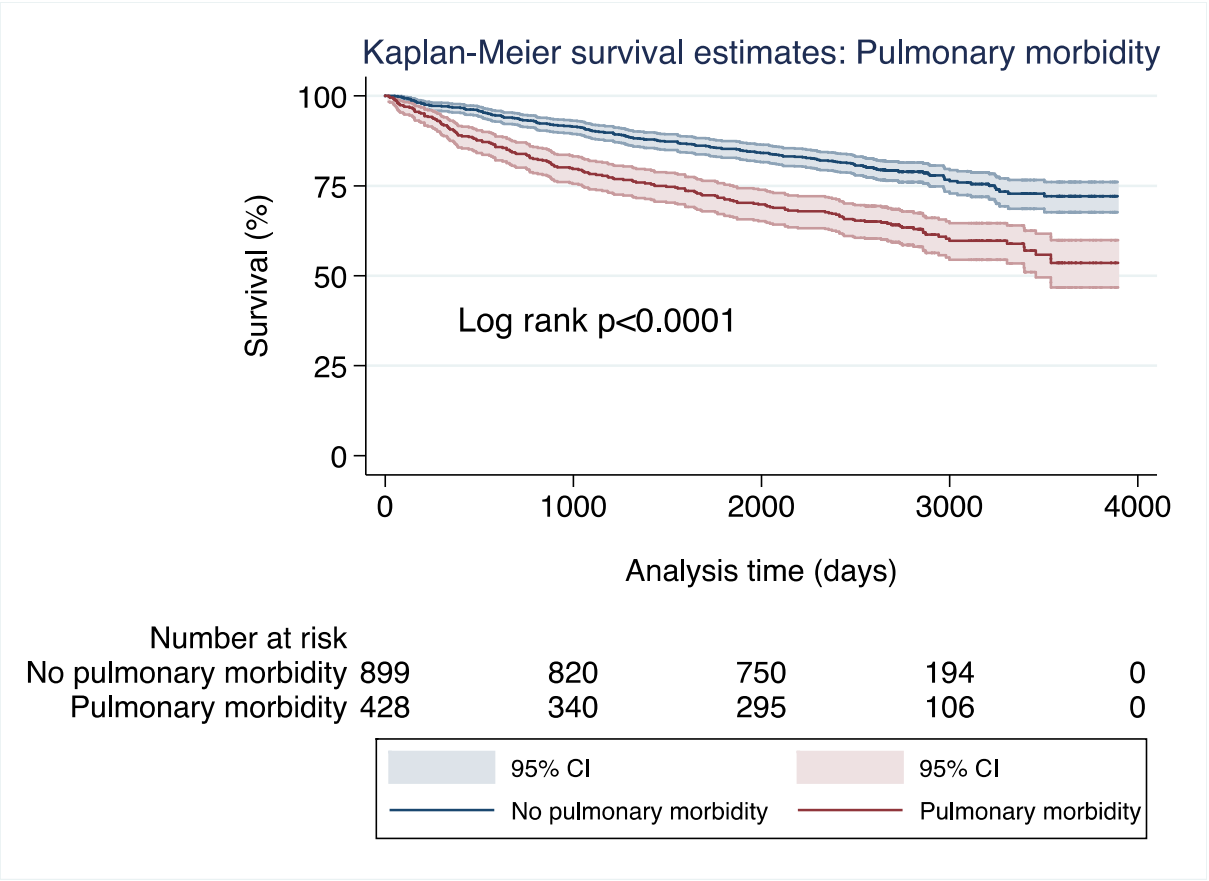

Figure 42: Long term survival by incidence of pulmonary morbidity

4.3.10 Postoperative length of stay and Long term survival by duration of postoperative morbidity (Final Morbidity Day)

Table 36 shows postoperative length of stay data based on the duration of postoperative morbidity.

Postoperative length of stay	No morbidity (n=368)	FMD 3 (n=266)	FMD 5 (n=278)	FMD 8 (n=238)	FMD 15 (n=212)
Mean (sd)	5.88 (3.04)	7.95 (4.43)	10.36 (10.70)	14.69 (8.77)	33.54 (25.27)
Median (IQR)	6 (4-7)	7 (6-9)	8 (7-11)	13 (11-15)	25 (19.5-36)

Table 36: Relationship between Final Morbidity Day (FMD) and postoperative length of stay

Figure 43 depicts the univariate analysis of long-term survival by duration of postoperative morbidity, as defined by the 'Final Morbidity Day' (FMD). Again, although the overall log-rank statistic indicates a significant difference between survival by the FMD category, visual inspection of the graph shows overlap in confidence intervals. Pair-wise analysis of FMD status reveals no significant difference in long-term survival between FMD 3 and 5 patients (p=0.0924) and between FMD5 and 8 patients (p=0.1244). All other pair-wise comparisons were statistically significant.

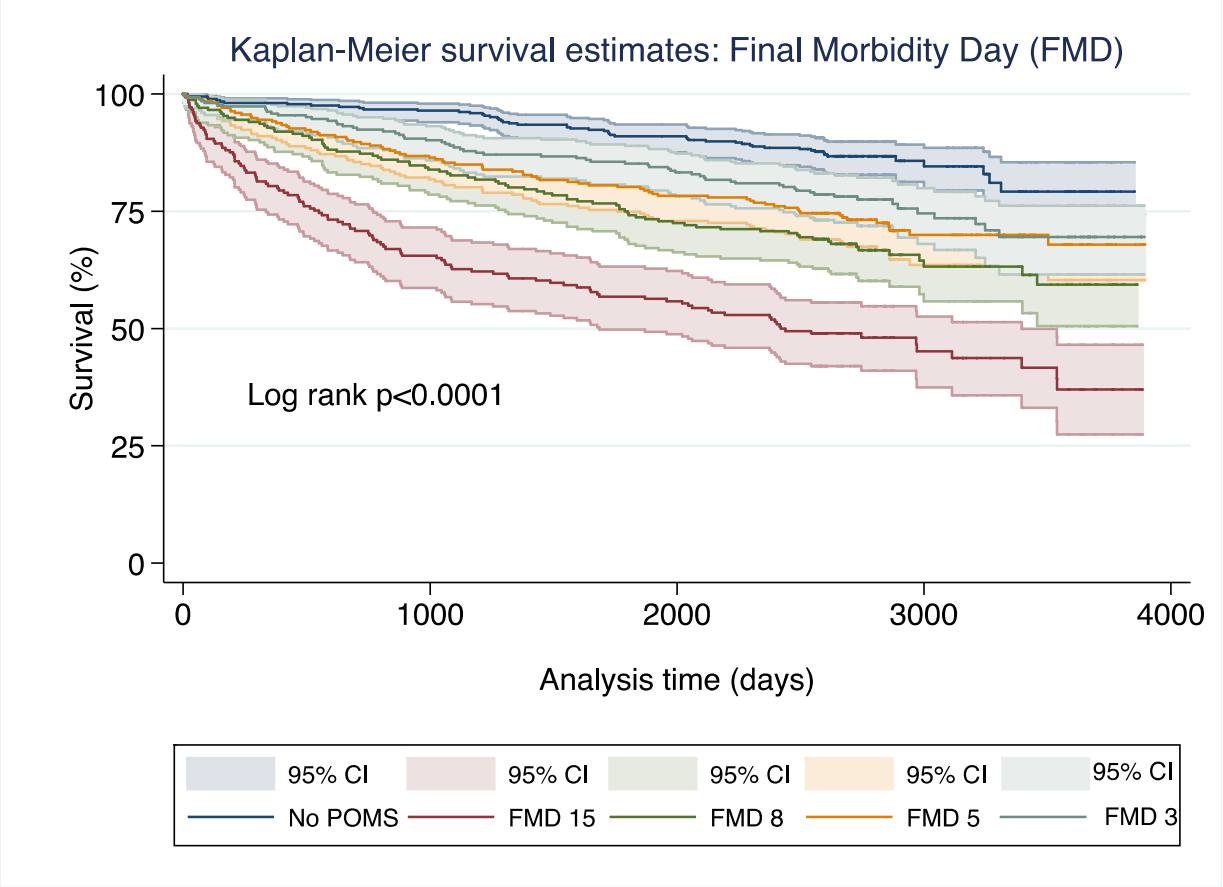

Figure 43: Univariate analysis of survival by final morbidity day (FMD)

4.3.11 Relationship between POMS status on POD5 and postoperative length of stay

Mean length of stay in patients who survived hospital admission and who were POMS positive on Day 5 postoperatively was significantly higher than those who were POMS negative (See Table 37).

	Median LOS	Mean LOS (sd)	p value
Overall	9	12.69 (14.55)	
Day 5 POMS-	6	7.64 (6.09)	p< 0.0001
Day 5 POMS+	13	17.87 (18.37)	

Table 37 Postoperative length of stay (inpatient survivors only)

4.3.12 Relationship between POMS status on Day 5 and postoperative mortality

Inpatient mortality was significantly higher in patients who had any of pulmonary, gastrointestinal or neurological morbidity on day 5 compared with patients who did not (see Table 38). Inpatient and thirty-day mortality were both significantly higher in patients who had renal morbidity on day 5 postoperatively (see Table 38 and 39). Thirty day mortality was significantly higher in the patient group with haematological morbidity on day 5. The significance levels of all these results were corrected for multiple comparisons.

POMS status on day 5 was predictive of one year mortality in all domains with the exception of wound and infectious morbidity (see table 40). Similarly at 5 years, Day 5 POMS status differentiated between survivors and non-survivors, with the exception of wound, infectious and cardiac morbidity (p=0.05 for cardiac morbidity). (See Table 41)

Morbidity domain	Number of patients POMS +	Mortality if negative; (%)	Mortality if positive; (%)	p' value (Bonferroni correction for 10 analyses)
Any morbidity	679	0.7	2.4	NS (p'=0.15)
Pulmonary	166	0.8	6.6	**<0.01**
Infection	288	1.3	2.4	NS (p'=1.0)
Renal	296	1.0	3.6	**<0.05**
Gastrointestinal	376	0.8	3.5	**<0.01**
Cardiac	50	1.4	6.0	NS (p'=0.09)
Neurological	18	1.3	22.2	**<0.01**
Wound	59	1.5	1.7	NS (p'=1.0)
Haematology	27	1.4	7.4	NS (p'=0.12)
Pain	149	1.2	4.0	NS (p'=0.09)

Table 388 Univariate analyses of Day 5 POMS defined morbidity and in-hospital mortality

Morbidity domain	Number of patients POMS +	Mortality if negative; n (%)	Mortality if positive; n (%)	p' value (Bonferroni correction for 10 analyses)
Any morbidity	679	0.6	1.6	NS (p'=0.67)
Pulmonary	166	0.8	3.6	NS (p'=0.059)
Infection	288	1.1	1.4	NS (p'=1.0)
Renal	296	0.7	2.6	**<0.05**
Gastrointestinal	376	0.7	2.1	NS (p'=0.25)
Cardiac	50	1.0	4.0	NS (p'=0.45)
Neurological	18	1.0	5.6	NS (p'=0.68)
Wound	59	1.2	0.0	NS (p'=1.0)
Haematology	27	1.0	7.4	**<0.05**
Pain	149	1.0	2.0	NS (p'=1.0)

Table 39 Univariate analyses of Day 5 POMS defined morbidity and 30 day mortality

Morbidity domain	Number of patients POMS +	Mortality if negative (%)	Mortality if positive (%)	p' value (Bonferroni correction for 10 analyses)
Any morbidity	679	3.2	10.5	**<0.01**
Pulmonary	166	5.1	19.5	**<0.01**
Infection	288	5.9	10.2	NS (p'=0.12)
Renal	296	5.0	13.3	**<0.01**
Gastrointestinal	376	4.5	13.0	**<0.01**
Cardiac	50	6.4	18.4	**<0.05**
Neurological	18	6.5	33.3	**<0.01**
Wound	59	7.0	3.5	NS (p'=1.0)
Haematology	27	6.4	29.6	**<0.01**
Pain	149	5.3	19.2	**<0.01**

Table 39 Univariate analyses of Day 5 POMS defined morbidity and one year mortality

Morbidity domain	Number of patients POMS +	Mortality if negative (%)	Mortality if positive (%)	p' value (Bonferroni correction for 10 analyses)
Any morbidity	679	12.30	29.22	**<0.01**
Pulmonary	166	17.94	40.49	**<0.01**
Infection	288	19.11	26.60	NS (p'=0.06)
Renal	296	16.87	34.12	**<0.01**
Gastrointestinal	376	15.64	34.06	**<0.01**
Cardiac	50	20.08	36.73	NS (p'=0.05)
Neurological	18	20.21	55.56	**<0.01**
Wound	59	20.92	15.52	NS (p'=1.0)
Haematology	27	20.20	40.44	**<0.05**
Pain	149	19.01	34.48	**<0.01**

Table 40 Univariate analyses of Day 5 POMS defined morbidity and five year mortality

4.4 Discussion

4.4.1 Summary of findings

There were no significant differences in mortality at hospital discharge, 30 days, 1 year or 5 years between the two cohorts of SOuRCe patients.

Shorter-term (30 day and inpatient) mortality varied between surgical specialities: it was significantly worse for vascular and general surgery than for orthopaedics or urology. There were also differences in mortality between surgical specialities at one and five years postoperatively.

There were differences in the prevalence and patterns of morbidity between surgical specialities. Overall, pulmonary, infection, gastrointestinal, renal and pain morbidity were more common than neurological, cardiac, wound or haematological complications. For most types of morbidity the prevalence was highest at the beginning of the postoperative period and dropped on successive postoperative days; however, cardiac complications peaked on day 5 and wound complications peaked on day 8.

Log-rank testing of Kaplan-Meier survival curves comparing outcome for patients in different categories of perioperative risk showed significant differences between risk groups for all the risk stratification tools which were tested (additive POSSUM, POSSUM physiology score, SRS and ASA-PS); however, there was overlap in confidence intervals for many of these comparisons.

Univariate analyses showed significant differences in mortality at one and five years between patients who did and did not develop any POMS defined morbidity at any point during the inpatient stay, and also based on POMS status on Days 3 and 5. POMS positive status on Days 8 or 15 was also associated with higher inpatient death rate, as well as higher one and five year mortality. Comparing survival for patients according to their POMS status on each of days 3, 5, 8 and 15, and overall POMS status showed significant differences in outcome. Survival distributions were also significantly different for patients who developed any of the individual types of morbidity recorded by the POMS, with the exception of wound morbidity.

Comparing patients with different durations of postoperative morbidity, as defined by the 'Final Morbidity Day', showed that FMD 15 was associated with significantly worse long-term survival when compared with the other classes.

Univariate analysis showed significant associations between each of pulmonary, renal, gastrointestinal, haematological, and neurological morbidity on Day 5 and shorter-term mortality. With the exception of infection and wound classes, complications in any of the individual POMS domains on Day 5 were all significantly associated with one year mortality. At five years, mortality was significantly higher for patients who had any type of morbidity on Day 5 apart from infectious, wound or cardiac.

