Implementation of a Structured Surgical Quality Improvement Programme

Richard Trafford Spence MBChB (Hons) MPhil

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfillment of the requirements for the degree
Doctor of Philosophy in Surgery

SUPERVISORS

Professor David Chang MPH MBA PhD, Eugenio Panieri MBChB FCS (SA), Andrew Nicol MBChB FCS (SA) PhD

“Comparisons are odious, but comparison is necessary in science.”

Ernest Amory Codman

Acknowledgements

This work would not have been possible without the support and guidance from a number of colleagues, friends and family members, who I would like to acknowledge. To the Department of General Surgery at Groote Schuur Hospital, thank you for your on-going teaching of the art of Surgery, your transparency in benchmarking our outcomes, the vote of confidence in letting me pursue this research and for supporting my leave of absence. Professor Del Kahn, your passion for teaching and dedication to Surgery is an inspiration. Professor Pradeep Navasaria, the uncompromising integrity with which you conduct your practice has been an example to strive towards. Professor Eugenio Panieri, thank you for your support from the start of this initiative, for always looking for avenues to promote my work and progress my career, as well as, your mentorship and supervision throughout my research. Professor Andrew Nicol, thank you for the opportunity to pursue this PhD, your supervision, and your support with the many applications it required. Sr. Colleen Sampson, thank you for your drive and dedication to improving the care we offer our patients and your contribution to the programme. Thank you to my many lecturers at the Harvard TH Chan School of Public Health for your pioneering work, dedication to teaching and granting me the privilege to pursue the Program of Clinical Effectiveness. A particular thanks to Dr. Fran Cook and Dr. John Orav for their many hours of statistical support. To my hosts and co-supervisors at the Codman Center for Clinical Effectiveness in Surgery, Professor Matt Hutter and Professor David Chang, thank you for your willingness to support

my work, for your significant contributions in the field of surgical outcomes research, providing me access to your national databases and for teaching me what is possible with sound research. To Dr. Kat Chu, for your critical review of my writing. To my sponsors, Discovery Health and Netcare, thank you for the opportunity to pursue this research. To my friends and colleagues at the University of British Columbia, Dr. Morad Hameed, Dr. Damon Ramsey and Dr. Eiman Zagaran, thank you for your pioneering work and for your support with mine. My parents, Rod and Lynne Spence, for providing me with every opportunity to pursue my passions. To my wife, Ingrid Spence, for your uncompromising support in every endeavour we pursue together. To our patients, for entrusting us to care for them. Finally, to the "father of outcomes measurement", Dr. Ernest Amory Codman, thank you.

Contents

Foreword

As surgery assumes a greater position in the global health agenda, the need to not only improve access to surgical care but also improve the quality of surgical care, is paramount. Surgical quality improvement programmes have been shown to reduce morbidity and mortality following surgery. A key first step to the design and implementation of a structured surgical quality improvement programme is the collection and analysis of high-quality data. To quote Dr. Margaret Chan, the director general of the World Health Organisation,

'…the real need (in global health) is to close the data gaps, especially in low and middle-income countries, so that we no longer have to rely heavily on statistical modeling for data on disease burden.'

In this thesis it was hypothesised that emerging m-Health technology, defined as medical and public health practices supported by the use of mobile devices, would provide a solution to close such data gaps.

Various m-Health applications were used to develop three databases, describing the outcomes of major surgery performed within the Cape Metro West health district during the study period. After reviewing the design and analytical rationale of the American College of Surgeons National Surgical Quality Improvement Programme and Trauma Quality Improvement Programme, these *de novo* databases were used to develop three quality improvement programmes, designed for local implementation: The Essentials programme for general and vascular surgery, a Procedure-targeted programme and a Trauma quality improvement programme.

Key to these programmes was the derivation and validation of prediction rules, which reliably estimate the probability of an adverse outcome following major surgery in a risk-adjusted manner. Such rules promote internal and external benchmarking over time to identify opportunities for quality improvement and critically appraise the impact of any corrective action implemented. In order to improve the quality of surgical care we provide, a continuous cycle of monitoring, assessment, and management should be routinely performed. This thesis provides some guidance on how this can be done within the Cape Metro West health district.

Glossary of abbreviated terms

Abbreviation	Definition
ACS-NSQIP	American College of Surgeons National Surgical Quality Improvement Programme
ACS-TQIP	American College of Surgeons Trauma Quality Improvement Programme
AIDS	Acquired Immune Deficiency Syndrome
AIS	Abbreviated Injury Score
ASA	American Society of Anaesthesia
COT	Committee of Trauma
CPT	Current procedural Terminology
CRM	Customer Relations Manager
eTHR	electronic Trauma Health Record
DALY	Disability Adjusted Life Years
EuSOS	European Surgical Outcomes Study
GBD	Global Burden of Disease
GOF	Goodness of Fit
GSH	Groote Schuur Hospital
GSTS	Groote Schuur Trauma Score
GSSRS	Groote Schuur Surgical Risk Score
HIV	Human Immunodeficiency Virus
IATSIC	International Association for Trauma and Intensive Care
ICD	International Classification of Disease
ISS	Injury Severity Score
LMIC	Low and middle-income countries
MPH	Mitchell's Plain District Hospital
MTOS	Multiple Trauma Outcomes Study
NTDB	National Trauma Data Bank
NSH	New Somerset Hospital
NVASRS	National Veterans Association Surgical Risk Study
PSS	Patient Safety in Surgery Study
ROC	Receiver Operating Characteristic/ Area under Receiver Operating Curve
RTS	Revised Trauma Score
RVU	Relative Value Unit
SASOS	South African Surgical Outcomes Study
TRISS	Trauma Score-Injury Severity Score
UCTSRS	University of Cape Town Surgical Risk Score
VAMC	Veterans Affairs Medical Centers
VGH	Vancouver General Hospital
VHA	Veterans Health Administration
VWH	Victoria War Memorial Hospital
WHO EES	World Health Organisation Emergency and Essential Services

Chapter 1

Executive summary

Measurement alone may improve outcomes - the so-called 'Hawthorne effect'. The framework defined by Avedis Donabedian to measure quality of medical care involves three concepts – structure, process, and outcomes.

In surgery, outcomes are the easiest to measure and the results are easily understood by surgeons. The main limitation of the use of outcome in the comparative assessment of the quality of surgical care, however, is the need to use adequate and validated models for risk-adjustment. The pre-surgical severity of illness must be adjusted, if outcome is to be used in the comparative assessment of the quality of surgical care.

To date, in the United States of America, the American College of Surgeons' National Surgical Quality Improvement Programme (ACS-NSQIP) remains the most robust risk-adjusted and reliable tool available to test the quality of surgical care. The ACS-NSQIP has been shown to reduce both morbidity and mortality in enrolled hospitals with initially worse performing hospitals having a greater likelihood of improvement. Following the success of the ACS-NSQIP, the American College of Surgeons Trauma Quality Improvement Programme (ACS-TQIP) was developed. External benchmarking, which allows direct inter-hospital performance comparisons, has been the cornerstone of

such programmes. However, a significant limitation is the need to retrospectively collect over 130 variables per patient, thereby limiting their generalisability to low-to-middle income countries (LMICs).

The primary aim of this thesis, to derive and validate prediction rules, which reliably predict the risk of mortality and morbidity following major surgery in the Cape Metro West health district in South Africa for benchmarking and quality improvement initiatives. It was hypothesised that emerging m-Health technology, defined as medical and public health practices supported by the use of mobile devices, would provide a solution to mitigate the paucity of reliable surgical outcomes research and surgical quality improvement programmes in LMICs.

Using various m-Health applications in general surgery, the findings of this thesis suggested that a patient undergoing a general surgery or vascular operation at Groote Schuur Hospital (GSH), was twice as likely to experience a major complication or almost ten times as likely to spend longer than 2 weeks in hospital compared to a patient in an average performing hospital in the ACS-NSQIP consortium, when controlling for a number of confounders. A patient undergoing an emergency exploratory laparotomy at a hospital in the Cape Metro West health district in South Africa, had a 10% increased risk of experiencing a major complication and a 20% increased risk of spending longer than 2 weeks in hospital, compared to, the standards set by the ACS-NSQIP.

In trauma, co-efficient based injury severity scores were used to compare the outcomes of all admissions, including severely injured and operatively managed patients, against what would be expected from the Multiple Trauma Outcomes Study (MTOS).

There was no difference in outcomes for the whole cohort and operatively managed patients between GSH and the expected MTOS data, according to the Trauma Revised Injury Severity Score (TRISS). However, according to both the TRISS and Revised Trauma Score (RTS), severely injured patients admitted to GSH were more likely to die than a participant of the MTOS. In contrast, according to the RTS, patients admitted to GSH who required an operation were 50% less likely to die compared to participants of the MTOS.

These findings further support the growing evidence that surgical outcomes vary by provider and that there are significant opportunities for quality improvement. Implementation of a structured surgical quality improvement programme within the Cape Metro West health district, in South Africa, is therefore warranted. Following the findings of this thesis, such a programme should consider the following recommendations:

1. Align surgical quality initiatives with the development of research collaboratives
2. Collect the proposed minimum dataset for general and vascular surgery
3. Collect the proposed minimum dataset for trauma
4. Use the derived and validated simple scoring algorithms to calculate risk-adjusted outcome predictions for pre-operative informed consent and resource allocation
5. Perform internal and external benchmarking using a meaningful objective surgical quality metric: the Observed to Expected (O/E) ratio

6. Adopt the proposed speciality-specific, procedure-specific programme at GSH
7. Use the proposed Codman Risk Calculator in the Cape Metro West surgical collaborative
8. Adopt the unexpected outcomes approach to the Cape Metro West Morbidity & Mortality meeting.

Implementing these recommendations will close the gap between data acquisition, analysis and action.

Chapter 2

Literature review

2.1 Surgery on the global agenda

In January 2014, Jim Kim, President of the World Bank, urged the global health community to challenge the injustice of global inequity in surgical care, stating that, "surgery is an indivisible, indispensable part of health care and of progress towards universal health coverage" (1).

During the past ten years, there has been increasing interest in defining the role of surgical care amongst other global health priorities (2). Surgery has been referred to as, the ''neglected stepchild'' of global health and considered amongst other ''neglected diseases''(3). It is now generally accepted, that there exists globally a significant burden of disease that requires surgical intervention; 234 million operations are currently performed each year (4). However, there are currently gross disparities in access to safe, essential surgical care worldwide, and an alarming lack of global focus on widespread provision of quality surgical services (1-4).

When analysed on a country level, the disparity between higher health expenditure countries and poor countries remains striking: 96.5% of all surgical procedures are performed on 4 billion people, whereas only 3.5% of all surgical procedures are performed on the poorest 2 billion people (4). In addition, 11-15 % of the world's disability is due to surgically treatable conditions (5). In 2010, an estimated 16.9

million lives (32.9% of all deaths worldwide) were lost from conditions, which needed surgical care. This well surpassed the number of deaths from HIV/AIDS (1.46 million), tuberculosis (1.20 million), and malaria (1.17 million) combined. Injuries alone cause 5.7 million deaths annually (6).

Despite these disparities, surgery has been identified as a cost effective intervention in resource-poor settings, on a par with vaccination programmes and 10-15 times more so than antiretroviral medication for HIV (5). This is not to say that surgery is any more important than other types of treatment, but surgery has a crucial role to play. Lack of treatment puts a significant economic burden on the millions of people who cannot work or function, due to conditions for which the treatment has been known for decades (6).

The World Health Organisation programme for emergency surgical, obstetric, and anaesthetic care (WHO EES) (7), was launched in December of 2005, to address the global disparities which exist in essential surgical services. The objective was to reduce death and disability from trauma, burns, pregnancy-related complications, domestic violence, disasters and other surgically treatable conditions. To date, the programme has been rolled out to over 35 countries. Recently, the Lancet Journal commissioned a group of experts in a variety of disciplines pertaining to Global Surgery to generate a collaborative effort to advance this rapidly growing field. The purpose of '*The Lancet Commission on Global Surgery*, was to make this vision a reality by embedding surgery within the global health agenda, catalysing political change, and defining scalable solutions for the provision of quality surgical and anaesthesia care for all (6).'