4.4.2 Results in context

4.4.2.1 Surgical speciality and postoperative morbidity

The differences in patterns of morbidity between surgical specialities are plausible and consistent with the literature.[44;315] One would expect that there would be less morbidity in patients having orthopaedic surgery (which in the majority of cases were joint replacement procedures aimed at alleviating pain and improving quality of life), than in patients having cancer surgery or vascular surgery. Similarly, vascular patients, who by definition have atherosclerotic disease, had higher rates of cardiac and neurological morbidity, which again is consistent with other reports.

4.4.2.2 Surgical speciality and long-term mortality

Again, it is not surprising that there were significant differences in long-term mortality between surgical specialities on univariate analyses. For example, it is probable that patients undergoing vascular surgery for example, are more likely to have significant cardiac or other atherosclerotic disease than those undergoing urological surgery. However, the main implication of this confirmed difference in outcome between surgical specialities is that it suggests that surgical speciality should be included as a co-variate in multivariable analysis of long-term survival, even if the POSSUM score is used as a risk adjuster. POSSUM, while considering the urgency and severity of the surgical procedure, does not include speciality as a variable; furthermore, the original POSSUM development cohort consisted of general, urological and vascular patients (no orthopaedics) and therefore it is possible that surgical speciality may not have been

included in the original multivariable analysis which developed the POSSUM. Even if surgical speciality was included, it may not have been associated with differences in outcome as the population of patients was more homogeneous.[38] In the ACS-NSQIP, different risk adjustment models are constructed for different surgical specialities, as the most significant predictors vary between groups.[316;317]

4.4.2.3 Patterns of postoperative morbidity

With two exceptions, different types of morbidity were most common on Day 3 and the prevalence fell steadily after that. Cardiac morbidity peaked on Day 5. This might be considered surprising given that previous reports suggest that postoperative myocardial infarction (MI) is most likely to occur within the first three days of surgery.[318;319] There are likely to be two main reasons for this discrepancy. The first, is that the cardiac domain of the POMS records the occurrence of cardiovascular events or interventions in the previous 24 hours (so the incidence of cardiac morbidity is actually highest on day 4 compared with days 2, 7 and 14). The second is that at least one of the diagnostic criteria for cardiac morbidity (hypotension requiring pharmacological therapy or >200ml/hour of fluid therapy) may have non-cardiac aetiology (for example sepsis or bleeding).

Wound morbidity peaked on Day 8. Although there is a great deal in the literature about the incidence and risk factors for wound complications,[320-322] I was unable to find any studies which specifically reported the pattern of timing of surgical wound morbidity.

4.4.2.4 Type and timing of morbidity and postoperative mortality

With the exception of wound complications, the occurrence of any of the other types of morbidity was associated with reduced long-term survival on univariate analysis. This is consistent with the ACS-NSQIP study examining the relationship between postoperative morbidity and long-term outcome.[151]

The occurrence of morbidity on any of the four postoperative days on which the POMS was recorded, was also associated with a reduction in long-term survival. There has been previous discussion about the validity of Day 3 POMS as a true reflection of morbidity, as some of the interventions on which POMS morbidity is defined (for example, administration of oxygen therapy or parenteral opioids) may under some circumstances be part of the normal package of postoperative care, rather than being evidence of harm.[56] Therefore, it might appear surprising that POMS on Day 3 was associated with reduced long-term survival, especially given that the majority of patients had POMS defined morbidity on Day 3 (67.8%). Explanations for this might include interaction with (confounding by) the operative type and severity or with preoperative risk. The magnitude of the risk which postoperative morbidity carries for long-term survival cannot be estimated on univariate analyses, and therefore multivariable analysis is required to estimate the magnitude of the hazard and whether this risk is still apparent when confounding factors are considered. This analysis is the subject of Chapter 5.

4.4.2.4 Duration of complications and long-term survival

Univariate analyses showed a significant difference in long-term survival between patients who remained in hospital with complications on Day 15 and those who recovered more quickly. The hypothesis that duration of illness is related to severity of illness and length of stay has face validity and is consistent with previous reports, both in surgery and in other settings (for example, critical care).[323;324]

4.4.2.5 Relationship between Day 5 POMS and other outcomes

In choosing a particular day on which to measure the POMS, one must bear in mind that changes in healthcare delivery over the past decade, and in particular the implementation of enhanced recovery pathways, have led to dramatic reductions in postoperative length of stay.[325] The median postoperative hospital length of stay for patients in this study (between 2001 and 2005) was 9 days, and for orthopaedic patients it was 8 days; current SOuRCe data (not within this thesis) show that median length of stay after primary joint replacement at UCLH is now 6 days. Hospital Episode Statistics data reveal that the national median hospital length of stay for primary hip and knee replacement is now just four days, for colectomy is 7.6 days, and for rectal excision is 9 days. Therefore, while the presence of POMS defined morbidity on Day 8 or 15 may be indicative of the most serious complications, if either of these were to be used as a single measure of postoperative outcome, it is possible that a low incidence may mean larger than feasible sample sizes are necessary for clinical trials. It is also

possible that significant morbidity (with long-term implications) might be missed in patients who have been discharged home within seven days.

When considering an outcome measure for clinical effectiveness studies in general, one must give consideration to the incidence (for the purposes of power calculations) and the clinical relevance of the proposed measure. Validation of outcome measures requires (amongst other things) establishing face and content validity ('on the face of it', does the instrument measure the relevant outcome and does it do so comprehensively?), construct validity (does the measure fit with our understanding of the outcome and the implications of the outcome?). The POMS as a whole has already been validated against these and other criteria;[56] Day 5 POMS on its own also fits well with these criteria. Nearly 50% of patients in this study had POMS defined morbidity on Day 5. There were significant associations on univariate analysis between Day 5 morbidity and shorter term mortality (inpatient, thirty day or both) for five out of the nine POMS domains. The Day 5 POMS positive group of patients also had a significantly longer length of hospital stay, and higher mortality at one and five years. Although measurement of POMS at four time points and recording the final morbidity day would be better descriptors of postoperative morbidity and its duration and severity, if a single point prevalence estimate of postoperative harm is required, then Day 5 POMS would be a valid choice.

4.4.3 Limitations of this study

The main limitation of this study, as with previous chapters, is lack of generalizability to cohorts which include patients having emergency surgery. The population is also unevenly distributed between surgical specialties – there are very few patients having vascular surgery, and the number of patients' undergoing orthopaedic surgery was disproportionately high; this is of importance, as these patients have been shown to have considerably lower morbidity than other patient groups.

4.5 Conclusions

1. The type and prevalence of morbidity varies with surgical speciality; postoperative mortality rates, both shorter and longer-term also varied between different surgical specialties. Postoperative outcomes were worse for patients undergoing vascular and general surgery compared with urology and orthopaedics.
2. Univariate analysis demonstrated differences in longer-term survival between patients in different risk groups as determined by POSSUM category.
3. There are significant relationships on univariate analyses between postoperative morbidity and longer-term mortality. The occurrence of morbidity at any stage postoperatively, as well as individual types of complication were all associated with reduced longer-term survival.
4. Increased duration of morbidity, as defined by the 'Final Morbidity Day' is associated with increased length of hospital stay and reduced longer-term survival on univariate analysis.
5. POMS defined morbidity on Day 5 after surgery is a valid measure of postoperative harm and would be a valid single outcome measure in clinical effectiveness studies.

Chapter Five: Survival after postoperative morbidity – a longitudinal prospective observational cohort study

5.1 Introduction and aims

This study explores the long-term outcome of patients who underwent major non-cardiac non-neurosurgical procedures at the Middlesex Hospital between 2001 and 2005. The cohort is the same as that discussed in Chapters 3 and 4.

The aims of this chapter were:

1. To build a Cox proportional hazards model for long-term survival of the SOuRCe cohort.
2. To determine whether postoperative morbidity is an independent risk factor for reduced long-term survival

5.2 Methods

5.2.1 Overview

This was an observational cohort (long-term follow-up) study. Perioperative patient data (demographics, the POSSUM predictor and the Post Operative Morbidity Survey, [POMS]) were linked with mortality data which was provided by the Medical Research

Information Service (NHS Information Centre). Multivariable analysis using Cox Proportional Hazards modelling was used to determine the relationship between perioperative risk and outcome, and long term survival.

5.2.2 Patients and data

The 1362 patients from the SOuRCe cohort were eligible for inclusion. Patient recruitment and selection is described in Chapter 3. Patients were excluded if data linkage with the Medical Research Information Service (MRIS) database was not possible for any reason. Patients who died within 15 days of surgery were also excluded from the analysis as Day 15 POMS status was included as an explanatory variable. Additionally, patients who were lost to follow up for any reason were excluded. The dataset is as described in Chapters 3 and 4. Linkage between the SOuRCe dataset and the NHS Information Centre MRIS dataset was achieved using patient name, date of birth, postcode and NHS number. Time zero was operation date; right-censoring occurred on 1 March 2012.

5.2.3 Statistical Approach

5.2.3.1 Determination of variables to enter into analysis

Independent variables were selected on the basis of known association with short- and long term mortality and /or morbidity from univariate analysis and / or the published literature. Based on the results of Chapters 3 and 4, P-POSSUM predicted mortality was

chosen as a marker of perioperative risk; neither ASA-PS nor Surgical Risk Scale were included in models as P-POSSUM was shown to be a more highly discriminant predictor for this cohort. P-POSSUM predicted mortality was entered as a continuous variable. Surgical speciality and gender were included as indicator variables; gender was considered a potential confounder, as men are known to have a shorter life expectancy than women.[326]

Postoperative morbidity was entered into the analysis in a number of ways, all of which were based on POMS results. The occurrence of morbidity at any stage during the postoperative stay was entered as separate indicator variables for each of the nine POMS domains, with an additional variable for the occurrence of any POMS-defined morbidity during the inpatient episode. While POMS is a validated measure of postoperative harm, it has not been validated as a 'score' (that is, it is not valid to consider that the number of POMS domains scored is directly related to the severity of morbidity).[56] However, the relationship between the duration of morbidity (Final Morbidity Day, FMD), length of stay and long-term survival was established in Chapter 4; therefore FMD was entered for each patient as multiple dichotomous variables.

5.2.3.2 Interactions

Based on established knowledge, and the results of univariate analysis, interactions between the following variables were tested:

- Gastrointestinal morbidity and general surgery
- Renal morbidity and urological surgery
- cardiac morbidity and vascular surgery
- wound morbidity and orthopaedic surgery

5.2.3.3 Post-estimation analyses

After constructing an initial model, the Proportional Hazards (PH) assumption was tested using Schoenfeld's partial residuals and the result confirmed by visual examination of a log-minus-log survival plot. Both tests showed that the PH assumption was seriously violated. Plotting baseline cumulative hazard for patients with and without morbidity on Day 15 postoperatively demonstrated a 'step' in the cumulative hazard at approximately three years postoperatively in the 'Day 15 positive' group (see Figure 44). It is also intuitive that the impact (if any) of surgery and / or postoperative morbidity on long term outcome is likely to vary over time. Therefore, time was included as an interaction with duration of postoperative morbidity, in order to avoid the need to comply with the PH assumption.[327] The postoperative era was split into three periods: 0-365days, 366-1095 days, more than 1095 days. The full model included the interactions of each of these time categories with each of the final morbidity day categories.

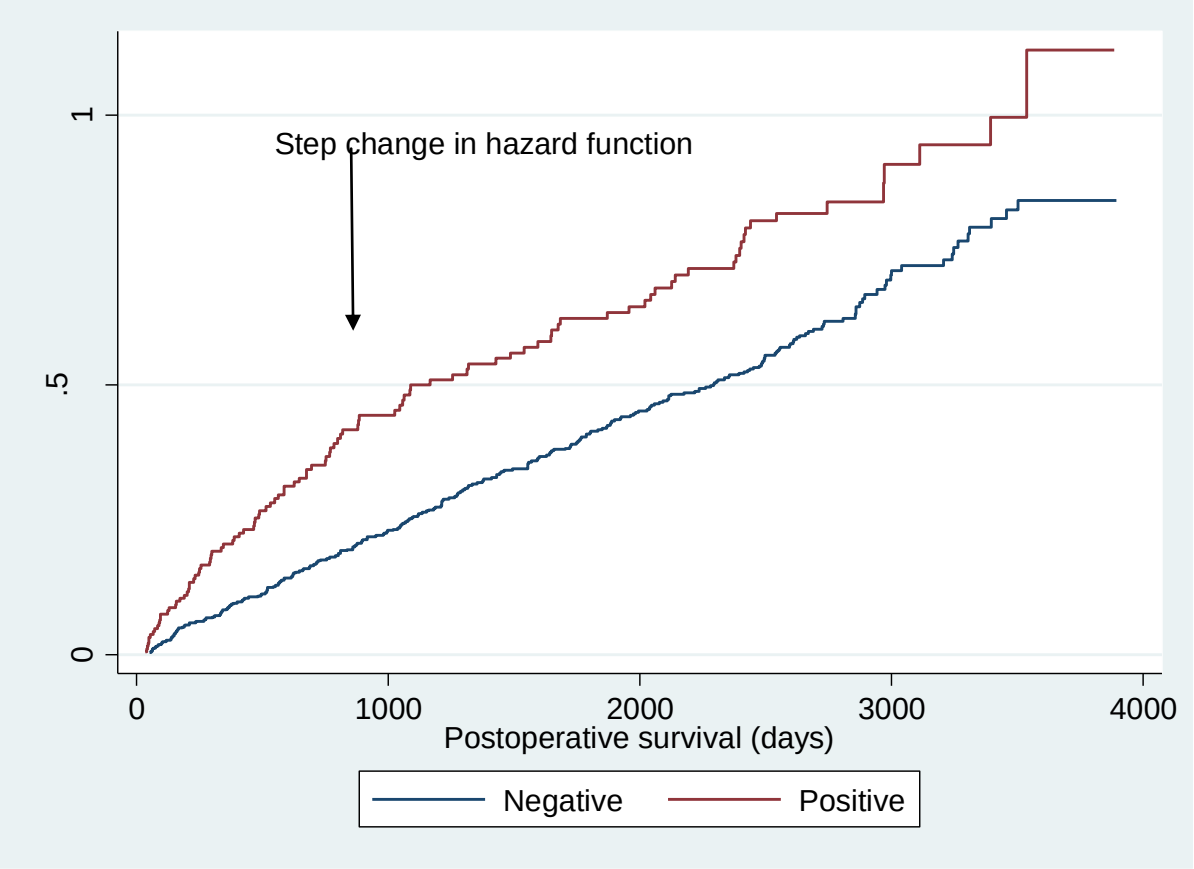

Figure 44: Baseline Cumulative Hazard graph for cohort based on Day 15 POMS status

5.2.3.4 Development of final model

The choice of variables to enter into the final model was based on significance testing; initially variables were dropped based on p values > 0.10 then p>0.05. Model fit was assessed with Likelihood Ratio testing against the full model.