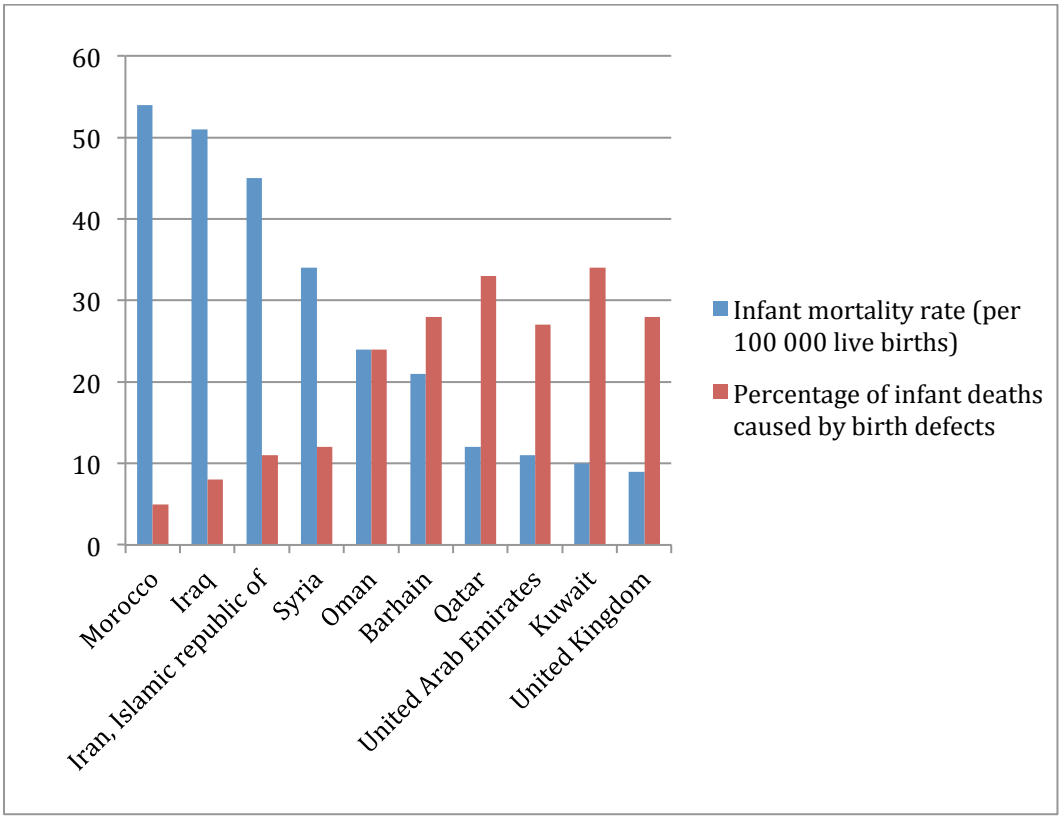
Subsequent to these global efforts, a new era of Global Health has begun, with the focus on broad-based health systems solutions, and the need to allocate resources accordingly. Surgical care has an incontrovertible, cross-cutting role in the achievement of local and global health challenges. As countries struggle to create comprehensive health care packages, more research is needed to determine the role of surgery in addressing national disease burdens (8). South Africa is no exception.

2.2 The epidemiological transition facing South Africa

In Omran's original description of the 'epidemiological transition', countries moved from high mortality, due largely to malnutrition and infectious disease to steadily declining mortality, peaking at older ages due largely to non-communicable diseases (9). The epidemiological transition occurs, as a country undergoes the process of modernisation from developing nation to developed nation status. In the 2010 GBD study, global deaths from non-communicable diseases rose by just under 8 million, between 1990 and 2010, accounting for two out of every three deaths (10).

The decline in infant and childhood mortality rates, which occurred in most countries in the 20th century, is a public health triumph (11). This health transition, is initially related to a decline in infant and 'under 5-years' mortality from infectious diseases and malnutrition (11). As illustrated in figure 2.1, there is an inverse relationship between infant mortality and the proportionate contribution of birth defects to infant mortality. Thus, birth defects increase in public health importance, as a country develops and moves through its health transition (11).

Figure 2.1 Relationship between infant mortality rate and percentage of infant deaths due to birth defects by country



Source: Adapted from WHO, 1997

In South Africa, the impact of the 'epidemiological transition' has lagged behind high-income countries by about 40 years. This has been related to the tragedy of the HIV/AIDS epidemic, and traditional views of the linear 'epidemiological transition' are, probably, too simplistic. This HIV/AIDS epidemic created what was then termed a 'counter-transition'. With the recent decline in deaths secondary to HIV, this counter-transition has unfolded, and resulted in a greater proportion of summed probability of disease-specific mortality to non-communicable diseases, than previously described (12).

The global burden of disease amenable to surgical intervention, such as, trauma, cancer, birth defects and complications from childbirth is significant and growing. South Africa is not free from this burden as this health transition is experienced. Surgery and anaesthesia must be an integral and indivisible component of a properly functioning health system in a developing country, and all people should have access to safe, high quality, affordable surgical and anaesthesia care. Whilst global interventions, such as, the WHO EES and the Lancet Commission of Global Surgery are catalysts to embed surgery within national health agendas, the quality and outcomes of surgical care must be measured at both a national and institutional level. Unfortunately, the expertise and knowledge of how to do this is lacking in South Africa, and this has to change.

2.3 A comparison of perioperative mortality rates

For the majority of patients the risks of surgery are low. However, it is well known that complications after surgery are an important cause of death (13). About 10% of patients undergoing surgery in the UK are at high risk of developing complications after surgery and account for 80% of post-operative deaths. If this rate is applicable worldwide, up to 25 million patients will undergo high-risk surgical procedures each year, of which 3 million will not survive until hospital discharge (14). This burden is likely to be much higher in developing countries, but clinical outcomes following major surgery are poorly described at a national and even institutional level.

International comparative data may provide important insights into the delivery of health care for surgical patients. The European Surgical Outcomes Study (EuSOS) was conducted, with the primary objective to describe mortality rates, and patterns of critical care resource usage for patients undergoing non-cardiac surgery across several European

countries. The EuSOS took place as a 7-day cohort study, between April 4 and April 11 2011. Data was collected, describing consecutive patients, aged 16 years and older, undergoing in-patient, non-cardiac surgery in 498 hospitals across 28 European countries. Patients were followed up for a maximum of 60-days. The primary endpoint was in-hospital mortality. Secondary outcome measures were duration of hospital stay and admission to critical care. Multilevel logistic regression models were constructed to adjust for the differences in mortality rates between countries (14).

The results of EuSOS showed mortality rates varied widely between countries (from 1·2% [95% CI 0·0–3·0] for Iceland to 21·5% [16·9–26·2] for Latvia). After adjustment for confounding variables, important differences remained between countries when compared with the UK, the country with the largest dataset (OR range from 0·44 [95% CI 0·19–1·05; p=0·06] for Finland to 6·92 [2·37–20·27; p=0·0004] for Poland). This meant that after adjustment for variations in perioperative factors, a patient was up to 7 times more likely to die post-operatively, simply, because of the location of the surgery (14).

Following the findings of this landmark study, in the emerging field of Surgical Outcomes research, EuSOS provided the impetus for Biccard *et al* to conduct a similar 7-day cohort study. This study was conducted during 2014, in South Africa, and was known as the South African Surgical Outcomes Study (SASOS) (15). The primary outcome was in-hospital mortality. Secondary outcomes, included duration of hospital stay, rate of admission to critical care following surgery, and the duration of critical care stay. In order to understand the proportional contribution of communicable diseases, non-communicable diseases and injuries to surgical outcomes, these variables were also

added to the adopted EuSOS dataset. Risk factors associated with in-hospital mortality and critical care admissions were also explored. In the SASOS study, the average crude in-hospital mortality rate was 123/3927 (3.1%, 95% CI 2.6 to 3.7). The number of post-operative admissions to critical care was 255/3927 (6.5%, 95% CI 5.7 to 7.3), with 43.5% being unplanned admissions. Urgent or emergent surgery occurred in 2120/3915 (54.2%), with a population attributable risk for mortality of 25.5% (95% CI 5.1 to 55.8) and for admission to critical care of 23.7% (95% CI 4.7 to 51.4).

The SASOS study concluded, that most operations in South Africa were urgent and emergent surgery, which was strongly associated with an increased mortality and unplanned critical care admission. Non-communicable diseases had a larger proportional contribution to mortality than communicable diseases and injuries (15). Based on the estimates of surgical volumes in South Africa by Weiser *et al* (16), the population statistics of South Africa for 2013 (17), and the SASOS data (15), the estimated mortality rate for adult surgery (≥20 years of age) in South Africa would be between 76 and 128 deaths per 100,000, which is equivalent to 7.2% and 12.1% of all deaths in South Africa. These estimates provide, a substantive supporting argument for a more in-depth analysis into the quality of surgical care provided in South Africa. They provide the rationale to further explore what perioperative factors, modifiable or not, predict an adverse outcome and more extensive resource use.

2.4 Measuring quality in surgery: structure, process, or outcomes

a. Definitions and basics of Quality Improvement

'Quality of care' is defined as, "the degree to which health services for individuals

and populations increase the likelihood of desired health outcomes, and are consistent with the current professional knowledge." (18).

'Quality Improvement ('QI') can be defined as, "the optimisation of resources, including knowledge, practical skills and material assets to produce good health."

QI is a method of improving medical care by monitoring the elements of diagnosis, treatment and outcome. QI involves, both prospective and retrospective review, and is aimed at improved outcomes – measuring the current status and figuring out ways to provide better care. QI evaluates the performance of both individual providers and the systems within which they work (18). Assessment and monitoring of quality in health care, has evolved considerably over the past 100 years and has been given many different names and associated acronyms (Table 2.1).

Table 2.1 The evolution of terminology for Quality Improvement

Timeline	Term	Definition
1900s	Medical Audit (MA)	A detailed *review* and evaluation of selected *clinical records* by qualified professional personnel for evaluating quality of *medical care*
1920s	Quality Assurance (QA)	A planned and systematic set of activities undertaken to ensure that standards and procedures are adhered to and that delivered products or services meet performance requirements.
1980s	Total Quality Management (TQM)	An organizational management approach that consists of making all individuals responsible for improving the quality of health care. The TQM approach to quality assurance emphasizes continuous product or service improvements through involvement of the workforce.
	Continuous Quality Improvement (CQI)	A management approach to improving and maintaining quality that emphasizes internally driven and relatively continuous assessments of potential causes of quality defects, followed by action aimed either at avoiding decrease in quality or at correcting it at an early stage.
1990s	Performance Improvement (PI)	The continuous evaluation of a system and the providers through structured review of the process of care as well as outcomes. PI has evolved from previous quality assurance paradigms and represents a more scientific and evidence-based continuation of those standards.
2000s	Quality Improvement (QI)	A method of evaluating and improving process of patient care which emphasizes a multidisciplinary approach to problem-solving and which focuses not on individuals but on systems of patients care that may be the cause of variations. QI consists of periodic scheduled evaluation of organizational activities, policies, procedures and performance to identify best practices and target areas in need of improvement and includes implementation of corrective actions or policy changes where needed.

Source: Adapted from WHO, Guidelines for trauma quality improvement programs (19)

During this time period from the 1900's to the 2000's, there has been a shift in the mindset with which the medical community has approached the topic of quality improvement. One of the original strategies introduced, to improve the quality of medical care, termed the "medical audit", originated in the 19th and early 20th centuries

and consisted of a system for counting procedures, complications, and deaths. This approach, was similar to the subsequent "quality assurance" (QA) approach. These earlier approaches were directed primarily towards defining standards of performance in medical care and identifying unacceptable levels of doctor performance in achieving these standards. These earlier approaches, involved determining fault after something went wrong. From this context, QA was viewed as reactive, policing, apportioning blame, and even punitive. The medical community developed negative perceptions of these activities and often resisted their implementation (19,20).