All statistical analyses were conducted using Stata InterCooled (Release 12.1) software (StataCorp LP, College Station, Texas, USA).

5.3 Results

5.3.1 Description of cohort

Of the SOuRCe cohort of 1362 patients, data were analysed for 1342. The reasons for 20 exclusions were as follows: missing follow-up data (n=13) and death within 15 days of surgery (n=7). There were 383 deaths in the final group (28.1%); maximum duration of follow up was 10.66 years (3895 days) median follow up was 6.51 years (2375 days, IQR 2696 to 2899).

5.3.2 Model development

The full model consisted of 49 variables; these are listed, along with individual Hazard Ratios, standard errors, p values and 95% confidence intervals in Table 42. The Likelihood Ratio test was associated with a Chi Squared of 316.99, giving a p value of <0.0001, thus demonstrating good fit. The variables which are listed as 'omitted' were excluded from the analysis on the basis of collinearity with other variables. The full model was narrowed down to a final model of six variables, which is described in Table 43. The LR Chi squared for the final model was 262.49, giving a p value of <0.0001. On Likelihood Ratio testing with the original full model, the p value was 0.0507, indicating that the final model was not significantly different from the full model and therefore also had good fit.

Variable	Hazard ratio	Standard error	p value	95% C.I.
Patient / Surgical risk factors				
P-POSSUM	1.072	0.010	<0.001	1.053- 1.091
Male gender	1.231	0.132	0.053	0.997 -1.520
Orthopaedics	1.552	0.619	0.271	0.710 -3.393
General	3.687	1.742	0.006	1.460 -9.310
Vascular	2.413	1.127	0.059	0.966 - 6.030
Urology	1 (reference)			
POMS defined morbidity domains				
Pulmonary	1.055	0.149	0.705	0.800 - 1.392
Infection	1.047	0.128	0.707	0.824 -1.331
Renal	1.108	0.145	0.434	0. 857 - 1.432
Gastrointestinal	0.794	0.127	0.150	0.581 - 1.087
Cardiac	1.274	0.243	0.204	0.877 - 1.852
Neurological	1.951	0.437	0.003	1.257 - 3.027
Haematological	1.129	0.193	0.477	0.808 -1.580
Pain	0.923	0.132	0.575	0.697 -1.221
Wound	0.790	0.193	0.335	0.489 -1.275
Time dependent interactions between duration of morbidity and longer-term survival				
FMD15 + post-op Year 1	8.956	4.625	<0.001	3.255 - 24.641
FMD15+ post-op Year 2&3	9.107	4.388	<0.001	3.541 - 23.418
FMD15+ post-op ≥ Year 4	1.668	0.504	0.091	0.922 - 3.014
FMD8 + post-op Year 1	3.278	1.774	0.028	1.134 -9.472
FMD8 + post-op Year 2&3	5.130	2.476	0.001	1.992 -13.214
FMD8 + post-op ≥ Year 4	1.426	0.388	0.192	0.837 - 2.432
FMD5 + post-op Year 1	2.891	1.558	0.049	1.006 - 8.312
FMD5 + post-op Year 2&3	4.778	2.273	0.001	1.881 - 12.140
FMD5 + post-op ≥ Year 4	1.140	0.302	0.620	0.679 -1.915
FMD3 + post-op Year 1	2.641	1.464	0.008	0.891 - 7.829
FMD3 + post-op Year 2&3	4.168	1.998	0.003	1.629 -10.666
FMD3 + post-op ≥ Year 4	1.221	0.300	0.416	0.755 - 1.975
Interactions between types of surgery and POMS defined morbidity				
General / GI morbidity	0.876	0.284	0.682	0.463 -1.654
Orthopaedics / Wound morbidity	1.268	0.406	0.459	0.677 -2.373
Urology / Renal morbidity	1.053	0.469	0.907	0.440 - 2.519
Vascular /Cardiac morbidity	1.780	0.769	0.177	0.771 - 4.107

Table 41: Full Cox Proportional Hazards models for long term survival after major surgery

5.3.3 Relationship between perioperative risk, postoperative morbidity and long-term survival

These data show that, independent of pre and intra-operative risk factors, the occurrence of neurological morbidity after surgery (prevalence 2.9%) is associated with a relative hazard for long term mortality of 2.53 ($p<0.001$; 95% C.I. 1.66 - 3.84). Furthermore, if a patient remains in hospital on Day 15 post-surgery with any type of morbidity as defined by the Post-Operative Morbidity Survey (prevalence 15.6%), the relative hazard for death in the first 12 months post surgery is 3.52 ($p<0.001$; 95% C.I. 2.23-5.43) and for the next 2 years is 2.33 ($p<0.001$; 95% C.I. 1.56-3.50).

Variable	Hazard Ratio	Standard error	p value	95% C.I.
p-POSSUM percentage risk	1.07	0.01	<0.001	1.06-1.09
General surgery	2.23	0.26	<0.001	1.84-2.85
Vascular surgery	2.41	0.49	<0.001	1.62-3.58
Neurological morbidity	2.53	0.54	<0.001	1.66 - 3.84
FMD15 + Postoperative Year 1	3.52	0.78	<0.001	2.23-5.43
FMD15 + Postoperative Year 2-3	2.33	0.48	<0.001	1.56-3.50

Table 42: Final Cox Proportional Hazards Model for long term mortality after major surgery

5.4 Discussion

5.4.1 Summary of findings

Cox Proportional Hazards Modelling demonstrates that, patients with prolonged postoperative morbidity (as defined by remaining in hospital with POMS defined morbidity on Day 15 post-surgery) have a risk- adjusted mortality rate over 3.5 times higher than those who do not in the first year after surgery, and more than twice as high in the second and third postoperative years. After 3 years, the hazard for early mortality of prolonged postoperative morbidity returns to baseline when adjusted for other risk factors. Of specific postoperative complications, neurological morbidity was the only POMS defined domain which was independently associated with reduced postoperative survival.

5.4.2 Results in context of previously published work

The results differ from a previously published report (Khuri's 2005 paper analysing long-term survival in over 100,000 patients from the ACS-NSQIP)[151] in so far as the significance of the impact of major complications on longer term outcome is dependent on the duration of time since the surgical (and morbidity) insult. By using time as an interaction term with postoperative morbidity the temporal relationship between short and long term outcome has been elucidated in greater depth than previous reports. This was necessary because the initial analysis revealed that proportionality could not be demonstrated, therefore implying that the difference between the hazards in different risk groups varied over time. This was subsequently clearly demonstrated in the final

analyses. Proportionality testing was not declared in Khuri's paper, and in fact, the 'step' change in mortality rate which occurred at varying times postoperatively for different surgical procedures was highlighted in the results (see Figure 1 in the Introduction). These diagrams indicate that even in the NSQIP cohort, the hazard may in fact vary (most likely, reduce) with time, as has been found in the SOuRCe cohort.

With the exception of Khuri's paper detailed above, all the studies identified on review of the literature which have previously examined the relationship between perioperative complications and long-term survival, were conducted in cohorts of patients undergoing cancer surgery.[328-335] As the majority of patients in our study were undergoing elective non-cancer surgery, these findings have important implications for the informed consent process for patients undergoing surgery which may be life-enhancing (such as joint replacement) rather than life-saving.

5.4.3 Clinical implications

The significance of these results can be gauged by comparing against the findings epidemiological studies of clinical and environmental risk factors which are frequently discussed in the medical literature and in the general press. The SOuRCe study population was predominantly middle-aged (median age 66 years; IQR 56-75); in a similarly aged population, the adjusted relative hazard conferred by smoking more than 20 cigarettes per day was 2.62 (95% C.I. 2.20 to 3.12) for women and 3.04 (95% C.I. 2.71 to 3.40) for men.[336] Diabetes mellitus was shown to confer an adjusted relative

hazard for all-cause mortality of 2.3 (95% C.I. 1.6 to 3.2) in one study with a median follow up time of 5.2 years. In Caucasian adults, obesity has been found to be associated with a progressive increase in relative hazard (compared with a BMI 22.5-24.9) of 1.44 (95% CI, 1.38 to 1.50) for a BMI of 30.0 to 34.9; 1.88 (95% CI, 1.77 to 2.00) for a BMI of 35.0 to 39.9; and 2.51 (95% CI, 2.30 to 2.73) for a BMI of 40.0 to 49.9.[337] With prolonged morbidity, the relative hazard for year after surgery is 3.52 and for the next two years it is 2.33. While after this, relative hazard falls back to baseline, postoperative morbidity is still a significant, and importantly, potentially preventable source of long-term harm to patients.

Short–term postoperative morbidity and mortality continues to vary across providers and healthcare systems,[2;338] and evidence that these variations can be minimised is accumulating. The ACS-NSQIP is one enterprise that shows that open and careful scrutiny of surgical outcome leads to improvements.[339;340] Initiatives such as the World Health Organisation Safer Surgery Checklist, have demonstrated that instigating simple measures to improve process in the perioperative period may lead to a reduction in short-term mortality and morbidity.[341] Wider implementation of strategies such as goal-directed fluid therapy,[144] enhanced recovery (or 'fast- track surgery') programmes,[342] and the expansion of critical care facilities so that a greater number of high risk patients can be managed in high-acuity ward areas[2] is required to reduce the disease burden which arises from the development of postoperative complications.

Furthermore, excess mortality may not be the only consequence of perioperative complications. It is known from long term follow up studies in other areas such as critical care that even if a patient survives their illness, general health may deteriorate.[343] A recent study of patients who survived 5 years after surgery for oesophageal cancer, has shown a significant difference in health-related quality of life (related principally to dyspnoea, fatigue and eating habits) between patients who did and did not have major perioperative complications. [344] A similar relationship between perioperative morbidity and longer term HRQOL has also been demonstrated in colorectal surgery.[345] Nevertheless, it remains the case that whatever the preoperative risk is, a significant proportion of perioperative morbidity may be avoidable, and therefore clinical and health economic strategies should focus on preventing harm.

Previous work has demonstrated substantial international variation in postoperative morbidity rates.[346] Despite this, perioperative morbidity is not routinely measured or reported in many healthcare systems, including the UK NHS. Surgical mortality is often used as a marker of quality; however for many procedures, short-term postoperative mortality remains low; therefore very large sample sizes are required to be able to conduct meaningful comparisons between healthcare providers.[2] [347] These data demonstrate that the adverse health impact of a complicated postoperative course persists far beyond the endpoints at which surgical mortality is usually measured (such as 30-day or even one-year). Therefore, it is proposed that prolonged postoperative morbidity measured in this way, using a validated and objective outcome measure, may

be an important indicator of the quality of surgical healthcare, for the purposes of research, benchmarking and quality improvement.

5.4.4 Other potential explanations

Can the results of this study be explained in any other way? Inaccurate risk adjustment may lead to spurious associations being found. The use of the P-POSSUM model to adjust for perioperative risk seems valid based on the systematic review detailed in Chapter 2 and the analysis of Receiver Operator Characteristic curves for this cohort in Chapter 3. Importantly, the final model is clinically plausible: for example, POMS defined neurological morbidity has also been shown to be associated with an increased long-term hazard (not time dependent) – this is consistent with the literature reporting the lasting harm associated with postoperative cognitive dysfunction and delirium.[178]

It is also possible that the inflammatory response which is presumed to lead to increased longer term mortality, may exist sub-clinically and therefore undetected, even before surgery, and may therefore be the cause rather than the result of postoperative complications, subsequently leading to long-term risk. For example, low serum endotoxin core antibody (EndoCAb) levels are predictive of postoperative complications both in cardiac [23] and non-cardiac surgery, [24] and independently of POSSUM risk score.[24] However, if the long-term outcome of patients with complications was 'pre-determined' by an unknown or unmeasured risk factor, then one might expect that risk to be consistent over time, rather than changing, as is the case with the effect of complications on mortality found in this study.

5.4.5 Limitations of this study

The first limitation is generalizability; this was a single centre study and the patient population was entirely elective; it is possible that different risk factors would have been significant if emergency surgical patients had been included. Second, no adjustment was made for the effect that social deprivation might have had on outcome. Although a previous study looking at a sub-group of the orthopaedic patients in this cohort concluded that there was minimal socio-demographic gradient in risk or postoperative complications,[348] life expectancy is strongly related to index of deprivation[349] and therefore this may have influenced long-term survival. Third, the risk adjustment was confined by the limitations of the original SOuRCe dataset, and there are risk factors for both perioperative and long-term outcome which were not included in the model (diabetes mellitus for example). It is therefore possible that the case-mix adjustment was not as precise as possible and that the presence of latent, unmeasured confounding may have lead to an overestimation of the association between postoperative morbidity and long-term outcome. However, the P-POSSUM model has been widely validated as a perioperative risk adjuster, and in this population, we found it to be highly discriminant for the prediction of short-term mortality, therefore justifying its use. Finally, it is possible that our study may have been insufficiently powered to detect some significant predictors of reduced long-term survival. In particular, cardiac morbidity looked on the initial model to be significant; however, the confidence intervals were wide and in the stepwise analyses this lost statistical

significance. Analysis of a larger cohort using similar indicator variables would be a useful aim of future work.

5.5 Conclusions

1. Independent of preoperative risk, postoperative neurological morbidity is associated with reduced long-term survival

2. Independent of preoperative risk, prolonged postoperative morbidity is associated with reduced survival for the first three years after surgery, with over three times higher than adjusted baseline hazard in the first postoperative year, and over twice the adjusted baseline hazard for two years after that.

3. Prolonged postoperative morbidity may be a valid indicator of the quality of perioperative healthcare.

Chapter Six: Conclusions and future work

6.1 Summary of contents of thesis

Predicting risk and measuring outcome of major surgery is important to both patients and providers of healthcare. A number of studies and reports from within the UK and beyond have highlighted failings in current systems both in measuring risk and outcome, and also in acting on the results of these measurements. Recent reports have highlighted that clinicians seem reluctant to undertake routine perioperative risk prediction. This is a missed opportunity, as conducting such measurements would provide an individualised estimate of perioperative risk, thereby facilitating a more informative consent process for the patient, and enabling the perioperative team to plan their strategy of care more appropriately. This is particularly important as the evidence shows us that there is significant variation in patient outcomes and standards of care between healthcare institutions. Accurate perioperative risk prediction would enable the implementation of strategies aimed at modifying surgical risk and allow limited resources –such as the provision of planned postoperative critical care – to be allocated to those patients who would be most likely to benefit. The failure of clinicians to widely adopt the use of risk prediction tools in perioperative medicine may in part be due to a lack of understanding of the accuracy of these tools, and of their strengths and limitations.