As a result, it became necessary to shift the focus away from the individual providers and their errors to a system-wide perspective. A pioneer, who worked to accomplish this, was Avedis Donabedian. He promoted a shift from focusing on the assessment of QA which centered on "doctor or human" performance to a more sophisticated model which embraced two major fundamental concepts – "systems" measures and "human" measures – in order to achieve optimal patient outcomes (21). This novel approach, destigmatised the individual as the target of "blame" for unfavourable outcomes, and emphasised a broader understanding, in which QA requires an awareness that the system also contributed to error. The model defined by Donabedian involved three concepts – structure, process and outcomes (22).

Birkmeyer JD *et al* (23) adopted the Donabedian paradigm to evaluate the quality of medical care, specifically in the context of surgery. "Structure" consisted of the components of the environment in which the health care was delivered. In surgery, such components included equipment and supplies, the members of the surgical team and their qualifications, the nature and ownership of the practice groups, review committees and

oversight mechanisms. "Process" comprised what the provider did to and for the patient in the course of the encounter. The pre-operative preparation of the patient, the choice of the surgical intervention and its execution, and the day-to-day post-operative care, were examples of process in surgery. "Outcome" referred to the patient's subsequent health status. Examples of frequently employed outcomes in surgery included post-operative morbidity, mortality, and long-term survival (23). The key points of the study by Birkmeyer *et al*, as well as, some other examples in the literature are described below.

b. Structural measures assessing quality

Procedure volume, measured at either the surgeon or hospital level, is a commonly used structural measure, and is often used as a surrogate for surgical quality. There has been much debate about the magnitude of volume-outcomes associations with various procedures (24). However, there is little doubt that high-volume providers have lower operative mortality, fewer complications and better long-term survival with some operations than their lower-volume counterparts (23).

Among other structural variables, subspecialty training by the operating surgeon is often cited as a predictor of improved surgical outcomes. For example, patients undergoing resection for rectal cancer had lower recurrence rates and improved survival rates, when treated by surgeons who were board certified in colorectal surgery (24). Structural variables, more broadly related to staff organisation and resource availability, may also influence surgical outcomes. For example, there is a suggestion that critically ill surgical patients had a lower mortality rate when treated in "closed" intensive care units i.e. those in which patients were managed primarily by dedicated, board-certified intensivists (25).

1. *Advantages of structural measures*

Compared with direct outcomes assessment, structural variables, including procedure volume, can be assessed easily and inexpensively, often with administrative data.

2. *Disadvantages of structural measures*

One of the problems is that the literature assessing structural measures is incomplete. It focuses on a small number of variables (e.g. volume) and outcomes measures (e.g. operative mortality). Little is known about the importance of relationships between structure and nonfatal outcomes. Unlike process measures, which can often be evaluated in randomised clinical trials, most structural measures can only be assessed in observational studies. It is often difficult, to rule out confounding variables as an explanation for observed associations between structure and outcomes. In contrast to process measures, many structural measures are not readily actionable, which limit their ultimate effectiveness as a means toward quality improvement. For example, a small hospital can increase the number of its high-risk patients receiving perioperative beta-blockers, but it cannot readily make itself a high-volume centre for a given procedure, or unless it has sufficient staff, can it convert to a closed-model intensive care unit (23).

Finally, and most importantly, structural variables are imperfect proxies for quality - they reflect average results for large groups of providers, not individuals. As an example, many low-volume hospitals have excellent performance, but many high-volume centres are poor performers. Even if all high-risk procedures were concentrated in high-volume hospitals, there would remain substantial variation in quality across hospitals and thus opportunity for improvement (23).

c. Process measures assessing quality

Hannan and colleagues performed a prospective clinical study of patients undergoing carotid endarterectomy at six hospitals in New York State (25). In this study, vascular surgeons had substantially lower 30-day rates of operative stroke or death than general surgeons or neurovascular surgeons. The investigators also found that use of intra-arterial shunting, eversion endarterectomy techniques, patching of the arteriotomy and protamine were associated with lower complication rates. Greater adoption of these four processes of care by vascular surgeons explained, to a large extent, their better outcomes (25).

1. Advantages of process measures

,

As potential quality indicators, process of care measures have several attractive features. In addition to the high level of evidence supporting their effectiveness (often randomised clinical trials), some process measures have very large potential benefits. Process of care measures are generally actionable, and link directly to quality improvement activities (23).

2. Disadvantages of process measures

Measurement systems focusing on process variables must be able to accurately identify eligible patient populations (i.e. the right denominator). Many processes known to be effective in general may not be appropriate for all patients undergoing a given procedure (e.g. beta blockers in patients with brad arrhythmias or severe left ventricular dysfunction). Ensuring the right denominator implies the need for clinical data, which may be labour intensive and therefore, a practical limitation of process measurement (23).

A second major limitation of process measures is the relative lack of evidence about which processes are important for specific procedures. The bulk of existing literature on process measures focuses on the medical management of surgical patients. Many of the most serious adverse events occurring after surgery are non-medical in nature, arising from technical problems associated with the procedure itself - anastomotic leaks, bleeding or wound complications. Although high leverage technical processes have been identified for some procedures (notably CABG and carotid endarterectomy), very few procedures have been studied in great detail leaving major knowledge gaps (23).

d. Direct outcomes measures assessing quality

Although operative mortality is most commonly used, other outcomes measures, which could be considered as quality indicators, include complication rates, length of stay, re-admission rates, patient satisfaction, functional health status and other measures of health-related quality of life.

There are many ongoing, large-scale initiatives aimed specifically at measuring and improving surgical outcomes. Clinical outcomes registries in cardiac surgery, including those launched in New York, Pennsylvania and northern New England, in the 1980s, were among the earliest and most successful (26,27). Although these registries vary in many respects, they all provide hospitals and cardiac surgeons with feedback on their risk-adjusted morbidity and mortality rates. Over the past decade, prospective outcomes registries have been implemented in numerous other fields. The largest example of these must be the National Surgical Quality Improvement Programme (NSQIP) of the American College of Surgeons which assesses hospital-specific morbidity and mortality rates collected across a wide range of surgical specialties and procedures

(28).

This programme is described in detail in the chapter to follow.

1. Advantages of direct outcome measures

Direct outcomes measures have at least two major advantages. First, because most consider patient outcomes as the "bottom line" of surgical practice, efforts assessing quality with direct outcomes measures have obvious face validity and are likely to get the greatest buy-in from surgeons. Secondly, measurement alone may improve outcomes, - the so-called 'Hawthorne effect'(23).

2. Disadvantages of direct outcome measures

The most important limitation of direct outcomes measurement relates to sample size. Outside of cardiac surgery, very few procedures have baseline mortality rates of 5% or higher and are performed frequently at individual hospitals (particularly low-volume ones). Most common operations tend to be associated with low baseline risks, and this substantially compounds problems with statistical power in measuring outcomes at the provider level. A summary of the above comments from Birkmeyer *et al* is presented in the following table.

Table 2.2 Measuring quality of surgical care using the Donabedian paradigm

	Structure	**Process**	**Outcome**
Examples	Procedure-volume	Perioperative B-blockers in high-risk surgical patients	Morbidity and mortality rates
	Fellowship-trained surgeons		Functional health status
	'Closed' intensive care	Use of internal mammary graft during coronary artery bypass	Patient satisfaction

	units	graft	Cost
Primary advantage(s)	Expedient, inexpensive proxies of surgical outcomes	Reflect care that patients actually receive- may seem 'fairer' to providers Actionable from provider perspective, clear link to quality improvement activities	Buy-in from surgeons- the 'bottom line' of what they do Outcomes measurement alone may improve outcomes
Disadvantages	Most variables not actionable from provider perspective Imperfect proxies for outcomes-reflect average results for large groups of providers, not individuals	Little information about which processes are important for specific procedures	Numbers too small to measure with adequate precision procedure-specific outcomes for most hospitals and procedures Outcomes measures that are not procedure-specific less useful for purposes of quality improvement

Source: Adapted from Birkmeyer et al [23].

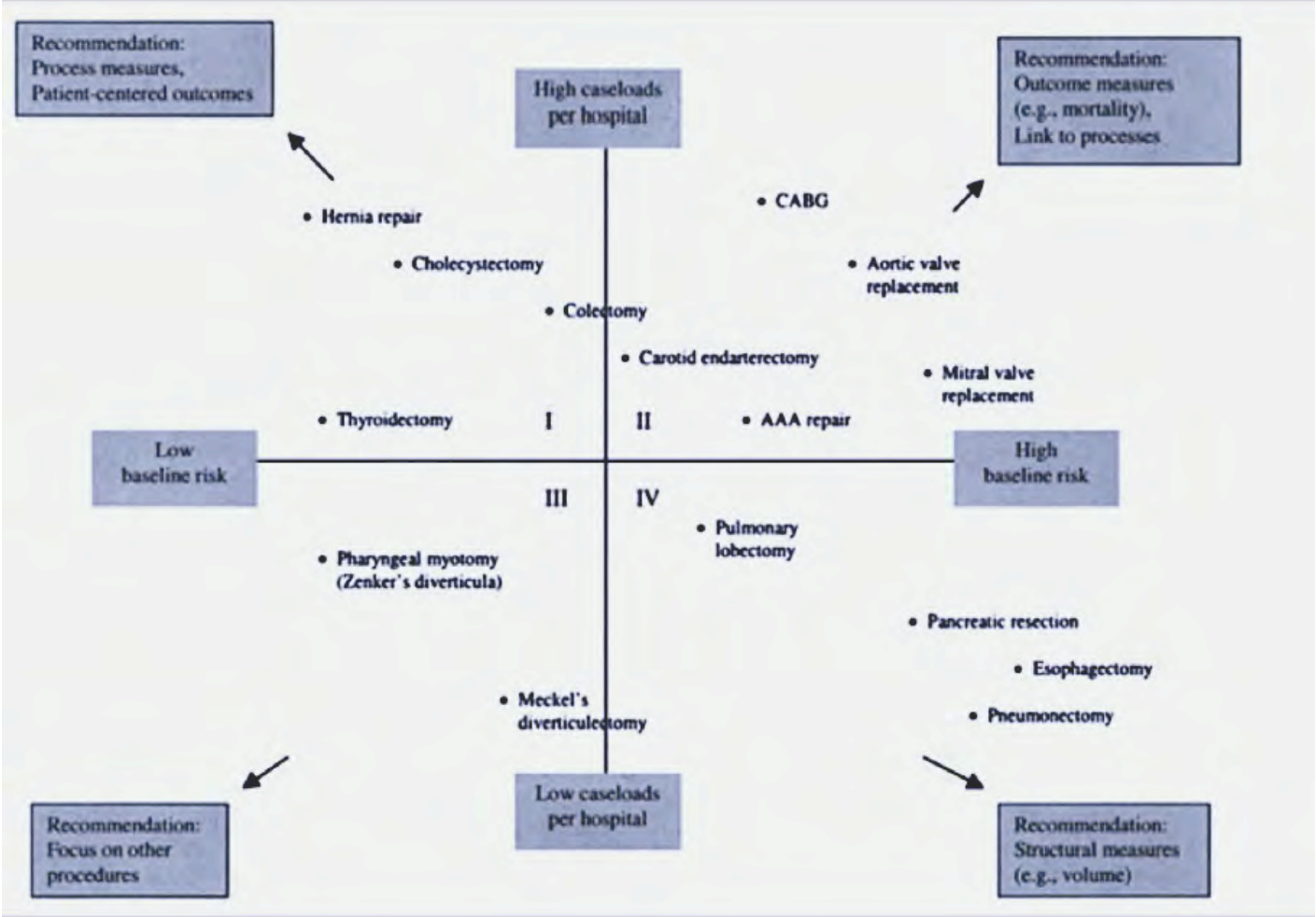
e. Choosing the right measure

Although structural, process and outcomes measures all have unique strengths, these three measures have distinct downsides, depending on how they are used. For these reasons, both surgeons and policy makers should be flexible in their approach to measuring quality and develop strategies best suited to meeting specific needs (23).

The procedure itself may be the most important factor in deciding about the most effective approach to quality measurement. Two attributes are particularly important: 1) The baseline risks of the procedure and 2) how commonly it is performed at individual hospitals (Figure 2.2). Measuring quality for procedures, which are both low risk and uncommonly performed (Quadrant III) should receive low priority. Many high-risk procedures, such as, esophagectomy and pancreatic resection (Quadrant IV) are

performed too infrequently at the vast majority of hospitals to support direct outcomes assessment. Procedure volume, a structural measure highly correlated with mortality for many of these procedures, is likely to be the only practical quality indicator. Quality for procedures, which are both common and relatively high risk (e.g. CABG, Quadrant II) is best assessed directly using risk-adjusted measures of morbidity and mortality.

Figure 2.2 Choosing the right quality measure based on procedure volume and baseline risk



Source: Adapted from Birkmeyer et al [23].

2.5 Quality improvement programmes

Quality improvement (QI) programmes are processes of patient care, which emphasise a multidisciplinary approach to problem solving and, which focus not on individuals, but on systems of patient care, which may be the cause of variations. QI programmes consist of periodic scheduled evaluation of organisational activities, policies, procedures and performance to identify best practices and target areas in need of improvement. QI programmes also include implementation of corrective actions or policy changes where needed (19). In general, formal QI programmes must have the essential attributes listed in Table 2.3.

Table 2.3 Principles fundamental to the success of a QI programme

The programme must be scheduled, planned and organized.
There must be a dedicated clinician leader who takes the lead in ensuring quality and is invested with power and the authority by the hospital administration (i.e. authority and accountability are essential components of QI).
Peer review processes must be uniform, nonpolitical and honest, and should incorporate evidence-based medicine.
Evaluations must be critical, but not destructive. A fair and nonpartisan approach that respects the opinions and role of deliverers of healthcare is essential.
The programme must be driven by predefined objective criteria and outcome definitions.
Infrastructure, logistical support, and investment are needed to ensure the improvement of quality.
Hard data must be incorporated
Data collection must be ongoing
The programme should incorporate methods not only for *identifying* problems, but also for *fixing* problems-often termed 'corrective strategies.'
The programme should measure what is achieved by the corrective strategies to confirm that they have had their intended effect-often termed 'closing the loop.'
The programme should be implemented with a commitment for sustained activity and improvement using ongoing data monitoring, data analysis and corrective strategies.

Source: Adapted from WHO, Guidelines for trauma quality improvement programs [19]

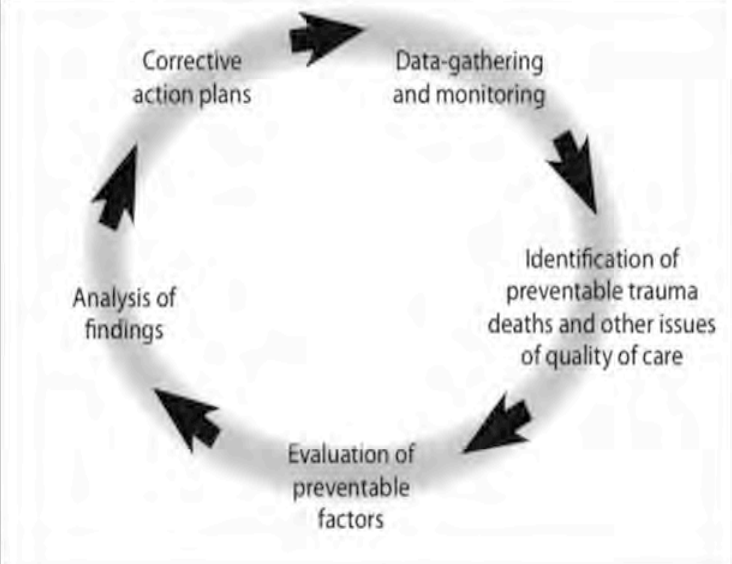
Essential to a QI programme is that it needs to be supported by appropriate infrastructure and reliable data collection. The goal in data collection and analysis is to obtain consistently valid and objective information identifying 'opportunities for improvement.' Results of analysis must define corrective strategies. A continuous cycle of monitoring, assessment and management should be performed routinely.

The formulation of corrective strategies or action plans in response to identified 'opportunities for improvement' is essential to QI. Corrective strategies or action plans are structured efforts to improve sub-optimal performance that is identified through the

QI process. Corrective strategies are basically solutions proposed for fixing a problem or process, which may be either case-specific or system-specific. Examples of corrective strategies include guidelines, pathways, protocols, education, peer reviews, monitoring and educational activities for specific clinical skills and resource upgrades and enhancements.

Confirmation and documentation of the impact of corrective actions is commonly termed, 'closing the loop.' (Figure 2.3). In addition to identifying problems and implementing solutions the QI process is dedicated to ensuring that there are measureable improvements in outcome that can be documented in response to the corrective strategies, which are implemented (19).

Figure 2.3 Closing the loop



Source: Adapted from WHO, Guidelines for trauma quality improvement programmes [19]

2.6 Evidence supporting the benefits of quality improvement programmes

It is important to assess the evidence base supporting the use of such surgical quality improvement programmes.

A systematic review published by Michael Fung-Kee-Fung *et al* (149) in the Annals of Surgery in 2009, and entitled 'Regional Collaborations as a Tool for Quality improvement in Surgery,' identified seven regional collaborative initiatives which were established to enhance quality improvement, quality of care, patient safety, knowledge transfer and common practice. Motivations for initiating collaborations were often in response to external demands for performance data. Changes in the processes of clinical care and improvements in clinical outcomes were reported on the basis of the collaborative efforts. Changes in clinical practice in line with regional guidelines on the use of chemotherapy and axillary surgery were observed in one collaborative initiative. Decreases in mortality rates, lower duration of post-operative intubations, and fewer surgical site infections were reported in three collaborative initiatives. Quality improvement process measures improved across all of the studies (149).

A collaborative review by WHO and the IATSIC, published in the World Journal of Surgery in 2009, entitled 'Establishing the evidence base for trauma quality improvement' (21), included both articles, reporting the evaluation results of a trauma QI programme and focusing on the identification and correction of specific problems. Thirteen articles reported on mortality as the main outcome, twelve reported on changes in morbidity (infection rates, complications), patient satisfaction, costs and other outcomes of tangible patient benefits, and eleven reported on changes in process of care.

Thirty articles addressed hospital-based care, four addressed system-wide care and two addressed pre-hospital care. Thirty-four articles reported an improvement in the outcome assessed, two reported no change, and none reported worsening of the outcome. Five articles also reported cost savings. The authors concluded that trauma QI programmes consistently showed improvement in the process of care, decreased mortality, and decreased costs (21).

The most cited article supporting the evidence base for quality improvement programmes for surgery, is the study published in the Annals of Surgery in 2009, which demonstrated improvement in the quality of surgery in the ACS-NSQIP enrolled hospitals over a 3-year period (2005-2007) (150). This analysis of 118 hospitals, showed a 66% reduction in risk-adjusted mortality and 82% improved risk-adjusted complication rates. In essence, an average of 52 complications were prevented per ACS-NSQIP enrolled hospital in 2007. Importantly, the initially worst performing hospitals had a greater likelihood of improvement (150).

Finally, an article by Cohen *et al* (42) published recently in the Annals of Surgery, reported on the long-term improvement in surgical outcomes associated with participation in ACS-NSQIP. ACS-NSQIP data (2006-2013) were used to create prediction models for mortality, morbidity (any of several distinct adverse outcomes) and surgical site infection (SSI). The primary performance metric was the within hospital trend in logged Observed versus Expected ratios over time for mortality, morbidity, and SSI. The authors found that for hospitals currently in the programme for at least 3 years, 69%, 79%, and 71% showed improvement in mortality, morbidity, and SSI, respectively. For these hospitals, they estimated 0.8%, 3.1%, and 2.6% annual reductions (compared

with the previous year's rates) for mortality, morbidity and SSI respectively. They concluded that participation in ACS-NSQIP was associated with reductions in adverse events after surgery (42).

After reviewing the literature, by far the most published QI programme is the ACS-NSQIP, and subsequently some background information on this programme, is warranted.

2.7 The American College of Surgeons National Quality Improvement Programme

Intuitively, measures of outcome are particularly applicable for the assessment of the quality of surgical care, because surgery involves an intervention with an expected outcome. For example, a repair of an abdominal aortic aneurysm is expected to prevent a subsequent fatal rupture, and a replacement of an osteoarthritic hip is expected to enable the patient to walk without pain (151,152). The main limitation of the use of outcome in the comparative assessment of the quality of surgical care, however, is the need to use adequate and validated models for risk-adjustment (151). The pre-surgical severity of illness must be adjusted, if outcome is to be used in the comparative assessment of the quality of surgical care.

Differences in pre-operative risk factors must be taken into account, before differences in surgical outcomes are attributed to the skill of the surgical team and the perioperative environment. Measuring comparative surgical outcomes using methods to adjust for patient risk has become an important part of measuring the quality of surgical care (153). Several models for risk adjustment and comparative assessment of outcome have been prospectively developed for cardiac surgery (35,63,85,135,154). Until 1994,

however, uniform and validated models for non-cardiac surgery did not exist. Prompted by a 1986 congressional mandate, the Veterans Health Administration (VHA) in the United States of America conducted the National VA Surgical Risk Study (NVASRS) between 1 October 1991 and 31 December 1993 (155). The NVASRS was conducted in 44 VA Medical Centres (VAMC's). The aim of the study was to develop and validate risk-adjustment models for the prediction of surgical outcomes, and the comparative assessment of the quality of surgical care in multiple facilities. A dedicated clinical nurse reviewer in each VAMC prospectively collected pre-surgical, surgical and 30-day outcome information after major surgery. Based on data from 87,078 major surgical procedures, risk-adjustment models for 30-day mortality and morbidity rates were developed for all non-cardiac surgery and for each of the following subspecialties: general surgery, vascular surgery, orthopedic surgery, urology, thoracic (non-cardiac) surgery, neurosurgery, plastic surgery and otolaryngology (155).

In 2001, following the success of the NVASRS, the American College of Surgeons decided to collaborate with the Veterans Association's NSQIP to investigate the applicability of the NSQIP methodology to the private sector. This study was based on the methodology and results of the VA-NSQIP and the results of the initial feasibility study, which had been conducted at three non-federal hospitals, namely, Emory University in Atlanta, the University of Michigan in Ann Arbor, and the University of Kentucky in Lexington (28).

The grant was approved and resulted in the inception of the Patient Safety in Surgery Study (PSS), a prospective cohort study (156). The primary endpoint of the study was that NSQIP, a risk-adjusted adverse reporting system for major surgical

operations in the Veterans Association, could be translated into non-federal institutions, resulting in measurable reductions in surgical morbidity and mortality. The secondary endpoints were that risk-adjusted outcomes of surgical care were comparable in the Veterans Association and the private sector. Thus, implementation of the NSQIP in the Veterans Association and the private sector would allow for identification of structures and processes, which would improve patient outcomes and enhance patient safety (156). The NSQIP methodology was successfully implemented in the 14 university medical centres during the three years, and overall morbidity and mortality declined by 43% and 47% respectively (156).

Based on the results of this study, the ACS-NSQIP was established in 2005, and at time of writing included over 250 hospitals in the developed world. Considering such staggering success, it is necessary to look at its design in more detail (157).