This thesis has attempted to advance knowledge in how to predict risk and measure outcome. A systematic review of risk stratification tools identified two tools in particular – the P-POSSUM model and the Surgical Risk Scale – as being the instruments which have been most consistently and well validated for the prediction of mortality after surgery in heterogeneous populations of patients. A single-centre study of 1362 patients who had major inpatient orthopaedic, general, urological or vascular surgery, demonstrated that the P-POSSUM model was an accurate predictor of short-term mortality, although it was poorly calibrated, consistently over-predicting risk of adverse outcome. The Surgical Risk Scale performed poorly in this population. A simple additive POSSUM score was also found to be an accurate predictor of short-term mortality, and use of such a scoring scheme, rather than a regression model, might be of greater appeal to clinicians in day-to-day practice. Additionally, the POSSUM physiology score was also found to be a moderately precise predictor of short-term mortality and merits further validation in multicentre cohorts.

The epidemiology of postoperative morbidity was described using a validated measure of morbidity (the Post Operative Morbidity Survey, POMS) and mortality measured at a variety of shorter and longer term endpoints. Mortality, morbidity and length of stay varied between surgical specialities. A clear relationship between postoperative morbidity as measured by the POMS, and resource utilisation (length of stay), short term mortality and long-term survival was demonstrated on univariate analyses. In addition, the Day 5 POMS was found to have face validity as a single measure of

postoperative morbidity, with significant associations observed between individual POMS domains measured on Day 5, and short and longer term mortality.

The various threads of this thesis were drawn together in a final study evaluating the independent relationships between perioperative risk (measured using the P-POSSUM), surgical speciality, and postoperative morbidity. POMS-defined neurological morbidity confers a clinically and statistically significant long-term risk of premature death, independent of premorbid status as measured by P-POSSUM. Prolonged postoperative morbidity, defined as remaining in hospital on Day 15 post-surgery, with POMS-defined morbidity, and which afflicted 15% of the study population, was associated with more than 3.5 times the baseline adjusted hazard of death for one year after surgery, and more than twice the hazard for the next two years after that. I believe this to be the first study to specifically link prolonged morbidity, of any severity, with a longer term risk of premature death. This may also be the first study to demonstrate a fluctuating temporal relationship between postoperative morbidity and the longer-term risk of reduced survival: although the risk persists long after the patient is discharged from hospital, the magnitude of this risk reduces over time, returning to baseline after three years. Prolonged postoperative morbidity, measured in this objective and validated manner, would be an important indicator of the quality of surgical healthcare, which could be used both for the purposes of comparative audit, and as an outcome measure in clinical studies.

6.2 Future directions

6.2.1 Risk stratification

The P-POSSUM model is a valid, accurate predictor of postoperative mortality. Impact studies should be the next step in the evaluation of this risk prediction system.[248] A suitable study design might be a stepped-wedge cluster randomised controlled trial, in which the intervention is the implementation of risk prediction using P-POSSUM (perhaps using POSSUM physiology preoperatively, and then recalculating P-POSSUM at the end of surgery) and the subsequently the individualised modification of perioperative management in accordance with the predicted level of patient risk.

6.2.2 Morbidity measurement

The relationship demonstrated between short term morbidity and longer term mortality supports the argument for the routine measurement of postoperative morbidity as an indicator of the quality of surgical healthcare. However, the rate-limiting step in such an initiative would be the identification of resources to support it – particularly with the current financial constraints on the NHS. Health economic studies are required to evaluate the cost-effectiveness of measuring morbidity (either POMS measured on Day 5 or morbidity in patients with prolonged length of hospital stay, or both). Evaluation of novel technology would enhance efforts to reduce the resources required to measure both risk adjustment variables and the POMS, such as introducing automated linkage with hospital Information Technology systems. Perhaps most effective, however, would be either a 'carrot' (for example CQUIN incentive) or a 'stick'

(mandatory implementation of outcomes data collection demanded by the Department of Health). The latter approach, albeit with some government investment, has led to the development of the NSQIP in the US. There is some hope in the UK that the 'carrot' approach may work: the enrolment of over 90% of general adult critical care units in England with the Case-Mix Programme of the Intensive Care National Audit & Research Centre is a good example of clinical engagement with the aim of measuring risk adjusted outcome, for the purposes of quality improvement. The anaesthesia profession in the UK has a great track record of engaging in short-term data collection exercises, such as EuSOS,[2] the NCEPOD audits,[241] or the Royal College of Anaesthetists' National Audit Projects (NAPs)[350;351] Such successes give hope that the aspiration of routine perioperative outcome collection is not an impossible dream.

6.2.3 Conclusion

The P-POSSUM and Surgical Risk Scale are at least moderately accurate predictors of short term postoperative mortality. The POSSUM physiology score warrants further evaluation as a preoperative risk predictor. Prolonged postoperative morbidity is an independent predictor of longer term mortality, and would be an important and relevant measure of the quality of surgical healthcare. The implementation of routine measurement of risk adjusted morbidity and mortality after major surgery should be prioritised as an aim of health services research into surgical outcome in the future.

Appendix: Peer reviewed publications from this thesis

Chapter One:

Moonesinghe SR, Mythen MG, Grocott MP. High risk surgery: epidemiology and outcomes. *Anesth Analg*. 2011 Apr; 112(4): 891-901

Moonesinghe SR, Mythen MG, Grocott MP. Patient related risk factors for postoperative adverse events. *Curr Opin Crit Care*. 2009 Aug; 15(4): 320-7

Chapter Two:

Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MPW. Risk stratification tools in major non-cardiac surgery: qualitative systematic review. *Anesthesiology.* In press.

Reference List

(1) Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, Berry WR et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008; 372(9633):139-144.

(2) Pearse RM, Moreno RP, Bauer P, Pelosi P, Metnitz P, Spies C et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012; 380(9847):1059-1065.

(3) Kable AK, Gibberd RW, Spigelman AD. Adverse events in surgical patients in Australia. *Int J Qual Health Care* 2002; 14(4):269-276.

(4) Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery* 1999; 126(1):66-75.

(5) Head J, Ferrie JE, Alexanderson K, Westerlund H, Vahtera J, Kivimaki M. Diagnosis-specific sickness absence as a predictor of mortality: the Whitehall II prospective cohort study. *British Medical Journal* 2008; 337(7674).

(6) Niskanen MM, Takala JA. Use of resources and postoperative outcome. *Eur J Surg* 2001; 167(9):643-649.

(7) Pearse RM, Harrison DA, James P, Watson D, Hinds C, Rhodes A et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006; 10(3):R81.

(8) Shroyer AL, McDonald GO, Wagner BD, Johnson R, Schade LM, Bell MR et al. Improving quality of care in cardiac surgery: evaluating risk factors, processes of care, structures of care, and outcomes. *Semin Cardiothorac Vasc Anesth* 2008; 12(3):140-152.

(9) McCulloch P, Ward J, Tekkis PP, ASCOT group of surgeons, British Oesophago-Gastric Cancer Group. Mortality and morbidity in gastro-oesophageal cancer surgery: initial results of ASCOT multicentre prospective cohort study. *BMJ* 2003; 327(7425):1192-1197.

(10) Tekkis PP, Poloniecki JD, Thompson MR, Stamatakis JD. Operative mortality in colorectal cancer: prospective national study. *BMJ* 2003; 327(7425):1196-1201.

(11) Roberts SE, Goldacre MJ. Time trends and demography of mortality after fractured neck of femur in an English population, 1968-98: database study. *BMJ* 2003; 327(7418):771-775.

(12) Hadjianastassiou VG, Tekkis PP, Goldhill DR, Hands LJ. Quantification of mortality risk after abdominal aortic aneurysm repair. *British Journal of Surgery* 2005; 92(9):1092-1098.

(13) Bennett-Guerrero E. Systemic Inflammation. In: Kaplan J, Reich DL, Konstadt SN, editors. Cardiac Anaesthesia. Philadelphia: Saunders; 1998. 297-318.

(14) Slade MS, Simmons RL, Yunis E, Greenberg LJ. Immunodepression after major surgery in normal patients. *Surgery* 1975; 78(3):363-372.

(15) Bone RC. Toward a theory regarding the pathogenesis of the systemic inflammatory response syndrome: what we do and do not know about cytokine regulation. *Crit Care Med* 1996; 24(1):163-172.

(16) Nesher N, Alghamdi AA, Singh SK, Sever JY, Christakis GT, Goldman BS et al. Troponin after cardiac surgery: a predictor or a phenomenon?[see comment]. *Annals of Thoracic Surgery* 2008; 85(4):1348-1354.

(17) Ugras B, Giris M, Erbil Y, Gokpinar M, Citlak G, Issever H et al. Early prediction of anastomotic leakage after colorectal surgery by measuring peritoneal cytokines: prospective study. *International Journal Of Surgery* 2008; 6(1):28-35.

(18) Sheeran P, Hall GM. Cytokines in anaesthesia. *Br J Anaesth* 1997; 78(2):201-219.

(19) Oka Y, Murata A, Nishijima J, Yasuda T, Hiraoka N, Ohmachi Y et al. Circulating interleukin 6 as a useful marker for predicting postoperative complications.[erratum appears in Cytokine 1993 Jan;5(1):89-90]. *Cytokine* 1992; 4(4):298-304.

(20) Sander M, von HC, von D, V, Spaethe C, Konertz WF, Jain U et al. Increased interleukin-6 after cardiac surgery predicts infection. *Anesthesia & Analgesia* 2006; 102(6):1623-1629.

(21) Hamano K, Gohra H, Noda H, Katoh T, Fujimura Y, Zempo N et al. Increased serum interleukin-8: correlation with poor prognosis in patients with postoperative multiple organ failure. *World Journal of Surgery* 1998; 22(10):1077-1081.

(22) Mythen MG, Barclay GR, Purdy G, Hamilton-Davies C, Mackie IJ, Webb AR et al. The role of endotoxin immunity, neutrophil degranulation and contact activation in the pathogenesis of post-operative organ dysfunction. *Blood Coagul Fibrinolysis* 1993; 4(6):999-1005.

(23) Bennett-Guerrero E, Ayuso L, Hamilton-Davies C, White WD, Barclay GR, Smith PK et al. Relationship of preoperative antiendotoxin core antibodies and adverse outcomes following cardiac surgery. *JAMA* 1997; 277(8):646-650.

(24) Bennett-Guerrero E, Panah MH, Barclay GR, Bodian CA, Winfree WJ, Andres LA et al. Decreased endotoxin immunity is associated with greater mortality and/or prolonged hospitalization after surgery. *Anesthesiology* 2001; 94(6):992-998.

(25) Mathew JP, Grocott HP, Phillips-Bute B, Stafford-Smith M, Laskowitz DT, Rossignol D et al. Lower endotoxin immunity predicts increased cognitive dysfunction in elderly patients after cardiac surgery. *Stroke* 2003; 34(2):508-513.

(26) Moretti EW, Newman MF, Muhlbaier LH, Whellan D, Petersen RP, Rossignol D et al. Effects of decreased preoperative endotoxin core antibody levels on long-term mortality after coronary artery bypass graft surgery. *Arch Surg* 2006; 141(7):637-641.

(27) Moller JT, Cluitmans P, Rasmussen LS, Houx P, Rasmussen H, Canet J et al. Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International Study of Post-Operative Cognitive Dysfunction. *Lancet* 1998; 351(9106):857-861.

(28) Newman MF, Kirchner JL, Phillips-Bute B, Gaver V, Grocott H, Jones RH et al. Longitudinal Assessment of Neurocognitive Function after Coronary-Artery Bypass Surgery. *N Engl J Med* 2001; 344(6):395-402.

(29) Buvanendran A, Kroin JS, Berger RA, Hallab NJ, Saha C, Negrescu C et al. Upregulation of prostaglandin E-2 and interleukins in the central nervous system and peripheral tissue during and after surgery in humans. *Anesthesiology* 2006; 104(3):403-410.

(30) Wan YJ, Xu J, Meng FZ, Bao YH, Ge YY, Lobo N et al. Cognitive decline following major surgery is associated with gliosis, beta-amyloid accumulation, and tau phosphorylation in old mice. *Critical Care Medicine* 2010; 38(11):2190-2198.

(31) Bland RD, Shoemaker WC. Common physiologic patterns in general surgical patients: hemodynamic and oxygen transport changes during and after

operation in patients with and without associated medical problems. *Surg Clin North Am* 1985; 65(4):793-809.

(32) Shoemaker WC, Appel PL, Waxman K, Schwartz S, Chang P. Clinical trial of survivors' cardiorespiratory patterns as therapeutic goals in critically ill postoperative patients. *Crit Care Med* 1982; 10(6):398-403.

(33) Jhanji S, Lee C, Watson D, Hinds C, Pearse RM. Microvascular flow and tissue oxygenation after major abdominal surgery: association with post-operative complications. *Intensive Care Medicine* 2009; 35(4):671-677.

(34) Venn R, Steele A, Richardson P, Poloniecki J, Grounds M, Newman P. Randomized controlled trial to investigate influence of the fluid challenge on duration of hospital stay and perioperative morbidity in patients with hip fractures. *Br J Anaesth* 2002; 88(1):65-71.

(35) Abbas SM, Hill AG. Systematic review of the literature for the use of oesophageal Doppler monitor for fluid replacement in major abdominal surgery. *Anaesthesia* 2008; 63(1):44-51.

(36) Sobol JB, Wunsch H. Triage of high-risk surgical patients for intensive care. *Crit Care* 2011; 15(2):217.

(37) Simpson JC, Moonesinghe SR. Introduction to the Post Anaesthetic Care Unit. *Perioperative Medicine* 2013; 2(5).

(38) Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; 78(3):355-360.

(39) Polanczyk CsA, Marcantonio E, Goldman L, Rohde LEP, Orav J, Mangione CM et al. Impact of Age on Perioperative Complications and Length of Stay in Patients Undergoing Noncardiac Surgery. *Annals of Internal Medicine* 2001; 134(8):637-643.

(40) Virkkunen J, Heikkinen M, Lep+ñntalo M, Mets+ñnoja R, Salenius JP. Diabetes as an independent risk factor for early postoperative complications in critical limb ischemia. *Journal of vascular surgery* 2004; 40(4):761-767.

(41) Ackland GL, Moran N, Cone S, Grocott MP, Mythen MG. Chronic Kidney Disease and Postoperative Morbidity After Elective Orthopedic Surgery. *Anesth Analg* 2010.

(42) Roche JW, Wenn RT, Sahota O, Moran CG. Effect of comorbidities and postoperative complications on mortality after hip fracture in elderly people: prospective observational cohort study. *BMJ* 2005; 331.

(43) Fleisher LA, Beckman JA, Brown KA, Calkins H, Chaikof E, Fleischmann KE et al. ACC/AHA 2007 Guidelines on Perioperative Cardiovascular Evaluation and Care for Noncardiac Surgery: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery): Developed in Collaboration With the American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Rhythm Society, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, and Society for Vascular Surgery. *Circulation* 2007; 116(17):1971-1996.

(44) Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; 100(10):1043-1049.

(45) Ford MK, Beattie WS, Wijeysundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med* 2010; 152(1):26-35.

(46) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002; 89(6):763-768.

(47) Donati A, Ruzzi M, Adrario E, Pelaia P, Coluzzi F, Gabbanelli V et al. A new and feasible model for predicting operative risk. *Br J Anaesth* 2004; 93(3):393-399.

(48) Constantinides VA, Tekkis PP, Senapati A, Association of Coloproctology of Great Britain and Ireland. Comparison of POSSUM scoring systems and the surgical risk scale in patients undergoing surgery for complicated diverticular disease. *Diseases of the Colon & Rectum* 2006; 49(9):1322-1331.

(49) Tambyraja AL, Kumar S, Nixon SJ. POSSUM scoring for laparoscopic cholecystectomy in the elderly. *ANZ Journal of Surgery* 2005; 75(7):550-552.

(50) Tamijmarane A, Bhati CS, Mirza DF, Bramhall SR, Mayer DA, Wigmore SJ et al. Application of Portsmouth modification of physiological and operative severity scoring system for enumeration of morbidity and mortality (P-POSSUM) in pancreatic surgery. *World Journal of Surgical Oncology* 2008; 6:39.

(51) Campillo-Soto A, Flores-Pastor B, Soria-Aledo V, Candel-Arenas M, ndres-Garcia B, Martin-Lorenzo JG et al. [The POSSUM scoring system: an instrument for measuring quality in surgical patients]. [Spanish]. *Cirugia Espanola* 2006; 80(6):395-399.

(52) Bollschweiler E, Lubke T, Monig SP, Holscher AH. Evaluation of POSSUM scoring system in patients with gastric cancer undergoing D2-gastrectomy. *BMC Surgery* 2005; 5:8.

(53) Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's Acute Physiology and Chronic Health Evaluation (APACHE II) study in Britain and Ireland: a prospective, multicenter, cohort study comparing two methods for predicting outcome for adult intensive care patients. *Crit Care Med* 1994; 22(9):1392-1401.

(54) Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med* 2007; 35(4):1091-1098.

(55) Ackland GL, Harris S, Ziabari Y, Grocott M, Mythen M. Revised cardiac risk index and postoperative morbidity after elective orthopaedic surgery: a prospective cohort study. *Br J Anaesth* 2010; 105(6):744-752.

(56) Grocott MP, Browne JP, Van der MJ, Matejowsky C, Mutch M, Hamilton MA et al. The Postoperative Morbidity Survey was validated and used to describe morbidity after major surgery. *J Clin Epidemiol* 2007; 60(9):919-928.

(57) Khuri SF, Henderson WG, DePalma RG, Mosca C, Healey NA, Kumbhani DJ et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Annals of Surgery* 2005; 242(3):326-341.

(58) Patterson BO, Holt PJE, Hinchliffe R, Loftus IM, Thompson MM. Predicting Risk in Elective Abdominal Aortic Aneurysm Repair: A Systematic Review of Current Evidence. *European Journal of Vascular and Endovascular Surgery* 2008;(6):637-645.

(59) Ramkumar T, Ng V, Fowler L, Farouk R. A comparison of POSSUM, P-POSSUM and colorectal POSSUM for the prediction of postoperative mortality in patients undergoing colorectal resection. *Diseases of the Colon and Rectum* 2006; 49(3):330-335.

(60) Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):328-340.

(61) Fink AS, Campbell DA, Jr., Mentzer RM, Jr., Henderson WG, Daley J, Bannister J et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Annals of Surgery* 2002; 236(3):344-353.

(62) Saklad M. Grading of Patients for Surgical Procedures. *Anesthesiology* 1941; 2(3).

(63) Wolters U, Wolf T, Stutzer H, Schroder T. ASA classification and perioperative variables as predictors of postoperative outcome. *Br J Anaesth* 1996; 77(2):217-222.

(64) Wolters U, Wolf T, Stutzer H, Schroder T, Pichlmaier H. Risk factors, complications, and outcome in surgery: a multivariate analysis. *Eur J Surg* 1997; 163(8):563-568.

(65) Grocott MP, Levett DZ, Matejowsky C, Emberton M, Mythen MG. ASA scores in the preoperative patient: feedback to clinicians can improve data quality. *J Eval Clin Pract* 2007; 13(2):318-319.

(66) Burgos E, Gomez-Arnau JI, Diez R, Munoz L, Fernandez-Guisasola J, Garcia d, V. Predictive value of six risk scores for outcome after surgical repair of hip fracture in elderly patients. *Acta Anaesthesiologica Scandinavica* 2008; 52(1):125-131.

(67) Nuttall M, Van Der MJ, Emberton M. Charlson scores based on ICD-10 administrative data were valid in assessing comorbidity in patients undergoing urological cancer surgery. *Journal of Clinical Epidemiology* 2006; 59(3):265-273.

(68) Li B, Evans D, Faris P, Dean S, Quan H. Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC Health Services Research* 2008; 12(8).

(69) Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 1998; 85(9):1217-1220.

(70) Ramesh VJ, Rao GS, Guha A, Thennarasu K. Evaluation of POSSUM and P-POSSUM scoring systems for predicting the mortality in elective neurosurgical patients. *British Journal of Neurosurgery* 2008; 22(2):275-278.

(71) Senagore AJ, Delaney CP, Duepree HJ, Brady KM, Fazio VW. Evaluation of POSSUM and P-POSSUM scoring systems in assessing outcome after laparoscopic colectomy. *British Journal of Surgery* 2003; 90(10):1280-1284.

(72) Horzic M, Kopljar M, Cupurdija K, Bielen DV, Vergles D, Lackovic Z. Comparison of P-POSSUM and Cr-POSSUM scores in patients undergoing colorectal cancer resection. *Archives of Surgery* 2007; 142(11):1043-1048.

(73) Reilly DF, McNeely MJ, Doerner D, Greenberg DL, Staiger TO, Geist MJ et al. Self-reported exercise tolerance and the risk of serious perioperative complications. *Arch Intern Med* 1999; 159(18):2185-2192.

(74) Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). *Am J Cardiol* 1989; 64(10):651-654.

(75) Murray P, Whiting P, Hutchinson SP, Ackroyd R, Stoddard CJ, Billings C. Preoperative shuttle walking testing and outcome after oesophagogastrectomy. *British Journal of Anaesthesia* 2007; 99(6):809-811.

(76) Win T, Jackson A, Groves AM, Sharples LD, Charman SC, Laroche CM. Comparison of shuttle walk with measured peak oxygen consumption in patients with operable lung cancer. *Thorax* 2006; 61(1):57-60.

(77) Milani RV, Lavie CJ, Mehra MR. Cardiopulmonary exercise testing: how do we differentiate the cause of dyspnea? *Circulation* 2004; 110(4):e27-e31.

(78) Older P, Smith R, Courtney P, Hone R. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing.[see comment]. *Chest* 1993; 104(3):701-704.

(79) Older P, Hall A, Hader R. Cardiopulmonary exercise testing as a screening test for perioperative management of major surgery in the elderly. *Chest* 1999; 116(2):355-362.

(80) Bolliger CT, Jordan P, Soler M, Stulz P, Gradel E, Skarvan K et al. Exercise capacity as a predictor of postoperative complications in lung resection candidates. *American Journal of Respiratory & Critical Care Medicine* 1995; 151(5):1472-1480.

(81) Kim HJ, Park SW, Cho BR, Hong SH, Park PW, Hong KP. The role of cardiopulmonary exercise test in mitral and aortic regurgitation: it can predict post-operative results. *Korean Journal of Internal Medicine* 2003; 18(1):35-39.

(82) Forshaw MJ, Strauss DC, Davies AR, Wilson D, Lams B, Pearce A et al. Is cardiopulmonary exercise testing a useful test before esophagectomy? *Annals of Thoracic Surgery* 2008; 85(1):294-299.

(83) Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR. Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. *Br J Anaesth* 2008; 101(6):774-780.

(84) Atkinson AJ, Colburn WA, DeGruttola VG, Demets DL, Downing GJ, Hoth DF et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* 2001; 69(3):89-95.

(85) Tsimikas S, Willerson JT, Ridker PM. C-reactive protein and other emerging blood biomarkers to optimize risk stratification of vulnerable patients. *J Am Coll Cardiol* 2006; 47(8 Suppl):C19-C31.

(86) Kullo IJ, Khaleghi M, Hensrud DD. Markers of inflammation are inversely associated with VO2 max in asymptomatic men. *J Appl Physiol* 2007; 102(4):1374-1379.

(87) Ridker PM, Buring JE, Shih J, Matias M, Hennekens CH. Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women. *Circulation* 1998; 98(8):731-733.

(88) Ackland GL, Scollay JM, Parks RW, de B, I, Mythen MG. Pre-operative high sensitivity C-reactive protein and postoperative outcome in patients undergoing elective orthopaedic surgery. *Anaesthesia* 2007; 62(9):888-894.

(89) Owens CD, Ridker PM, Belkin M, Hamdan AD, Pomposelli F, Logerfo F et al. Elevated C-reactive protein levels are associated with postoperative events in patients undergoing lower extremity vein bypass surgery. *J Vasc Surg* 2007; 45(1):2-9.

(90) Heider P, Poppert H, Wolf O, Liebig T, Pelisek J, Schuster T et al. Fibrinogen and high-sensitive C-reactive protein as serologic predictors for perioperative cerebral microembolic lesions after carotid endarterectomy. *Journal of Vascular Surgery* 2007; 46(3):449-454.

(91) Groschel K, Ernemann U, Larsen J, Knauth M, Schmidt F, Artschwager J et al. Preprocedural C-reactive protein levels predict stroke and death in patients undergoing carotid stenting. *Ajnr: American Journal of Neuroradiology* 2007; 28(9):1743-1746.

(92) Shiu YC, Lin JK, Huang CJ, Jiang JK, Wang LW, Huang HC et al. Is C-reactive protein a prognostic factor of colorectal cancer? *Diseases of the Colon & Rectum* 2008; 51(4):443-449.

(93) Komai Y, Saito K, Sakai K, Morimoto S. Increased preoperative serum C-reactive protein level predicts a poor prognosis in patients with localized renal cell carcinoma. *BJU International* 2007; 99(1):77-80.

(94) Crumley AB, McMillan DC, McKernan M, Going JJ, Shearer CJ, Stuart RC. An elevated C-reactive protein concentration, prior to surgery, predicts poor

cancer-specific survival in patients undergoing resection for gastro-oesophageal cancer. *British Journal of Cancer* 2006; 94(11):1568-1571.

(95) Cappabianca G, Paparella D, Visicchio G, Capone G, Lionetti G, Numis F et al. Preoperative C-reactive protein predicts mid-term outcome after cardiac surgery. *Annals of Thoracic Surgery* 2006; 82(6):2170-2178.

(96) Kangasniemi OP, Biancari F, Luukkonen J, Vuorisalo S, Satta J, Pokela R et al. Preoperative C-reactive protein is predictive of long-term outcome after coronary artery bypass surgery. *European Journal of Cardio-Thoracic Surgery* 2006; 29(6):983-985.

(97) Pedrazzini GB, Masson S, Latini R, Klersy C, Rossi MG, Pasotti E et al. Comparison of brain natriuretic peptide plasma levels versus logistic EuroSCORE in predicting in-hospital and late postoperative mortality in patients undergoing aortic valve replacement for symptomatic aortic stenosis. *American Journal of Cardiology* 2008; 102(6):749-754.

(98) Akazawa T, Nishihara H, Iwata H, Warabi K, Ohshima M, Inada E. Preoperative plasma brain natriuretic peptide level is an independent predictor of postoperative atrial fibrillation following off-pump coronary artery bypass surgery. *J Anesth* 2008; 22(4):347-353.

(99) Sarzi BS, Vaninetti R, Pedretti RF. Plasma B-type natriuretic peptide predicts atrial fibrillation during rehabilitation after cardiac surgery. *Eur J Cardiovasc Prev Rehabil* 2008; 15(4):460-466.

(100) Leibowitz D, Planer D, Rott D, Elitzur Y, Chajek-Shaul T, Weiss AT. Brain natriuretic peptide levels predict perioperative events in cardiac patients undergoing noncardiac surgery: a prospective study. *Cardiology* 2008; 110(4):266-270.

(101) Yun KH, Jeong MH, Oh SK, Choi JH, Rhee SJ, Park EM et al. Preoperative plasma N-terminal pro-brain natriuretic peptide concentration and perioperative cardiovascular risk in elderly patients. *Circulation Journal* 2008; 72(2):195-199.

(102) Cuthbertson BH, Card G, Croal BL, McNeilly J, Hillis GS. The utility of B-type natriuretic peptide in predicting postoperative cardiac events and mortality in patients undergoing major emergency non-cardiac surgery. *Anaesthesia* 2007; 62(9):875-881.

(103) Dernellis J, Panaretou M. Assessment of cardiac risk before non-cardiac surgery: brain natriuretic peptide in 1590 patients. *Heart* 2006; 92(11):1645-1650.

(104) Redfield MM, Rodeheffer RJ, Jacobsen SJ, Mahoney DW, Bailey KR, Burnett JC, Jr. Plasma brain natriuretic peptide concentration: impact of age and gender. *J Am Coll Cardiol* 2002; 40(5):976-982.

(105) Goei D, Schouten O, Boersma E, Welten GM, Dunkelgrun M, Lindemans J et al. Influence of renal function on the usefulness of N-terminal pro-B-type natriuretic peptide as a prognostic cardiac risk marker in patients undergoing noncardiac vascular surgery. *Am J Cardiol* 2008; 101(1):122-126.

(106) Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE et al. Clinical assessment incorporating a personal genome. *Lancet* 2010; 375(9725):1525-1535.

(107) Schrijver I, Galli SJ. Between hype and hope: whole-genome sequencing in clinical medicine. *Personalized Medicine* 2012; 9(3):243-246.

(108) Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; 44(3):Suppl-206.

(109) Ghaferi AA, Osborne NH, Birkmeyer JD, Dimick JB. Hospital Characteristics Associated with Failure to Rescue from Complications after Pancreatectomy. *Journal of the American College of Surgeons* 2010; 211(3):325-330.

(110) Hartz AJ, Krakauer H, Kuhn EM, Young M, Jacobsen SJ, Gay G et al. Hospital Characteristics and Mortality-Rates. *N Engl J Med* 1989; 321(25):1720-1725.

(111) Silber JH, Romano PS, Rosen AK, Wang Y, Even-Shoshan O, Holpp KG. Failure-to-rescue - Comparing definitions to measure quality of care. *Medical Care* 2007; 45(10):918-925.

(112) Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA* 2002; 288(16):1987-1993.

(113) Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009; 360(5):491-499.

(114) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: Site visit assessment of process and structure. *Journal of the American College of Surgeons* 1997; 185(4):341-351.

(115) Mahajan RP. Critical incident reporting and learning. *Br J Anaesth* 2010; 105(1):69-75.