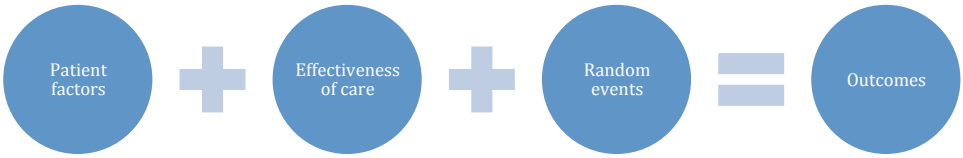
a. The Vision for the NSQIP

The NSQIP was always meant to have dual responsibilities to both surgical service management and surgical research. It was believed that good science informs good management and performance improvement and that the programme should be "owned" and managed by the participating surgeons and surgical services. Data quality, reliability and validity were of paramount importance to ensure that the programme would deliver information, which the surgeons could trust. The main purpose of the reporting system was for the surgical team to be able to use the information in their own practices for quality improvement purposes. The NSQIP had never been intended to be used as a source of pay-for-performance measure or for public reporting (157).

b. Conceptual model

The ACS-NSQIP was designed on the conceptual model of Iezzoni (see Figure 2.5), in which the outcome of care is determined by patient factors, the effectiveness of care and random events (158).

Figure 2.5 Iezzoni's model for effectiveness of care



If high quality data were collected on patient pre-operative factors, and good risk-adjustment statistical models were used then differences in risk adjusted patient outcomes among surgical services were likely to be due to variations in the effectiveness of care.

c. Process or outcome as a quality measure in surgery?

As discussed in previous chapters, a quality measure in surgery may be structure, process or outcome of care. The ACS-NSQIP founders believed that post-operative outcomes, i.e. a change in the patient's health status, were particularly relevant when measuring quality in surgery. The view held was that surgery was particularly amenable to the use of patient outcomes prior to surgery as a surgical operation was a predictable event with an expected outcome in most cases. Table 2.5 presents possible outcome measures, which could be used in surgical quality improvement, and the advantages and disadvantages of each. The 30-day post-operative mortality and morbidity rates were chosen as the primary outcomes for the ACS-NSQIP, because of the relative ease of data

collection, and the importance of these outcomes. The intention was to also incorporate other outcomes listed in Table 2.5 (157).

Table 2.5 Potential outcome measures to assess surgical quality

Measure	Advantage	Disadvantage
1) Post-operative mortality	Easily ascertained	Low event rate
		Low statistical power
		Primarily determined by patient factors
2) Post-operative morbidity	Higher event rate	More difficult to ascertain consistently
	Higher statistical power	
3) Long-term mortality	More meaningful for some operations (eg, cancer)	More difficult to ascertain
		Determined by many factors - patient factors, type of therapeutic intervention, patient response
	High statistical power	
4) Functional status or quality of life	More meaningful for some operations (eg orthopedic, urologic)	More difficult to measure
		More resources needed
		Requires pre-operative and post-operative assessment
5) Patient satisfaction	Important to the patient. High statistical power	More difficult to measure
		More resources needed
		May be considered subjective and subject to halo effects
6) Post-operative length of stay	Easy to ascertain	Skewed distribution
	High statistical power	More difficult to analyse
		May be related to factors other than quality (eg, socioeconomic status)
7) Cost	Important	More difficult to measure
	High statistical power	Skewed distribution
		More difficult to analyse
		Non-uniformity of cost accounting between centers

d. Selection of the patient population

One of the most contentious issues in the design of the ACS-NSQIP was whether to evaluate selected operations or to take a random sample of all operations. The methodologists believed that a random sample was better, because a sample would provide knowledge about different types of operations and their outcomes. This would cover operations performed by the different surgical sub-specialties and the surgical service as a whole. Differences in the operations could be accounted for by measures of the complexity of the operations, and by coding the subspecialties from which they came. An overall risk model for mortality or morbidity combining all operations together could provide a meaningful measure of quality for a surgical sub-specialty or a surgical service as a whole. As the size of the ACS-NSQIP clinical registry grew, individual higher-volume operations could be assessed as well.

Consequently, the ACS-NSQIP patient population was defined as "a systematic sample of major operations performed under general, spinal or epidural anaesthesia". Cases were selected through a list of appropriate Current Procedural Terminology [CPT] codes. It was thought to be important to select cases from each day of the week, and consecutive eligible cases were selected starting with a different day of the week in each 8-day cycle with 40 cases per cycle. Minor operations were excluded, and some high-volume operations e.g. inguinal hernia repair, breast lumpectomies, were limited to the first 3 or 5 cases in each 8-day cycle. This sampling procedure resulted in about 1,600 cases assessed per hospital per year (157).

e. Data collection

The data collected for each ACS-NSQIP assessed case included about 105

variables - 70 pre-operative risk factors, 11 variables about the operation and 24 post-operative outcome variables. Because the ACS-NSQIP was intended to cover many different types of operations from many different surgical subspecialties, more generic types of variables were collected rather than more disease- or operation-specific variables. The generic variables were selected based on ease of data collection, reliability, and their importance in evaluating the risk and outcomes of patients undergoing many different types of operations (157).

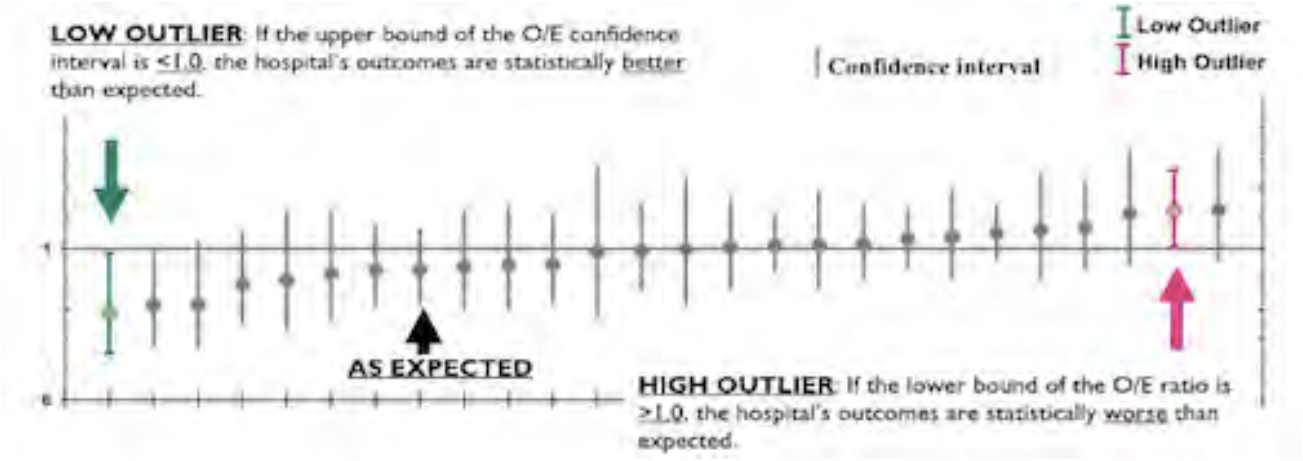
A surgical nurse reviewer was assigned to collect and transmit the data at each centre. Trained in clinical medicine and quality assurance, these clinical nurses completed in-depth training on the data collection procedures and detailed definitions of each of the variables. Regular conference calls and periodic inter-rater reliability site visits were conducted to maintain data uniformity and reliability. A manual of operations specified all aspects of data collection, including selection of patients, definitions of variables, methods for entering data into the computer system and a list of who to contact for answers to questions (157).

f. Statistical analysis

Multiple logistic regression analysis was used in which the adverse event was the dependent variable and the patient pre-operative risk factors were the independent variables. Relative value unit (RVU) was included in each model to account for differences in operation complexity. Once the logistic regression equation was computed, the equation could be used to calculate a probability of the adverse event for each patient.

These probabilities were aggregated across a surgical specialty or a surgical service to obtain the expected number of adverse events (E) for the patient sample for that surgical unit based on the patient characteristics. An O/E ratio and confidence interval were then calculated for the patient sample for that surgical unit where O is the number of patients observed to have the adverse event. If the surgical unit's O/E ratio was above one and the lower bound of the confidence interval was greater than one then the surgical unit had experienced a statistically significant larger number of adverse events than would be expected based on its patient characteristics. If the surgical unit's O/E ratio was below one and the upper bound of the confidence interval was less than one then the surgical unit had experienced a statistically significant smaller number of adverse events than would be expected on the basis of its patient characteristics. An example of such a caterpillar plot is shown in figure 2.5.

Figure 2.5 Caterpillar plot benchmarking hospitals by O/E ratio



Source: Adapted from Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? (157)

g. Feedback to participating hospitals

In the ACS-NSQIP, 6-monthly statistical reports have been distributed to participating hospitals since June 2006. Each semi-annual report used data from the previous 12- month reporting period.

The success of ACS-NSQIP could be attributed to a number of core strengths of the programme. Data extraction was conducted by trained nurses according to well tested procedures and rigorously defined variables. A comprehensive set of clinical and laboratory risk factors were collected on every patient. These form the basis of well validated, risk- adjustment models. Submitted data were externally audited to ensure their completeness and accuracy. For all these reasons participating hospitals could expect extremely robust risk-adjusted estimates of their surgical morbidity and mortality expressed relative to other hospitals as observed to expected (O/E) ratios (158). Following the success of the ACS-NSQIP in the United States of America, the ACS Committee of Trauma considered a QI programme specifically for trauma care.

2.8 The American College of Surgeons Trauma Quality Improvement Programme

It is well known that trauma systems improve outcomes among seriously injured adults and children (59). Previous research has highlighted that there was variability in the quality of care between trauma centres (159). It is most likely that the differences in patient selection (selection bias), case mix, data quality, geography and other factors inherent to different injured populations, most likely contributed to this variability.

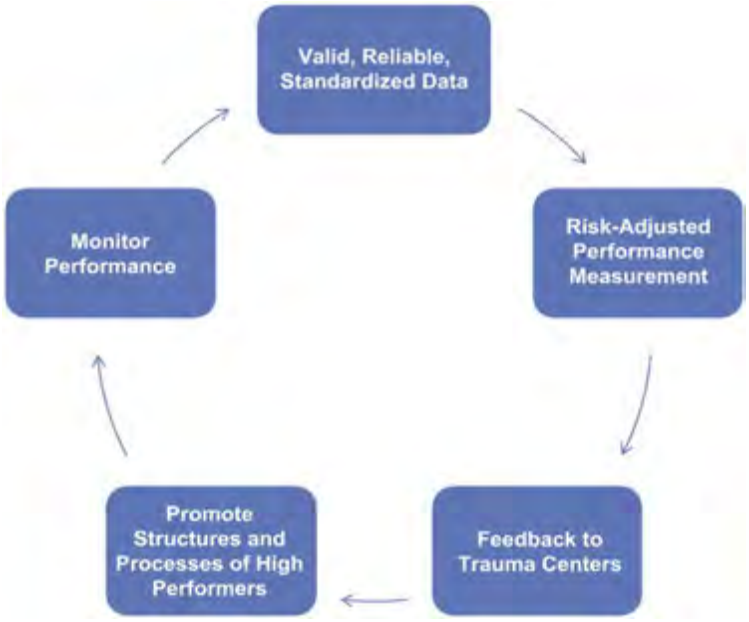
However, variability in the processes and quality of care at different trauma centres could also contribute to outcomes variations among hospitals (160). Assessing outcomes objectively was challenging in trauma care, but the introduction of the Trauma Injury Severity Score (TRISS) allowed centers to identify patients with unexpected outcomes. Between 1982 and 1989, the Major Trauma Outcome Study (MTOS) established national standards for trauma care, which could be used to benchmark trauma centers (161).