(116) Zare MM, Itani KM, Schifftner TL, Henderson WG, Khuri SF. Mortality after nonemergent major surgery performed on Friday versus Monday through Wednesday. *Ann Surg* 2007; 246(5):866-874.

(117) Bell CM, Redelmeier DA. Mortality among patients admitted to hospitals on weekends as compared with weekdays. *N Engl J Med* 2001; 345(9):663-668.

(118) Magid DJ, Wang Y, Herrin J, McNamara RL, Bradley EH, Curtis JP et al. Relationship between time of day, day of week, timeliness of reperfusion, and in-hospital mortality for patients with acute ST-segment elevation myocardial infarction. *JAMA* 2005; 294(7):803-812.

(119) Wunsch H, Mapstone J, Brady T, Hanks R, Rowan K. Hospital mortality associated with day and time of admission to intensive care units. *Intensive Care Med* 2004; 30(5):895-901.

(120) Goldfrad C, Rowan K. Consequences of discharges from intensive care at night. *Lancet* 2000; 355(9210):1138-1142.

(121) Jen MH, Bottle A, Majeed A, Bell D, Aylin P. Early In-Hospital Mortality following Trainee Doctors' First Day at Work. *PLoS ONE* 2009; 4(9):e7103.

(122) Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009; 361(14):1368-1375.

(123) Using the Commissioning for Quality and Innovation Payment (CQUIN) Framework. *http://www dh gov uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_091443* 2012.

(124) Regional Enhanced Recovery CQUIN. *http://www london nhs uk/what-we-do/our-current-projects/enhanced-recovery-programme/further-information-on-enchanced-recovery* 2012.

(125) IHI Global Trigger Tool for measuring adverse events. *http://www ihi org/knowledge/Pages/Tools/IHIGlobalTriggerToolforMeasuringAEs aspx* 2012.

(126) Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002; 137(6):511-520.

(127) Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002; 346(15):1128-1137.

(128) Reavis KM, Smith BR, Hinojosa MW, Nguyen NT. Outcomes of esophagectomy at academic centers: an association between volume and outcome. *American Surgeon* 2008; 74(10):939-943.

(129) Miyata H, Motomura N, Ueda Y, Matsuda H, Takamoto S. Effect of procedural volume on outcome of coronary artery bypass graft surgery in Japan: implication toward public reporting and minimal volume standards.[see comment]. *Journal of Thoracic & Cardiovascular Surgery* 2008; 135(6):1306-1312.

(130) Lin HC, Xirasagar S, Tsao NW, Hwang YT, Kuo NW, Lee HC. Volume-outcome relationships in coronary artery bypass graft surgery patients: 5-year major cardiovascular event outcomes. *Journal of Thoracic & Cardiovascular Surgery* 2008; 135(4):923-930.

(131) Holt PJ, Poloniecki JD, Loftus IM, Thompson MM. The relationship between hospital case volume and outcome from carotid endartectomy in England from 2000 to 2005. *European Journal of Vascular & Endovascular Surgery* 2007; 34(6):646-654.

(132) Hamilton BH, Ho V. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Med Care* 1998; 36(6):892-903.

(133) Khuri SF, Henderson WG. The case against volume as a measure of quality of surgical care. *World J Surg* 2005; 29(10):1222-1229.

(134) Birkmeyer JD, Dimick JB, Birkmeyer NJO. Measuring the quality of surgical care: Structure, process, or outcomes? *Journal of the American College of Surgeons* 2004; 198(4):626-632.

(135) Helen B, Rosalind R. Hospital service reconfiguration: the battle for hearts and minds. *BMJ* 2012; 344.

(136) Holt PJE, Poloniecki JD, Loftus IM, Michaels JA, Thompson MM. Epidemiological study of the relationship between volume and outcome after abdominal aortic aneurysm surgery in the UK from 2000 to 2005. *Br J Surg* 2007; 94(4):441-448.

(137) Mayer EK, Bottle A, Aylin P, Darzi AW, Athanasiou T, Vale JA. The volume-outcome relationship for radical cystectomy in England: an analysis of outcomes other than mortality. *Bju International* 2011; 108(8B):E258-E265.

(138) Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. *British Medical Journal* 2010; 340.

(139) Pal N, Axisa B, Yusof S, Newcombe RG, Wemyss-Holden S, Rhodes M et al. Volume and outcome for major upper GI surgery in england. *Journal of Gastrointestinal Surgery* 2008; 12(2):353-357.

(140) Locker TE, Mason SM. Analysis of the distribution of time that patients spend in emergency departments. *British Medical Journal* 2005; 330(7501):1188-1189.

(141) Haller G, Stoelwinder J, Myles PS, McNeil J. Quality and safety indicators in anesthesia: a systematic review. *Anesthesiology* 2009; 110(5):1158-1175.

(142) Walsh SR, Tang T, Bass S, Gaunt ME. Doppler-guided intra-operative fluid management during major abdominal surgery: systematic review and meta-analysis. *Int J Clin Pract* 2008; 62(3):466-470.

(143) Abbas SM, Hill AG. Systematic review of the literature for the use of oesophageal Doppler monitor for fluid replacement in major abdominal surgery. *Anaesthesia* 2008; 63(1):44-51.

(144) Kuper M, Gold SJ, Callow C, Quraishi T, King S, Mulreany A et al. Intraoperative fluid management guided by oesophageal Doppler monitoring. *BMJ* 2011; 342:d3016.

(145) Taketomi A, Kitagawa D, Itoh S, Harimoto N, Yamashita Y, Gion T et al. Trends in morbidity and mortality after hepatic resection for hepatocellular carcinoma: an institute's experience with 625 patients. *Journal of the American College of Surgeons* 2007; 204(4):580-587.

(146) Ho V, Heslin MJ, Yun H, Howard L. Trends in hospital and surgeon volume and operative mortality for cancer surgery. *Annals of Surgical Oncology* 2006; 13(6):851-858.

(147) Ivanov J, Weisel RD, David TE, Naylor CD. Fifteen-year trends in risk severity and operative mortality in elderly patients undergoing coronary artery bypass graft surgery. *Circulation* 1998; 97(7):673-680.

(148) Walsh SR, Tang T, Bass S, Gaunt ME. Doppler-guided intra-operative fluid management during major abdominal surgery: systematic review and meta-analysis. *Int J Clin Pract* 2008; 62(3):466-470.

(149) Wakeling HG, McFall MR, Jenkins CS, Woods WG, Miles WF, Barclay GR et al. Intraoperative oesophageal Doppler guided fluid management shortens postoperative hospital stay after major bowel surgery. *Br J Anaesth* 2005; 95(5):634-642.

(150) Pearse R, Dawson D, Fawcett J, Rhodes A, Grounds RM, Bennett ED. Early goal-directed therapy after major surgery reduces complications and duration of hospital stay. A randomised, controlled trial [ISRCTN38797445]. *Crit Care* 2005; 9(6):R687-R693.

(151) Khuri SF, Henderson WG, DePalma RG, Mosca C, Healey NA, Kumbhani DJ. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg* 2005; 242(3):326-341.

(152) Schiesser M, Chen JW, Maddern GJ, Padbury RT. Perioperative morbidity affects long-term survival in patients following liver resection for colorectal metastases. *Journal of Gastrointestinal Surgery* 2008; 12(6):1054-1060.

(153) Laurent C, Sa CA, Couderc P, Rullier E, Saric J. Influence of postoperative morbidity on long-term survival following liver resection for colorectal metastases.[see comment]. *British Journal of Surgery* 2003; 90(9):1131-1136.

(154) Lehrke S, Steen H, Sievers HH, Peters H, Opitz A, Muller-Bardorff M et al. Cardiac troponin T for prediction of short- and long-term morbidity and mortality after elective open heart surgery. *Clinical Chemistry* 2004; 50(9):1560-1567.

(155) Toumpoulis IK, Anagnostopoulos CE, DeRose JJ, Jr., Swistel DG. The impact of deep sternal wound infection on long-term survival after coronary artery bypass grafting. *Chest* 2005; 127(2):464-471.

(156) Rodrigo JP, Suarez C. Prognostic significance of postoperative wound infection on head and neck cancer. *Otolaryngology - Head & Neck Surgery* 1998; 118(2):272-275.

(157) Kertai MD, Boersma E, Klein J, van Urk H, Bax JJ, Poldermans D. Long-term prognostic value of asymptomatic cardiac troponin T elevations in patients after major vascular surgery. *European Journal of Vascular and Endovascular Surgery* 2004; 28(1):59-66.

(158) Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012; 307(21):2295-2304.

(159) Wright JG. Outcomes research: what to measure. *World J Surg* 1999; 23(12):1224-1226.

(160) Brenneman FD, Wright JG, Kennedy ED, McLeod RS. Outcomes research in surgery. *World J Surg* 1999; 23(12):1220-1223.

(161) Ellis H, Calne R, Watson C. General Surgery (Lecture Notes). Oxford: Wiley-Blackwell; 2006.

(162) Russell RGC, Williams NS, Bulstrode C.J.K. Bailey and Love's Short Practice of Surgery. London: Hodder and Arnold; 2004.

(163) Mythen MG. Postoperative gastrointestinal tract dysfunction. *Anesthesia and Analgesia* 2005; 100(1):196-204.

(164) O'Neil-Callahan K, Katsimaglis G, Tepper MR, Ryan J, Mosby C, Ioannidis JPA et al. Statins decrease perioperative cardiac complications in patients undergoing noncardiac vascular surgery: The Statins for Risk Reduction in Surgery (StaRRS) study. *Journal of the American College of Cardiology* 2005; 45(3):336-342.

(165) Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. *Ann Surg* 2010; 251(3):535-541.

(166) Poldermans D, Bax JJ, Boersma E, De HS, Eeckhout E, Fowkes G et al. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. *Eur Heart J* 2009; 30(22):2769-2812.

(167) Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008; 371(9627):1839-1847.

(168) Poldermans D, Schouten O, Vidakovic R, Bax JJ, Thomson IR, Hoeks SE et al. A clinical randomized trial to evaluate the safety of a noninvasive approach in high-risk patients undergoing major vascular surgery: the DECREASE-V Pilot Study. *J Am Coll Cardiol* 2007; 49(17):1763-1769.

(169) McFalls EO, Ward HB, Moritz TE, Goldman S, Krupski WC, Littooy F et al. Coronary-artery revascularization before elective major vascular surgery. *N Engl J Med* 2004; 351(27):2795-2804.

(170) Mythen M. Pre-operative coronary revascularisation before non-cardiac surgery: think long and hard before making a pre-operative referral. *Anaesthesia* 2009; 64(10):1048-1050.

(171) Older P, Smith R, Courtney P, Hone R. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing. *Chest* 1993; 104(3):701-704.

(172) Dyer CB, Ashton CM, Teasdale TA. Postoperative delirium. A review of 80 primary data-collection studies. *Arch Intern Med* 1995; 155(5):461-465.

(173) Kalisvaart KJ, Vreeswijk R, De Jonghe JFM, Van Der Ploeg T, Van Gool WA, Eikelenboom P. Risk Factors and Prediction of Postoperative Delirium in Elderly Hip-Surgery Patients: Implementation and Validation of a Medical Risk Factor Model. *J Am Geriatr Soc* 2006; 54(5):817-822.

(174) Litaker D, Locala J, Franco K, Bronson DL, Tannous Z. Preoperative risk factors for postoperative delirium. *Gen Hosp Psychiatry* 2001; 23(2):84-89.

(175) Marcantonio ER, Goldman L, Orav EJ, Cook EF, Lee TH. The association of intraoperative factors with the development of postoperative delirium. *Am J Med* 1998; 105(5):380-384.

(176) Robinson TN, Raeburn CD, Tran ZV, Angles EM, Brenner LA, Moss M. Postoperative Delirium in the Elderly: Risk Factors and Outcomes. *Annals of Surgery* 2009; 249(1).

(177) Edelstein DM, Aharonoff GB, Karp A, Capla EL, Zuckerman JD, Koval KJ. Effect of Postoperative Delirium on Outcome after Hip Fracture. *Clinical Orthopaedics and Related Research* 2004; 422.

(178) Steinmetz J, Christensen KB, Lund T, Lohse N, Rasmussen LS, the ISPOCD Group. Long-term Consequences of Postoperative Cognitive Dysfunction. *Anesthesiology* 2009; 110(3).

(179) Hou SH, Bushinsky DA, Wish JB, Cohen JJ, Harrington JT. Hospital-acquired renal insufficiency: A prospective study. *Am J Med* 1983; 74(2):243-248.

(180) Sear JW. Kidney dysfunction in the postoperative period. *Br J Anaesth* 2005; 95(1):20-32.

(181) Novis BK, Roizen MF, Aronson S, Thisted RA. Association of Preoperative Risk-Factors with Postoperative Acute-Renal-Failure. *Anesthesia and Analgesia* 1994; 78(1):143-149.

(182) Coca SG, Yusuf B, Shlipak MG, Garg AX, Parikh CR. Long-term Risk of Mortality and Other Adverse Outcomes After Acute Kidney Injury: A Systematic Review and Meta-analysis. *American Journal of Kidney Diseases* 2009; 53(6):961-973.

(183) Bennett-Guerrero E, Welsby I, Dunn TJ, Young LR, Wahl TA, Diers TL et al. The use of a postoperative morbidity survey to evaluate patients with prolonged hospitalization after routine, moderate-risk, elective surgery. *Anesth Analg* 1999; 89(2):514-519.

(184) Mythen MG. Postoperative gastrointestinal tract dysfunction: An overview of causes and management strategies. *Cleveland Clinic Journal of Medicine* 2009; 76(Suppl 4):S66-S71.

(185) Mythen MG, Webb AR. The Role of Gut Mucosal Hypoperfusion in the Pathogenesis of Postoperative Organ Dysfunction. *Intensive Care Medicine* 1994; 20(3):203-209.

(186) Mythen MG, Webb AR. Intraoperative Gut Mucosal Hypoperfusion Is Associated with Increased Postoperative Complications and Cost. *Intensive Care Medicine* 1994; 20(2):99-104.

(187) Chieveley-Williams S, Hamilton-Davies C. The role of the gut in major surgical postoperative morbidity. *Int Anesthesiol Clin* 1999; 37(2):81-110.

(188) Vincent JL. Clinical sepsis and septic shock-definition, diagnosis and management principles. *Langenbecks Arch Surg* 2008; 393(6):817-824.

(189) Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995; 23(10):1638-1652.

(190) Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013; 41(2):580-637.

(191) Jablonski S. Syndrome--a changing concept. *Bull Med Libr Assoc* 1992; 80(4):323-327.

(192) Martin RCG, Brennan MF, Jaques DP. Quality of complication reporting in the surgical literature. *Ann Surg* 2002; 235(6):803-812.

(193) Sinha S, Sinha S, Ashby E, Jayaram R, Grocott MP. Quality of reporting in randomized trials published in high-quality surgical journals. *J Am Coll Surg* 2009; 209(5):565-571.