The MTOS database was used to develop the Trauma Injury Severity Score (TRISS) methodology for predicting the probability of survival for an individual trauma patient (162). The prediction equation was based on indices of physiological derangement, anatomical injury severity and age. Trauma Injury Severity Score (TRISS) allowed trauma centers to compare their observed outcomes against the predicted outcomes based on the patient's presenting clinical status. However, advances in trauma and critical care have made the co-efficients, which were calculated and validated with the MTOS data obsolete (163).

It was within this context that the ACS-TQIP was conceived in 2008, through a small working group assembled by the ACS Committee on Trauma (ACS-COT). The impetus for change came from accumulating evidence of the effectiveness of large national collaboratives, which, were based on: (1) standardised data, (2) feedback to centres, and (3) a network, which would allow for the sharing of challenges and best practices (160). High-quality data provided the opportunity to provide centres and risk-adjusted performance measures. These institutions were then able to review their performance in the context of their environment, and seek areas to improve relying on strategies used by high-performing centres. Performance was monitored and the loop

continues Figure 2.6 (19).

Fig 2.6 The conceptual framework of continuous quality improvement underlying ACS TQIP and other quality collaboratives.



Source: Adapted WHO. Guidelines for Trauma Quality Improvement Programs [19]

The ACS-TQIP expanded on a foundation of quality improvement programmes already implemented by the ACS, including the NSQIP performance improvement, patient safety and trauma centre verification. The primary goal of ACS-TQIP was to improve the quality of trauma care through outcomes-based, risk-adjusted benchmarking of trauma centres and feedback reports.

However, simply incorporating trauma patients into ACS-NSQIP was not possible. ACS-NSQIP had created a data infrastructure where none previously existed, and it required a well-trained surgical clinical reviewer to collect and submit data. By contrast, each trauma centre had a trauma registry, a team of registrars to collect the data

and a means of aggregating these data through the National Trauma Data Bank. To avoid creating a parallel data infrastructure and costly duplication, it was decided to use the existing infrastructure and work toward data standardisation. This decision moved the process forward rapidly to allow the development of a 2-year pilot study involving 23 level 1 and level 2 trauma centres (164). This pilot study, similar to the Patient Safety in Surgery study, was a success, and concluded that using the National Trauma Data Bank infrastructure to provide risk-adjusted benchmarking of trauma centre mortality was feasible and perceived as useful (164).

In a subsequent study, the authors examined the relationship between outcomes and expenditure for trauma patients treated in hospitals participating in a Collaborative Quality Initiative (CQI) and those who did not (166). The risk-adjusted rate of serious complications declined from 14.9% to 9.1% ($p < 0.001$) in participating hospitals (Post-CQI, n = 26). The average episode payments decreased by $2 720 (from $36 043 to $33 323 $p = 0.08$) among patients treated in Post-CQI centres, whereas patients treated at Never-CQI institutions had a significant year-to-year increase in payments (from $23.47 to $28.446 $p < 0.001$). A savings of $6.5 million in total episode payments from 2010 to 2011 was achieved for payer-covered Post-CQI treated patients. This study confirmed the hypothesis that participation in a regional CQI programmes improves outcomes, and reduces costs for trauma patients. Support for a regional CQI for trauma represented an effective investment to achieve health care value (165).

Since the inception of ACS-TQIP, 143 hospitals across the United States of America and Canada have successfully adopted participant status. A closer look into the methodology of the ACS TQIP is therefore warranted (160).

a. Study design and setting

ACS-TQIP used a retrospective cohort of trauma patients meeting specific inclusion criteria and cared for in designated and ACS-verified Level I and II hospitals across the United States of America and Canada. Trauma centre participation in ACS-TQIP was voluntary, entailed the use of existing trauma registry data, conforming to specific standards and required an annual fee to offset the costs of the programme.

Patient population and inclusion criteria

ACS-TQIP used a broad, heterogeneous group of seriously injured patients with focused assessment of several distinct subset populations (Table 2.6). Analysis of these subset cohorts addressed different aspects of trauma care. These groups included blunt multisystem injury (Abbreviated Injury Severity [AIS] ≥3 in at least 2 body regions), penetrating truncal injury (AIS greater than 3 in the neck, chest or abdomen), shock (systolic blood pressure [SBP] <90 mmHg), isolated traumatic brain injury and elderly. These cohorts were selected to focus performance and treatment efforts, target distinct types of trauma patients with different needs and management strategies, highlight injury populations with varying representation and experience among centres, and to increase comparability among hospitals.

Table 2.6 Inclusion and Exclusion Criteria for Trauma Quality Improvement Programme

Inclusion criteria	**Age 16 y or older**
	At least 1 valid trauma ICD-9 code in the range of 800 to 959.9 (excluding late effects (905-909.9), superficial injuries (910-924.9), and foreign bodies (930-930.9)
	Primary mechanism of injury classified as either blunt or penetrating:
	Blunt is defined as an injury where the primary E-code is mapped to the following categories: fall, machinery, motor vehicle traffic, pedestrian, cyclist, and struck by or against. Penetrating is defined as an injury where the primary E-code is mapped to the following categories: cut/pierce and firearm
	Severely injured patients with at least one AIS ≥3 injury: For blunt injuries: at least 1 injury in any of the following AIS body regions: head, face, neck, thorax, abdomen, spine, or upper and lower extremity
	For penetrating injuries: at least one AIS ≥3 injury in any of the following AIS body regions: neck, thorax, and abdomen
	Injury severity score ≥ 9
	ED discharge disposition and hospital discharge disposition cannot both be unknown
Exclusion criteria	Comorbidity: pre-existing advanced directive to withhold life-sustaining interventions
	Isolated hip fractures for patients 65 years or older with an injury with mechanism of fall is defined as any traumatic injury with at least one of the following diagnosis codes: 851810.3 Femur, fracture, intertrochanteric851812.3 Femur, fracture, neck 851818.3 Femur, fracture, subtrochanteric and all other injuries in AIS body region ”external” (i.e., bruise, abrasion, or laceration)

AIS, Abbreviated Injury Scale; ED, emergency department; SBP, systolic blood pressure.

Outcomes measures

Primary outcomes included mortality (on arrival, in the ED and in-hospital), complications and resource usage. Although in-hospital mortality is influenced by many factors, it is a well-recognized outcome in trauma care, which is reliably captured in trauma registries and useful for ACS-TQIP. For complications, ACS-TQIP focused on addressing potentially preventable events, which caused disability, additional resource utilisation and deviations from the expected clinical course after injury. Measures of resource utilisation (e.g. length of stay [LOS], duration ICU stay and ventilator days) were selected based on feasibility of data capture, association with quality of care, relation to other ACS-TQIP outcomes (e.g. complications), responsiveness to evidence-based practice guidelines and a direct relationship with cost (160).

Variables

Multiple data elements were captured for ACS-TQIP and considered in risk-adjustment models. The variables included patient demographics, co-morbid conditions, initial ED physiology, ED disposition, transfer status, mechanism of injury, ICD-9-CM diagnosis codes, procedures, AIS scores, derived injury severity measures, LOS, ICU stay, complications and in-hospital mortality (Table 2.7.).

Table 2.7. Variables Considered in Trauma Quality Improvement Programme Multivariable Models

Mortality model
Initial GCS motor score in ED, Initial systolic BP in ED, Initial pulse rate in ED
Mechanism of injury: Pedestrian/pedal: motor vehicle-pedal cyclist, motor vehicle-pedestrian, pedal cyclist/other, pedestrian/other, motor vehicle occupant and other motor vehicle related event, motorcyclist, fall, struck by or against, firearm, cut/pierce, other
Transfer status, age, gender, race and ethnicity
AIS severity by individual body region (except for external)
Individual comorbidities: Heart disease, cancer, liver disease, alcoholism, smoking, stroke, diabetes, hypertension, renal disease, impaired sensorium, respiratory disease, functional dependence, bleeding disorder, peripheral vascular disease, steroid use (included if prevalence >2%)
Region, payment type
Derived variables: Injury Severity Score, ICD9-based Injury Severity Score, SWI (based on ICD9 injury codes), Maximum AIS by body region, Lowest AIS =lowest AIS score, Serious AIS = maximum AIS ≥3 for specific body regions, Arrest SBP =emergency department SBP ≤ 40 mmHg
Length of stay model
Same covariates noted above, plus complications
Cardiovascular (cardiac arrest with CPR, myocardial infarction, stroke)
Surgical infections (organ/space surgical site infection, deep surgical site infection, superficial surgical site infection, and wound disruption)
Acute respiratory distress syndrome, pulmonary embolism, renal failure, pneumonia, sepsis

AIS, Abbreviated Injury Scale; ED, emergency department; SBP, systolic blood pressure; SWI, Single Worst Injury.

Once AIS scores had been generated for all ACS-TQIP patients, an ISS was calculated from the AIS values. ICD- 9-CM diagnosis codes were also used to generate a separate measure of injury severity termed the ICD-9 Injury Severity Score. This score was calculated by first creating survival risk ratios for every ICD-9 injury diagnosis in a reference population. For patients in TQIP the Single Worst Injury Severity Score,

ICD-9, was used for modeling. The goal in generating injury severity measures between hospitals was consistency in injury coding and therefore comparability of results between institutions (160).

b. Statistical analysis

In-hospital mortality

A central goal of ACS-TQIP was to provide valid risk-adjusted mortality estimates for trauma centre comparison and benchmarking. Patient populations vary across hospitals by demographics, acuity, mechanism of injury, timing of presentation and comorbidities, and methods were needed to account for these differences. Multiple approaches to risk-adjusted modelling were considered for ACS-TQIP, including logistic regression, hierarchical models, generalised estimating equations, Bayesian analysis, linear regression, and Poisson regression.

Although each of these approaches has certain advantages, there were concerns that overly complicated approaches to modelling would reduce the face validity and interpretability of ACS-TQIP reports, and potentially create analytic obstacles. After evaluating the merits and limitations of different types of models, ACS-TQIP selected multivariate logistic regression for the primary mortality model. This selection was based on its face validity (widely recognisable and easily understood), its generation of risk-adjusted estimates which preserve centre level differences as quality targets, its compatibility with multiple imputation and its equivalent performance with less complexity for risk-adjusting trauma care (160). Several multivariate logistic regression models were developed and tested before deciding on the final TQIP risk-adjusted

mortality model.

Length of stay

Similar models were developed to produce risk-adjusted estimates for LOS as a measure of resource utilisation. Risk factors similar to those described for the mortality model were considered in the LOS model (Table 2.7). Because LOS could be affected by in-hospital mortality rates (e.g. a hospital with high in-hospital mortality may appear to have low LOS), they opted to restrict LOS models to survivors for simplicity and clarity (160).

c. Presentation of results

One of the key goals of ACS-TQIP was to provide readily interpretable and informative comparisons of trauma center performance. Initial ACS-TQIP reports used rank plots, such as the caterpillar plots used in ACS-NSQIP. However, the ACS-COT noted that these plots had important limitations. These included a rank-order list, which was not necessarily meaningful (the majority of hospitals end up being ranked as equal), the potential for misinterpretation, the lack of hospital sample size information (inability to compare outcomes among similar-volume hospitals) and the difficulty in differentiating hospitals, which were close to outlier status. For these reasons, ACS-TQIP also used funnel plots. Funnel plots allow direct assessment of trauma centre volume, improved visual assessment of outlier hospitals (high and low), elimination of non-meaningful hospital rankings and easier identification of hospitals close to outlier status (e.g. early recognition of quality issues that can prompt behaviour change even if not yet statistically significant) (160).

2.9 Conclusions of the NSQIP and TQIP Programs

The ACS-NSQIP and subsequently the ACS-TQIP, has substantially improved the quality of surgical care and had a considerable influence on the culture of quality improvement in the profession.