(194) Dindo D, Hahnloser D, Clavien PA. Quality assessment in surgery: riding a lame horse. *Ann Surg* 2010; 251(4):766-771.

(195) Dindo D, Demartines N, Clavien PA. Classification of surgical complications - A new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Annals of Surgery* 2004; 240(2):205-213.

(196) Clavien PA, Sanabria JR, Mentha G, Borst F, Buhler L, Roche B et al. Recent results of elective open cholecystectomy in a North American and a European

center. Comparison of complications and risk factors.[see comment]. *Annals of Surgery* 1992; 216(6):618-626.

(197) Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD et al. The Clavien-Dindo Classification of Surgical Complications Five-Year Experience. *Annals of Surgery* 2009; 250(2):187-196.

(198) Strasberg SM, Linehan DC, Hawkins WG. The Accordion Severity Grading System of Surgical Complications. *Annals of Surgery* 2009; 250(2):177-186.

(199) Clavien PA, Strasberg SM. Severity Grading of Surgical Complications. *Annals of Surgery* 2009; 250(2):197-198.

(200) Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric and Psychometric Strategies for Development of a Health Measurement Scale. *J Clin Epidemiol* 1999; 52(2):105-111.

(201) Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G et al. The National Veterans-Administration Surgical Risk Study - Risk Adjustment for the Comparative-Assessment of the Quality Surgical Care. *Journal of the American College of Surgeons* 1995; 180(5):519-531.

(202) Alvord LA, Rhoades D, Henderson WG, Goldberg JH, Hur K, Khuri SF et al. Surgical morbidity and mortality among American Indian and Alaska Native veterans: A comparative analysis. *Journal of the American College of Surgeons* 2005; 200(6):837-844.

(203) Sanders J, Keogh BE, Van der Meulen J, Browne JP, Treasure T, Mythen MG et al. The development of a postoperative morbidity score to assess total morbidity burden after cardiac surgery. *J Clin Epidemiol* 2012; 65(4):423-433.

(204) Cecconi M, Fasano N, Langiano N, Divella M, Costa MG, Rhodes A et al. Goal-directed haemodynamic therapy during elective total hip arthroplasty under regional anaesthesia. *Critical Care* 2011; 15(3).

(205) Lebuffe G, Vallet B, Takala J, Hartstein G, Lamy M, Mythen M et al. A European, multicenter, observational study to assess the value of gastric-to-end tidal Pco(2) difference in predicting postoperative complications. *Anesthesia and Analgesia* 2004; 99(1):166-172.

(206) Ackland GL, Scollay JM, Parks RW, de Beaux I, Mythen MG. Pre-operative high sensitivity C-reactive protein and postoperative outcome in patients undergoing elective orthopaedic surgery. *Anaesthesia* 2007; 62(9):888-894.

(207) Scollay JM, de Beaux I, Parks RW. Prospective Study of Intra-Abdominal Pressure Following Major Elective Abdominal Surgery. *World Journal of Surgery* 2009; 33(11):2372-2377.

(208) Davies SJ, Yates D, Wilson RJT. Dopexamine Has No Additional Benefit in High-Risk Patients Receiving Goal-Directed Fluid Therapy Undergoing Major Abdominal Surgery. *Anesthesia and Analgesia* 2011; 112(1):130-138.

(209) Phypers BJ, Robiony-Rogers D, Pickering RM, Garden AL. Test-retest reliability of the oxygen uptake efficiency slope in surgical patients. *Anaesthesia* 2011; 66(8):659-666.

(210) Hennis PJ, Meale PM, Hurst RA, O'Doherty AF, Otto J, Kuper M et al. Cardiopulmonary exercise testing predicts postoperative outcome in patients undergoing gastric bypass surgery. *Br J Anaesth* 2012; 109(4):566-571.

(211) Ausania F, Vallance AE, Manas DM, Prentis JM, Snowden CP, White SA et al. Double bypass for inoperable pancreatic malignancy at laparotomy: postoperative complications and long-term outcome. *Ann R Coll Surg Engl* 2012; 94(8):563-568.

(212) Junejo MA, Mason JM, Sheen AJ, Moore J, Foster P, Atkinson D et al. Cardiopulmonary exercise testing for preoperative risk assessment before hepatic resection. *Br J Surg* 2012; 99(8):1097-1104.

(213) Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. *Health & Quality of Life Outcomes* 2008; 6:55.

(214) Browne J, Jamieson L, Lewsey J, van der Meulen J, Black N, Cairns J et al. Patient Reported Outcome Measures (PROMs) in Elective Surgery; Report to the Department of Health. 12-12-2007.

(215) Juul D. Using Rasch analysis to analyze performance on patient management problems. *Res Med Educ* 1988; 27:190-194.

(216) Hays RD, Kim S, Spritzer KL, Kaplan RM, Tally S, Feeny D et al. Effects of Mode and Order of Administration on Generic Health-Related Quality of Life Scores. *Value Health* 2009.

(217) Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006; 12(5):559-568.

(218) Darzi A. High Quality Care for All. *Department of Health* 2008.

(219) Black N. Patient Reported Outcome Measures: How is the PROMs programme progressing? *Lecture to the Royal Statistical Society* 2012.

(220) Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016.

(221) Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338.

(222) Aylin P, Bottle A, Azeem M. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007; 334.

(223) Kay P. Patient Reported Outcome Measurement Data. *Newsletter of the British Orthopaedic Association* 2011; 49:1-2.

(224) Zosia K. Surgeons condemn Lansley for misinformation on outcomes of knee operations. *BMJ* 2012; 344.

(225) Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and program planning* 1983; 6(3-4):185-210.

(226) Rubin HR, Ware JE, Jr., Hays RD. The PJHQ questionnaire. Exploratory factor analysis and empirical scale construction. *Med Care* 1990; 28(9 Suppl):S22-S29.

(227) Barnett SF, Alagar RK, Grocott MPW, Giannaris S, Dick JR, Moonesinghe SR. Patient satisfaction measures in anaesthesia: qualitative systematic review. *Anesthesiology* 2013; In press.

(228) Institute of Medicine. Evaluation of biomarkers and surrogate endpoints in chronic disease. *www iom edu/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease aspx* 2010.

(229) Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ* 2011; 343:d7995.

(230) Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, Rushakoff RJ et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006; 296(4):427-440.

(231) Albrecht E, Kirkham KR, Endersby RV, Chan VW, Jackson T, Okrainec A et al. Ultrasound-Guided Transversus Abdominis Plane (TAP) Block for

Laparoscopic Gastric-Bypass Surgery:a Prospective Randomized Controlled Double-Blinded Trial. *Obes Surg* 2013.

(232) Enko D, Wallner F, von-Goedecke A, Hirschmugl C, Auersperg V, Halwachs-Baumann G. The impact of an algorithm-guided management of preoperative anemia in perioperative hemoglobin level and transfusion of major orthopedic surgery patients. *Anemia* 2013; 2013:641876.

(233) Fisher DM. Surrogate outcomes: meaningful not! *Anesthesiology* 1999; 90(2):355-356.

(234) Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. *National Academy Press* 2001; 337.

(235) Cook TM, Coupe M, Ku T. Shaping quality: the use of performance polygons for multidimensional presentation and interpretation of qualitative performance data. *Br J Anaesth* 2012; 108(6):953-960.

(236) Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999; 16(1):9-13.

(237) Anderson ID. The Higher Risk General Surgical Patient: towards improved care for a forgotton group. *Royal College of Surgeons and the Department of Health* 2011.

(238) Adams ST, Leveson SH. Clinical prediction rules. *BMJ (Clinical research ed)* 2012; 344:d8312.

(239) Grobman WA, Stamilio DM. Methods of clinical prediction. *Am J Obstet Gynecol* 2006; 194(3):888-894.

(240) Hennis PJ, Meale PM, Grocott MP. Cardiopulmonary exercise testing for the evaluation of perioperative risk in non-cardiopulmonary surgery. *Postgrad Med J* 2011; 87(1030):550-557.

(241) Findlay GP, Goodwin APL, Protopapa K, Smith NCE, Mason M. Knowing the risk: a review of the perioperative care of surgical patients. *National Confidential Enquiry into Patient Outcome and Death* 2011.

(242) Liao L, Mark DB. Clinical prediction models: Are we building better mousetraps? *J Am Coll Cardiol* 2003; 42(5):851-853.

(243) Noble D, Dent T, Greenhalgh T. Re: Comparisons of established risk prediction models for cardiovascular disease: systematic review. (rapid response). *BMJ* 2012; 345:e4357.

(244) Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol* 2006; 59(8):765-769.

(245) Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001; 323(7306):224-228.

(246) Altman DG. Systematic reviews of evaluations of prognostic variables. *Systematic Reviews in Healthcare: Meta-analysis in Context* 2005; Chapter 13:228-247.

(247) Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338:b605.

(248) Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; 338:b606.

(249) Swets JA. Measuring the Accuracy of Diagnostic Systems. *Science* 1988; 240(4857):1285-1293.

(250) Arvidsson S, Ouchterlony J, Sjostedt L, Svardsudd K. Predicting postoperative adverse events. Clinical efficiency of four general classification systems - The project perioperative risk. *Acta Anaesthesiol Scand* 1996; 40(7):783-791.

(251) Ding L-A, Sun L-Q, Chen S-X, Qu L-L, Xie D-F. Modified physiological and operative score for the enumeration of mortality and morbidity risk assessment model in general surgery. *World J Gastroenterol* 2007;(38):5090-5095.

(252) Carneiro AV, Leitao MP, Lopes MG, De PF. [Risk stratification and prognosis in critical surgical patients using the Acute Physiology, Age and Chronic Health III System (APACHE III)]. *Acta Med Port* 1997; 10(11):751-760.

(253) Zhang H, Zhu D-M, Xue Z-G, Luo J-F, Jiang H. Performance of APACHE II models in surgical intensive care unit. [Chinese]. *Fudan Univ J Med Sci* 2004; 31(4):417-420.

(254) Saba V, Goffi L, Jassem W, Ghiselli R, Necozione S, Mattei A et al. Prognostic value of the Apache II scoring system daily preoperative use in major general surgery. [Italian]. *Chirurgia* 1997; 10(2):187-194.

(255) Martin Graczyk AI, Molina Hernandez MJ, Vazquez PC, Mora FJ, Hierro VM, Gomez PJ et al. [Preoperative geriatric assessment in major surgery in the aged]. [Spanish]. *Anales de Medicina Interna* 1995; 12(6):270-274.

(256) Kuo HS, Chuang JH, Tang GJ, Hou CC, Chou SS, Lui WY et al. Development of a new prognostic system and validation of APACHE II for surgical ICU mortality: a multicenter study in Taiwan. *Chung Hua i Hsueh Tsa Chih - Chin Med J* 1999; 62(10):673-681.

(257) Krenzien J, Roding H, Mummelthey R. [Surgical risk in old age. Prospective evaluation of a prognosis index]. *Zentralblatt fur Chirurgie* 1990; 115(12):717-727.

(258) Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13(10):818-829.

(259) Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Brit J Surg* 1996;(6):812-815.

(260) Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg* 2007; 204(2):201-208.

(261) Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994; 47(11):1245-1251.

(262) Jones DR, Copeland GP, de CL. Comparison of POSSUM with APACHE II for prediction of outcome from a surgical high-dependency unit. *Br J Surg* 1992; 79(12):1293-1296.

(263) Osler TM, Rogers FB, Glance LG, Cohen M, Rutledge R, Shackford SR. Predicting survival, length of stay, and cost in the surgical intensive care unit: APACHE II versus ICISS. *J Trauma Inj Inf Crit Care* 1998;(2):234-238.

(264) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40(5):373-383.

(265) Atherly A, Fink AS, Campbell DC, Mentzer RM, Jr., Henderson W, Khuri S et al. Evaluating alternative risk-adjustment strategies for surgery. *Am J Surg* 2004; 188(5):566-570.

(266) Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004; 57(12):1288-1294.

(267) Brooks MJ, Sutton R, Sarin S. Comparison of Surgical Risk Score, POSSUM and p-POSSUM in higher-risk surgical patients. *Br J Surg* 2005; 92(10):1288-1292.

(268) Organ N, Morgan T, Venkatesh B, Purdie D. Evaluation of the P-POSSUM mortality prediction algorithm in Australian surgical intensive care unit patients. *ANZ J Surg* 2002; 72(10):735-738.

(269) Neary WD, Prytherch D, Foy C, Heather BP, Earnshaw JJ. Comparison of different methods of risk stratification in urgent and emergency surgery. *Br J Surg* 2007; 94(10):1300-1305.

(270) Haga Y, Ikejiri K, Wada Y, Takahashi T, Ikenaga M, Akiyama N et al. A Multicenter Prospective Study of Surgical Audit Systems. *Ann Surg* 2011; 253(1):194-201.

(271) Regenbogen SE, Ehrenfeld JM, Lipsitz SR, Greenberg CC, Hutter MM, Gawande AA. Utility of the surgical apgar score: Validation in 4119 patients. *Arch Surg* 2009;(1):30-36.

(272) Haynes AB, Regenbogen SE, Weiser TG, Lipsitz SR, Dziekan G, Berry WR et al. Surgical outcome measurement for a global patient population: validation of the Surgical Apgar Score in 8 countries. *Surgery* 2011; 149(4):519-524.

(273) Goffi L, Saba V, Ghiselli R, Necozione S, Mattei A, Carle F. Preoperative APACHE II and ASA scores in patients having major general surgical operations: prognostic value and potential clinical applications. *Eur J Surg* 1999; 165(8):730-735.

(274) Hightower CE, Riedel BJ, Feig BW, Morris GS, Ensor JE, Jr., Woodruff VD et al. A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: Physiological capacity compared with the ASA physical status classification system. *Br J Anaes* 2010; 104(4):465-471.

(275) Makary MA, Segev DL, Pronovost PJ, Syin D, Bandeen-Roche K, Patel P et al. Frailty as a predictor of surgical outcomes in older patients. *J Am Coll Surg* 2010; 210(6):901-908.

(276) Dasgupta M, Rolfson DB, Stolee P, Borrie MJ, Speechley M. Frailty is associated with postoperative complications in older adults with medical problems. *Arch Gerontol Geriatr* 2009; 48(1):78-83.

(277) Davenport DL, Bowe EA, Henderson WG, Khuri SF, Mentzer RM, Jr. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists Physical Status Classification (ASA PS) levels. *Ann Surg* 2006; 243(5):636-641.

(278) Hadjianastassiou VG, Tekkis PP, Poloniecki JD, Gavalas MC, Goldhill DR. Surgical mortality score: risk management tool for auditing surgical performance. *World J Surg* 2004; 28(2):193-200.

(279) Haynes AB, Regenbogen SE, Weiser TG, Lipsitz SR, Dziekan G, Berry WR et al. Surgical outcome measurement for a global patient population: validation of the Surgical Apgar Score in 8 countries. *Surgery* 2011; 149(4):519-524.

(280) Hobson SA, Sutton CD, Garcea G, Thomas WM. Prospective comparison of POSSUM and P-POSSUM with clinical assessment of mortality following emergency surgery. *Acta Anaesthesiol Scand* 2007;(1):94-100.

(281) Kuzu MA, Terzioglu H, Genc V, Erkek AB, Ozban M, Sonyurek P et al. Preoperative nutritional risk assessment in predicting postoperative outcome in patients undergoing major surgery. *World J Surg* 2006; 30(3):378-390.

(282) Liebman B, Strating RP, van WW, Mulder W, Oomen JL, Engel AF. Risk modelling of outcome after general and trauma surgery (the IRIS score). *Brit J Surg* 2010; 97(1):128-133.

(283) Nathanson BH, Higgins TL, Kramer AA, Copes WS, Stark M, Teres D. Subgroup mortality probability models: Are they necessary for specialized intensive care units? *Crit Care Med* 2009; 37(8):2375-2386.

(284) Pillai SB, Van Rij AM, Williams S, Thomson IA, Putterill MJ, Greig S. Complexity- and risk-adjusted model for measuring surgical outcome. *Br J Surg* 1999;(12):1567-1572.

(285) Stachon A, Becker A, Kempf R, Holland-Letz T, Friese J, Krieg M. Re-evaluation of established risk scores by measurement of nucleated red blood cells in blood of surgical intensive care patients. *J Trauma Inj Inf Crit Care* 2008;(3):666-673.

(286) Stachon A, Becker A, Holland-Letz T, Friese J, Kempf R, Krieg M. Estimation of the mortality risk of surgical intensive care patients based on routine laboratory parameters. *Eur Surg Res* 2008;(3):263-272.

(287) Story DA, Fink M, Leslie K, Myles PS, Yap SJ, Beavis V et al. Perioperative mortality risk score using pre- and postoperative risk factors in older patients. *Anaesth Intensive Care* 2009; 37(3):392-398.

(288) Aust JB, Henderson W, Khuri S, Page CP, Jones RS, Jurkiewicz MJ et al. The impact of operative complexity on patient risk factors. *Ann Surg* 2005;(6):1024-1028.

(289) Copeland GP, Sagar P, Brennan J, Roberts G, Ward J, Cornford P et al. Risk-adjusted analysis of surgeon performance: A 1-year study. *Brit J Surg* 1995;(3):408-411.

(290) Covinsky KE, Martin GE, Beyth RJ, Justice AC, Sehgal AR, Landefeld CS. The relationship between clinical assessments of nutritional status and adverse outcomes in older hospitalized medical patients. *J Am Geriatr Soc* 1999; 47(5):532-538.

(291) Wolinsky FD, Coe RM, Chavez MN, Prendergast JM, Miller DK. Further assessment of the reliability and validity of a Nutritional Risk Index: analysis of a three-wave panel study of elderly adults. *Health Serv Res* 1986; 20(6 Pt 2):977-990.

(292) Ensrud KE, Ewing SK, Taylor BC, Fink HA, Cawthon PM, Stone KL et al. Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women. *Arch Intern Med* 2008; 168(4):382-389.

(293) Lee DH, Buth KJ, Martin BJ, Yip AM, Hirsch GM. Frail Patients Are at Increased Risk for Mortality and Prolonged Institutional Care After Cardiac Surgery. *Circulation* 2010; 121(8):973-978.

(294) Whitson HE, Purser JL, Cohen HJ. Frailty thy name is ... Phrailty? *J Gerontol A Biol Sci Med Sci* 2007; 62(7):728-730.

(295) Oliveira M, Fogaca K, Leandro-Merhi V. Nutritional status and functional capacity of hospitalized elderly. *Nutrition Journal* 2009; 8(1):54.

(296) Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. *Annals of Surgery* 2010; 251(3):535-541.

(297) Haynes SR, Lawler PGP. An Assessment of the Consistency of Asa Physical Status Classification Allocation. *Anaesthesia* 1995; 50(3):195-199.

(298) Grocott MP, Levett DZ, Matejowsky C, Emberton M, Mythen MG. ASA scores in the preoperative patient: feedback to clinicians can improve data quality. *J Eval Clin Pract* 2007; 13(2):318-319.

(299) Aronson WL, McAuliffe MS, Miller K. Variability in the American Society of Anesthesiologists Physical Status Classification Scale. *AANA J* 2003; 71(4):265-274.

(300) Mak PHK, Campbell RCH, Irwin MG. The ASA physical status classification: Inter-observer consistency. *Anaesth Intensive Care* 2002; 30(5):633-640.

(301) Robinson PJ, Wilson D, Coral A, Murphy A, Verow P. Variation between experienced observers in the interpretation of accident and emergency radiographs. *Br J Radiol* 1999; 72(856):323-330.

(302) Trzeciak S, Erickson T, Bunney EB, Sloan EP. Variation in patient management based on ECG interpretation by emergency medicine and internal medicine residents. *Am J Emerg Med* 2002; 20(3):188-195.

(303) Hall BL, Hirbe M, Waterman B, Boslaugh S, Dunagan WC. Comparison of Mortality Risk Adjustment Using a Clinical Data Algorithm (American College of Surgeons National Surgical Quality Improvement Program) and an Administrative Data Algorithm (Solucient) at the Case Level Within a Single Institution. *J Am Coll Surg* 2007;(6):767-777.

(304) Copeland GP. The POSSUM system of surgical audit. *Arch Surg* 2002; 137(1):15-19.

(305) Tilford JM, Roberson PK, Lensing S, Fiser DH. Differences in pediatric ICU mortality risk over time. *Crit Care Med* 1998; 26(10):1737-1743.

(306) Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007; 35(9):2052-2056.

(307) Parikh P, Shiloach M, Cohen ME, Bilimoria KY, Ko CY, Hall BL et al. Pancreatectomy risk calculator: an ACS-NSQIP resource. *HPB* 2010; 12(7):488-497.

(308) Gupta PK, Franck C, Miller WJ, Gupta H, Forse RA. Development and validation of a bariatric surgery morbidity risk calculator using the prospective, multicenter NSQIP dataset. *Journal of the American College of Surgeons* 2011; 212(3):301-309.

(309) Cohen ME, Bilimoria KY, Ko CY, Hall BL. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg* 2009; 208(6):1009-1016.

(310) Gupta PK, Gupta H, Sundaram A, Kaushik M, Fang X, Miller WJ et al. Development and Validation of a Risk Calculator for Prediction of Cardiac Risk After Surgery / Clinical Perspective. *Circulation* 2011; 124(4):381-387.

(311) Grocott MP. Improving outcomes after surgery. *BMJ* 2009; 339:b5173.

(312) Beattie WS, Karkouti K, Wijeysundera DN, Tait G. Risk associated with preoperative anemia in noncardiac surgery: a single-center cohort study. *Anesthesiology* 2009; 110(3):574-581.

(313) Haga Y, Ikejiri K, Wada Y, Takahashi T, Ikenaga M, Akiyama N et al. A multicenter prospective study of surgical audit systems. *Annals of Surgery* 2011; 253(1):194-201.

(314) General Medical Council. Consent: Patients and Doctors making decisions together. *http://www gmc-uk org/static/documents/content/Consent_0510 pdf* 2008.

(315) Goldman L, Caldera DL, Nussbaum SR, Southwick FS, Krogstad D, Murray B et al. Multifactorial Index of Cardiac Risk in Noncardiac Surgical Procedures. *N Engl J Med* 1977; 297(16):845-850.

(316) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: Results of the National Veterans Affairs Surgical Risk Study. *Journal of the American College of Surgeons* 1997; 185(4):315-327.

(317) Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: Results of the National Veterans Affairs Surgical Risk Study. *Journal of the American College of Surgeons* 1997; 185(4):328-340.

(318) Fleischmann KE, Goldman L, Young B, Lee TH. Association between cardiac and noncardiac complications in patients undergoing noncardiac surgery: outcomes and effects on length of stay. *Am J Med* 2003; 115(7):515-520.

(319) Badner NH, Knill RL, Brown JE, Novick TV, Gelb AW. Myocardial infarction after noncardiac surgery. *Anesthesiology* 1998; 88(3):572-578.

(320) Olson MM, Lee J. Continuous, 10-Year Wound Infection SurveillanceResults, Advantages, and Unanswered Questions. *Arch Surg* 1990; 125(6):794-803.

(321) Riou JP, Cohen JR, Johnson J. Factors influencing wound dehiscence. *The American Journal of Surgery* 1992; 163(3):324-330.

(322) Giacometti A, Cirioni O, Schimizzi AM, Del Prete MS, Barchiesi F, D'Errico MM et al. Epidemiology and microbiology of surgical wound infections. *J Clin Microbiol* 2000; 38(2):918-922.

(323) Hein OV, Birnbaum J+, Wernecke K, England M, Konertz W, Spies C. Prolonged Intensive Care Unit Stay in Cardiac Surgery: Risk Factors and Long-Term-Survival. *The Annals of Thoracic Surgery* 2006; 81(3):880-885.

(324) Mahesh B, Choong CK, Goldsmith K, Gerrard C, Nashef SAM, Vuylsteke A. Prolonged Stay in Intensive Care Unit Is a Powerful Predictor of Adverse Outcomes After Cardiac Operations. *The Annals of Thoracic Surgery* 2012; 94(1):109-116.

(325) Wind J, Polle SW, Fung Kon Jin PHP, Dejong CHC, von Meyenfeldt MF, Ubbink DT et al. Systematic review of enhanced recovery programmes in colonic surgery. *Br J Surg* 2006; 93(7):800-809.

(326) Leon DA. Trends in European life expectancy: a salutary view. *International Journal of Epidemiology* 2011; 40(2):271-277.

(327) Hernan MA. The hazards of hazard ratios. *Epidemiology* 2010; 21(1):13-15.

(328) Chok KS, Ng KK, Poon RT, Lo CM, Fan ST. Impact of postoperative complications on long-term outcome of curative resection for hepatocellular carcinoma. *Br J Surg* 2009; 96(1):81-87.

(329) Manku K, Bacchetti P, Leung JM. Prognostic significance of postoperative in-hospital complications in elderly patients. I. Long-term survival. *Anesth Analg* 2003; 96(2):583-589.

(330) Rueth NM, Parsons HM, Habermann EB, Groth SS, Virnig BA, Tuttle TM et al. The long-term impact of surgical complications after resection of stage I nonsmall cell lung cancer: a population-based survival analysis. *Ann Surg* 2011; 254(2):368-374.

(331) Xia BT, Rosato EL, Chojnacki KA, Crawford AG, Weksler B, Berger AC. Major Perioperative Morbidity Does Not Affect Long-Term Survival in Patients Undergoing Esophagectomy for Cancer of the Esophagus or Gastroesophageal Junction. *World J Surg* 2012.

(332) Tokunaga M, Tanizawa Y, Bando E, Kawamura T, Terashima M. Poor Survival Rate in Patients with Postoperative Intra-Abdominal Infectious Complications Following Curative Gastrectomy for Gastric Cancer. *Ann Surg Oncol* 2012.

(333) Mrak K, Eberl T, Laske A, Jagoditsch M, Fritz J, Tschmelitsch J. Impact of Postoperative Complications on Long-term Survival After Resection for Rectal Cancer. *Diseases of the Colon & Rectum* 2013; 56(1).

(334) Kamphues C, Bova R, Schricke D, Hippler-Benscheidt M, Klauschen F, Stenzinger A et al. Postoperative Complications Deteriorate Long-Term Outcome in Pancreatic Cancer Patients. *Annals of Surgical Oncology* 2012; 19(3):856-863.

(335) Laurent C, Sa Cunha A, Couderc P, Rullier E, Saric J. Influence of postoperative morbidity on long-term survival following liver resection for colorectal metastases. *Br J Surg* 2003; 90(9):1131-1136.

(336) Vollset SE, Tverdal A, Gjessing HK. Smoking and Deaths between 40 and 70 Years of Age in Women and Men. *Annals of Internal Medicine* 2006; 144(6):381-389.

(337) Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ et al. Body-Mass Index and Mortality among 1.46 Million White Adults. *N Engl J Med* 2010; 363(23):2211-2219.

(338) Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009; 361(14):1368-1375.

(339) Merkow RP, Bilimoria KY, Cohen ME, Richards K, Ko CY, Hall BL. Variability in Reoperation Rates at 182 Hospitals: A Potential Target for Quality Improvement. *Journal of the American College of Surgeons* 2009; 209(5):557-564.

(340) Stachler RJ, Yaremchuk K, Ritz J. Preliminary NSQIP results: A tool for quality improvement. *Otolaryngology - Head and Neck Surgery* 2010; 143(1):26-30.

(341) de Vries EN, Prins HA, Crolla RM, den Outer AJ, van AG, van Helden SH et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med* 2010; 363(20):1928-1937.

(342) Kehlet H. Fast-track colorectal surgery. *Lancet* 2008; 371(9615):791-793.

(343) Desai SV, Law TJ, Needham DM. Long-term complications of critical care. *Critical Care Medicine* 2011; 39(2):371-379.

(344) Derogar M, Orsini N, Sadr-Azodi O, Lagergren P. Influence of Major Postoperative Complications on Health-Related Quality of Life Among Long-Term Survivors of Esophageal Cancer Surgery. *Journal of Clinical Oncology* 2012.

(345) Anthony T, Long J, Hynan LS, Sarosi GA, Nwariaku F, Huth J et al. Surgical complications exert a lasting effect on disease-specific health-related quality of life for patients with colorectal cancer. *Surgery* 2003; 134(2):119-125.

(346) Bennett-Guerrero E, Hyam JA, Shaefi S, Prytherch DR, Sutton GL, Weaver PC et al. Comparison of P-POSSUM risk-adjusted mortality rates after surgery between patients in the USA and the UK. *Br J Surg* 2003; 90(12):1593-1598.

(347) Vonlanthen R, Clavien PA. What factors affect mortality after surgery? *Lancet* 2012; 380(9847):1034-1036.

(348) Hollowell J, Grocott MP, Hardy R, Haddad FS, Mythen MG, Raine R. Major elective joint replacement surgery: socioeconomic variations in surgical risk, postoperative morbidity and length of stay. *J Eval Clin Pract* 2010; 16(3):529-538.

(349) Woods LM, Rachet B, Riga M, Stone N, Shah A, Coleman MP. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. *Journal of Epidemiology and Community Health* 2005; 59(2):115-120.

(350) Cook TM, Counsell D, Wildsmith JA. Major complications of central neuraxial block: report on the Third National Audit Project of the Royal College of Anaesthetists. *Br J Anaesth* 2009; 102(2):179-190.

(351) Woodall NM, Cook TM. National census of airway management techniques used for anaesthesia in the UK: first phase of the Fourth National Audit Project at the Royal College of Anaesthetists. *Br J Anaesth* 2011; 106(2):266-271.