The success of the ACS NSQIP and ACS-COT TQIP was the result of providing hospitals with rigorous clinical data, networking opportunities and resources to improve their risk-adjusted outcomes. In this way, the ACS-NSQIP and ACS-COT TQIP programs challenged its hospitals and health care providers to improve the care they provide, continuously. In addition to reducing the complications and mortality experienced by patients after surgical procedures, hospitals, which participated in these programmes, have seen the financial rewards of their quality improvement efforts. Therefore, the question must be asked, are these or similar programmes feasible in a LMIC?

2.10 Quality Improvement programs in LMIC's

To date there has been limited experience of QI programmes in LMICs. A review of the recent QI literature found that 4.1% of 121 articles took place in a LMIC (based on primary author institution). A review of the articles would help to understand the limitations and identify potential solutions for implementing QI programmes for surgery in LMIC's

a. Improving institutional maternal mortality in rural Mali, Senegal

Some of the best examples of the role of QI programs in LMICs have come from the field of obstetric care. A specific type of QI for obstetric care is the maternal death

audit. This has proved to be instrumental in improving obstetric care globally (159). A model programme for improving maternal care through QI processes (maternal death audit) in a district hospital in Senegal was based on a daily review of cases by senior specialists in obstetrics and gynaecology (166).

For each maternal death the senior staff interviewed the staff involved with the case and the patient's family. Standardised information obtained and reviewed at weekly meetings. Two senior specialists reviewed the charts of all maternal deaths annually to classify causes of death and contributing factors. This led to detailed recommendations for corrective action. Each subsequent year, the manager of district health services evaluated how well the recommendations for action had been implemented. The main recommendations focused on improving the 24-hour availability of essential drugs and blood, and the availability of basic emergency obstetric care at both hospitals and clinics. Over a four- year period, the case-fatality rate for women delivering at the hospital decreased from 6.0% to 2.6%, primarily due to decreases in deaths as a result of haemorrhage and hypertensive disorders (166).

b. The impact of the International Quality Improvement Collaborative on outcomes after congenital heart surgery in India

The International Quality Improvement Collaborative (IQIC) for Congenital Heart Surgery in Developing Countries was initiated to decrease mortality and major complications after congenital heart surgery in the developing world (167). A unit in India participated in this collaborative, and then sought to assess the impact of IQIC on post-operative outcomes after congenital heart surgery at their institution. The key components of the IQIC programme included the creation of a robust worldwide database

on key outcome measures and nurse education on quality driven 'best practices' using telemedicine platforms.

The participating institutions in India evaluated 1702 consecutive patients under the age of 18 years undergoing congenital heart surgery, from January 2010 - December 2012. The overall in-hospital mortality was 3.1%. Over the subsequent three years, there was a significant decline in bacterial sepsis (from 15.1%, to 9.6%, $P< 0.001$), surgical site infection (11.1% to 2.4%, $P< 0.001$) and average duration of ICU stay from 114 hours to 72 hours ($P< 0.001$). The decline in mortality from (4.3% to 2.2%) did not reach statistical significance. They concluded, that inclusion of their institution in the IQIC programme was associated with improvement in key outcome measures, following congenital heart surgery over a three- year period (167).

c. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population

A landmark study by Atul Gawande's group at the Centre for Public Health and Surgery at Harvard University, described the implementation of a 19-item surgical safety checklist. This checklist was designed to improve team communication and consistency of care in order to reduce complications and deaths associated with surgery (168). Between October 2007 and September 2008, eight hospitals in eight cities, which represented a variety of economic circumstances and diverse populations of patients, participated in the World Health Organisation's Safe Surgery Saves Lives programme. These cities included Toronto, Canada; New Delhi, India; Amman, Jordan; Auckland, New Zealand; Manila, Philippines; Ifakara, Tanzania; London, England; and Seattle, United States of America. The study collaborators, prospectively collected data on

clinical processes and outcomes from 3,733 consecutively enrolled patients, 16 years of age or older, who were undergoing non-cardiac surgery. They subsequently, collected data on 3,955 consecutively enrolled patients after the introduction of the Surgical Safety Checklist. The primary endpoint was the rate of complications, including death, during hospitalisation and within the first 30 days after the operation. The mortality rate was 1.5% before the checklist was introduced and declined to 0.8% afterwards ($p = 0.003$). In-patient complications occurred in 11.0% of patients at baseline and in 7.0% after introduction of the checklist ($P<0.001$). The study concluded that the implementation of the checklist was associated with concomitant reductions in the rates of death and complications among patients over 16 years of age who were undergoing non-cardiac surgery in a diverse group of hospitals.

d. Reviewing effectiveness of trauma QI's in LMIC's

A mixed methods research study involving thematic analysis of a meeting, as well as a pre-meeting structured survey to explore experiences with trauma QI activities in LMICs showed that standardising injury data is the first step in improving injury care (138). Injury control activities cannot be effectively conducted without adequate assessments of their impact. Development of reliable and valid minimal data sets, which include key performance indicators, were essential for effective injury surveillance, targeting interventions and assessing their success or failure. Implementation of simple and standardised data collection forms and basic software programmes were not necessarily expensive (138).

e. Cost effectiveness of QI's in LMIC's

Improving the quality of medical care through measures such as QI programmes,

have been shown by the Disease Control Priorities Project (DCPP) to be very cost-effective. Cost-effectiveness ratios for such efforts range from US$ 4 to US$ 28 per disability adjusted life year (DALY) averted, in circumstances where disease prevalence was high and existing quality of care was low. These cost effectiveness ratios were in the range of the most cost effective interventions studied by the DCPP. The cost effectiveness ratios of interventions studied by the DCPP from US$1 per DALY averted (very cost-effective) to over US$ 20 000 (not cost-effective). Moreover, QI programmes themselves could also lead to cost savings (19).

After an extensive review of the literature, the evidence supporting QI programmes in High-Income Countries was strong, particularly after the history and development of the ACS-NSQIP. In LMICs there was a significant paucity of evidence supporting CI programmes. A common message in the literature was that the acquisition of standardised, reliable, data collected by a sustainable means in order to provide risk-adjusted outcome measures was key in the implementation of a structured-surgical quality programme anywhere in the world. In a LMIC context, including in South Africa, this would require novel means of data-capture to be explored.

2.11 Finding novel solutions

In the preceding chapters, both the ACS-NSQIP and TQIP programmes showed that to be successful a large national collaborative requires interested participants. The success of ACS-NSQIP and the long-standing interest in performance improvement in the trauma community assured that there was a place for ACS-TQIP.

However, nationwide and even international success of these programmes has

been hampered by several problems. Firstly, performance measures may not be sufficiently "granular" to inform quality improvement optimally. They provide specialty-wide morbidity and mortality estimates, but contain little information about procedure-specific outcomes, actionable processes of care or other data for guiding local quality improvement efforts. Secondly, measures are not surgeon-specific, and therefore are not designed to help individual surgeons assess their own performance. Finally, participation is very expensive. Annual subscription fees and the cost of a full-time nurse for data abstraction exceed $100,000 annually for most hospitals participating in ACS-NSQIP (158).

There is subsequently, very little uptake of such QI initiatives in more resource-limited environments despite good evidence of their effectiveness and potential cost saving. Potential strategies whereby QI initiatives in surgery can be implemented in more resource-limited environments are described below.

a. Creating global databases

The International Quality Improvement Collaborative for Congenital Heart Surgery, as well as the review on trauma QI's in LMIC's of the Asia-Pacific Region, showed conclusively that standardising data was the first step in improving injury care. The feasibility of international/global databases and their potential impact have been very elegantly described in a paper by Adil Haider and colleagues, 'Benchmarking of Trauma Care Worldwide: The Potential Value of an International Trauma Data Bank (ITDB)' (169,170).

The authors used observed/expected (O/E) mortality ratios to compare two trauma

centres (one European high [HIC] and one Asian [LMIC]) with centres in the North American National Trauma Data Bank (NTDB). Patients (>16 years) with blunt/penetrating injuries were included. Multivariate logistic regression adjusting for known predictors of trauma mortality was used to predict the expected deaths at each centre and to calculate O/E mortality ratios for benchmarking. A total of 375,433 patients from 301 centres were included from the NTDB (2002–2010). The LMIC trauma centre had 806 patients (2002–2010), whereas the HIC reported 1,003 patients (2002–2004).

The most important known predictors of trauma mortality were adequately recorded in all datasets. Mortality benchmarking revealed that the HIC Centre performed similarly to the NTDB centres [O/E = 1.11 (95 % confidence interval (CI) 0.92–1.35)], whereas the LMIC centre showed significantly worse survival [O/E = 1.52 (1.23–1.88)]. The authors concluded that using only a few key co-variates aggregated global trauma data could be used to adequately perform international trauma center benchmarking. The creation of the ITDB was feasible and recommended, as it could be a pivotal step towards improving global trauma outcomes (170).

b. More efficient risk-adjustment models

For the primary purpose of risk adjustment, ACS-NSQIP currently collects information about patient demographics (e.g. age, gender), procedure acuity, American Society of Anesthesiologists class, and more than 40 co-morbidities, ranging from the commonplace (e.g. diabetes, hypertension) to the relatively esoteric (e.g. esophageal varices, quadriplegia). It also requires 25 pre-operative and post-operative laboratory values. These variables allow for excellent risk-adjustment of both morbidity and

mortality rates, with C-statistics generally exceeding 0.85. Nonetheless, a large majority of risk factors currently collected may be unnecessary (158).

David Chang and colleagues recently undertook a study to identify a model with the fewest number of variables necessary to perform an adequate risk adjustment to predict any in-patient adverse event for use in resource-limited settings (171). All patients from the ACS-NSQIP database from 2005 to 2010 were included. Outcomes were in-patient mortality or any surgical complication captured by NSQIP. Models were built by sequential addition of pre-operative risk variables, selected by their area under the receiver operator characteristic curve (ROC). Among 863,349 patients, the single variable with the highest ROC was American Society of Anesthesiologists (ASA) classification (ROC 0.7127). ROC values reached 0.7923 with five variables (ASA classification, wound classification, functional status prior to surgery, albumin, and age) and 0.7945 with six variables. The sixth variable was one of the following: alkaline phosphatase, weight loss, principal anaesthesia technique, gender, or emergency status. The model with the highest discrimination, which did not require laboratory data, included ASA classification, functional status prior to surgery, wound classification, and age (ROC=0.7810). Inclusion of all 66 pre-operative variables produced little additional gain (ROC=0.8006). The authors concluded that six variables were sufficient to develop a risk adjustment tool for in-patient surgical mortality and morbidity (171). This study has important implications for the field of surgical outcomes research by improving the efficiency of data collection. This limited model can aid the expansion of risk-adjusted analyses to resource-limited settings worldwide.

c. Towards a more procedure-targeted approach

Procedure-specific outcomes assessment has obvious advantages for quality improvement. As currently provided by ACS-NSQIP, specialty level outcomes measures provide surgical leaders with a "bottom line" of their overall performance. They are also useful for monitoring specific outcomes relevant to almost any procedure (e.g. surgical site infection). However, specialty-level outcomes measures are not sufficiently granular for targeting specific procedures or subspecialty areas for improvement. They may even be falsely reassuring and result in missed opportunities for improvement. For example, a hospital's poor performance in colorectal surgery may be masked by better than average outcomes in bariatric surgery, or vice versa. Procedure- specific performance measures would alleviate such problems and better engage surgeons in their areas of interest or specialisation (158).

Focusing on procedure-specific outcomes assessment would also reduce the amount of information needed for risk adjustment. Examples include anastomotic leak after gastrointestinal surgery or vocal cord paralysis after thyroidectomy. In LMICs, efforts need to be focused on receiving the greatest return on investment of time and resources. For this reason, high-risk or high-volume procedures must be prioritised for audit.

d. Impact of emerging mobile health technology

The single major limiting factor for surgical outcomes research and quality improvement initiatives in the developing world, is the lack of reliable collection of actionable data. Where surgical outcomes databases do exist in LMICs, they are often rudimentary and incomplete (172). The lack of trained data capturers, the reliance on

retrospective folder reviews and poor record keeping are contributory. However, with the recent and rapid advancement of high functionality smart phones, and the growth of mobile phone subscriptions across the globe, there is widespread interest in using mobile health (m-Health) applications for routine collection of health data. m-Health, defined as medical and public health practices supported by the use of mobile devices, is widely considered to be a transformative force in the evolution of global health service delivery.

By exploiting the internet-capabilities of smart phones and mobile devices, near real-time transfer of data collected, using electronic forms on mobile applications can be achieved (173). Prospectively generated, clinician-entered electronic data is therefore, a realistic prospect. This should reduce the costs related to data processing and data entry in LMICs, and also increase the accuracy of data collected. This may provide the foundation for reliable surgical outcomes research and quality improvement initiatives in LMICs.

A practical, user friendly, mobile electronic Trauma Health Record (eTHR) for point of care data collection by front-line clinicians has recently been designed and implemented in the Trauma Centre at Groote Schuur Hospital. eTHR was designed to populate standard clinical reports, while wirelessly populating an electronic trauma registry, in real time, with standardised data (172). Within fifteen months after the implementation of eTHR, electronically generated records had replaced all previous hand-written record keeping from April 2014.

The web-based application REDCap (Research Electronic Data Capture) is a free, secure, web-based application designed to support data capture for research studies and is

available on a portable tablet or smart-phone (174). This application can also be used as a tool for clinician-entered data capture and for the efficient development of surgical outcomes databases.

e. Risk calculators

Following the acquisition of very large databases by the ACS-NSQIP, investigators have managed to develop risk calculators for clinicians to use remotely in order to calculate individual risk based on risk-adjustment models from their consortium. These have been done with relevance to colorectal, hepatobiliary, breast reconstruction and surgical oncology (176). The Universal Surgical Risk calculator, developed by Bilimoria *et al* in 2013, uses 21 pre-operative factors (demographics, co-morbidities, procedure) and regression models to predict 8 outcomes, based on the pre-operative risk factors. The development was based on 1,414,006 patients encompassing 1,557 unique Current Procedural Terminology (CPT) codes. The final calculator has excellent performance for mortality (c-statistic = 0.944), morbidity (c-statistic = 0.816) and 6 additional complications (c-statistics>0.8) (177).

For trauma cases, the Trauma and Research Network have made their probability-scoring tool, which is updated on an annual basis, available on-line. Using a combination of abbreviated injury score (AIS) per body region, age, GCS, intubation status, gender and co-morbidities, a probability for survival according to the Trauma and Research Network (TARN) experience is calculated (178). These remotely accessible risk calculators may prove to be useful tools to surgeons interested in surgical outcomes research in LMIC's

2.12 Summary

This literature review has demonstrated that there is a current movement away from seeing surgery as the, 'Neglected step-child of global health', and towards an, 'indivisible, indispensable part of health care and of progress towards universal health coverage.' It is in this era, that the importance of not only increased access and coverage to surgical services is highlighted, but also, ensuring that quality care is received. Quality in healthcare has been defined and various means explored to assess this pertaining to the practice of surgery. The current understanding of the essential components of quality improvement programmes, have been described and the literature reviewed, to ascertain their effectiveness in improving measured outcomes. The two gold standard programmes identified are the ACS-NSQIP and ACS-TQIP, and the methodology and analytical rationale of both have been described in detail.

It is clear that despite the evidence, suggesting QI programmes have the potential to save lives, improve the quality of life, and significantly reduce costs; there has been limited uptake in LMICs. After generating a better understanding of the strengths and limitations of the more established QI programmes, suggestions for implementing a structured surgical quality improvement programme in a more resource-limited environment have been explored.

The intention of this thesis is to learn from the ACS-NSQIP and ACS-COT TQIP programmes, as well as, the landmark studies reviewed, and to design and implement a structured surgical quality improvement programme, applicable for a health district in South Africa.

Chapter 3

Methods

3.1 Primary aim

To derive and validate prediction rules, which reliably predict the risk of mortality following major surgery in the Cape Metro West health district for benchmarking and quality improvement initiatives.

Without a reliable method to estimate risk-adjusted outcomes, benchmarking internally over time or externally is not possible. Benchmarking is critical for quality improvement initiatives. Adverse events are an inevitable consequence of major surgery. However, not all major events are expected or acceptable. The primary aim is to identify a method that takes the heterogeneity of surgical patients into account, and reliably discriminates expected versus unexpected adverse events.

3.2 Secondary aims

In order for the primary aim to be achieved within the Cape Metro West health district, a sustainable means of accurate data collection needs to be implemented, as this is currently not available. Using this newly acquired data, a prediction rule for adverse events needs to be identified or developed. Once developed, this prediction rule can be used to identify the expected (E) adverse events for a given surgical cohort, which can then be compared to what is observed (O). While risk-adjusted outcomes are important for patient-level analysis, hospital-level QI efforts also use this O/E metric for monitoring hospital- national- and global-level QI efforts.

A validated prediction rule will ensure that the outcome of any corrective action, which is subsequently implemented as a quality improvement initiative, is reliably measured.

3.3 Study hypothesis

Emerging m-Health technology, defined as medical and public health practices supported by the use of mobile devices, provides a solution to mitigate the lack of reliable surgical outcomes research and surgical quality improvement programmes in LMICs. This statement can be tested by addressing the following hypothesis:

Clinician-entered data generated using emerging m-Health technology can be used to predict major in-hospital adverse events following trauma or general surgery with adequate precision (c statistic >0.7) and calibration (p>0.05).

This hypothesis will be tested separately for trauma and general surgery patients following a common framework of objectives:

1. Develop a derivation dataset
2. Use the derivation dataset to develop a prediction rule, which reliably predicts an adverse event in a surgical patient, managed within our health district
3. Validate the prediction rule
4. Develop a simple scoring system from the prediction rule to pre-operatively identify high-risk surgical patients
5. Perform a risk-adjusted, benchmarking analysis to identify areas for quality improvement.

3.4 Study setting

Groote Schuur Hospital (GSH) is a government-funded, tertiary teaching hospital situated in Cape Town, South Africa. It is the main academic hospital of the University of Cape Town. Groote Schuur houses one of the busiest trauma referral units in the world, with an estimated 11,000 patients being seen in the trauma unit annually. An estimated 10 000 operations are performed annually in GSH; the risk-adjusted outcomes of these surgical operations are not formally audited or known. GSH is the central referral hospital in the Cape Metro West health district, and accepts referrals from three government-funded surgical units within this health district i.e. two district level hospitals (Mitchell's Plain District Hospital [MPH] and Victoria War Memorial Hospital [VWH]) and one regional level hospital (New Somerset Hospital [NSH]). Collectively, these hospitals make up the surgical referral base within the district, which serves an estimated catchment area of 2, 292, 000 uninsured patients.

This study describes the design and implementation of three structured surgical quality improvement programs within the Cape Metro West health district:

1. An Essentials Programme for Groote Schuur Hospital
2. A Procedure-targeted Programme for the Cape Metro West health district
3. A Trauma Quality Improvement Programme for Groote Schuur Hospital

3.5 Generic concepts employed for prediction rule development and validation

Prediction rules provide estimates for the risk (P) of a subject belonging to one of the two categories of a binary outcome. Values for the estimated risk from a prediction rule can theoretically range from zero to one (157,175). Regression modeling is the most

commonly used method for developing prediction rules. Through appropriate model assumption (e.g. structure, linearity and additivity), regression models can account for the relationship between multiple predictors and an outcome. All outcome measures audited and predicted in this thesis will be binary, and therefore, logistic regression analysis will be the primary method for developing the final co-efficients for inclusion in the prediction rules (175).

a. Methods employed to ensure a stable and valid model

The following techniques and principles of model building will be adhered to during the development of the prediction rules, whilst determining the number of potential predictors to include:

1. Univariate screen

All candidate predictors will be tested for association with outcome using either a chi-squared test or Fisher's exact test, preferably, with significance set at 0.1. Only those predictors with a significant association to the outcome will be eligible for inclusion in the final model.

2. Ensuring face validity and limiting use of computer algorithms.

Clinical acumen will override the result of the univariate screen for certain variables under consideration. Generic variables, such as age, will be included in all models to ensure face validity. Furthermore, predictors that have gained credibility in the developed world context and in the literature reviewed will be prioritised and in this manner, models will be built manually with very limited use of computer-generated algorithms.

3. Assessing for interaction and collinearity

An interaction may arise when considering the relationship among three or more variables, and describes a situation in which the simultaneous influence of two variables on a third is not additive. Collinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data.After each variable addition the change of the risk estimates and their associated standard errors were reviewed to screen for collinearity and a Wald test was performed to assess for significant interaction between variables.

4. The principle of parsimony

Parsimony pertains to the number of predictors to include in a prediction rule to avoid overtraining and over-fitting the model. In general, parsimony increases the simplicity and interpretability of a prediction rule, and preference will be given to simpler models.

5. 10 to 1 rule

In keeping with the principle of parsimony, a rule of thumb of 10 outcome events per predictor included in the model will be utilised.

6. Handling missing data

Only variables, which were greater than 80% complete in the derivation datasets were considered for inclusion in the prediction rules. For those variables, which were greater than 80% complete, but still had greater than 5% missing data, multiple imputation methods were used to complete the datasets. Instead of filling

in a single value, the distribution of the observed data is used to estimate multiple values that reflect the uncertainty around the true value. These values are then used in the analysis of interest and the results are combined. Each imputed value includes a random component whose magnitude reflects the extent to which other variables in the imputation model cannot predict it's true values (Johnson and Young, 2011; White et al, 2010). Thus, building into the imputed values a level of uncertainty around the "truthfulness" of the imputed values. Missing values were assumed to be missing at random. Outcome measures, which were not complete, were entered into the regression as missing in all datasets (175).

b. Methods considered for validating the prediction rule

Validation of a clinical prediction rule, involves obtaining 'honest' estimates of the performance of the rule in actual practice. This is done by illustrating transportability of the prediction rule in a different, but plausibly related, population or setting (geographical transportability) or in data collected by using slightly different methods than those used to create the original dataset (methodological transportability). Transportability refers to the validity and reproducibility, when applied to other datasets, from the same underlying population (175).

Probably the most accepted way to evaluate the merits of a prediction rule is by its performance in another dataset (the validation set) assessing discrimination and calibration. Measures of discrimination refer to the ability of the prediction rule to separate subjects with different outcomes into categories according to their values for the prediction rule. The ability of the prediction rule developed to discriminate will be assessed by constructing the ROC curve (Receiver Operating Characteristic).

Measures of calibration refer to the ability of the model's estimated risk to agree with actual outcomes within groups of subjects of similar predictive risk. Measures of the degree of calibration, commonly, take on the form of "observed versus expected" comparisons of the outcome. The sum of the predicted risks in a category provides an estimate of the expected number of outcomes for that category (E), which can be compared to the actual number of outcomes for that category (O). Measure of calibration of prediction rules developed here will be assessed by the Hosmer and Lemeshow goodness-of-fit statistic (GOF). A low statistic translates to a high p-value ($p>0.05$) and a better fitting model (175). This statistic is usually based, on grouping subjects by percentiles of the prediction rule into K categories. The form of this goodness-of-fit statistic is:

$$\mathbf{X2HL = \Sigma \{(Oi - Ei)\ 2 / Vi\}}$$

Where Oi = the observed number of outcomes in the ith category,
Ei = the expected number of outcomes in the ith category
Ei = nipi
ni = number of subjects in the ith row
pi = mean predicted outcome for the ni subjects in the ith row
Vi = nipi (1-pi)

An a priori level of >0.7 for discrimination, and >0.05 for calibration, has been set in the hypothesis.

3.6 Ethics

This study has been conducted with the necessary ethical approvals and considerations. The original protocol and subsequent amendments (17th April, 2015) for the PhD was approved by the University of Cape Town's Faculty of Health Sciences

Human Research Ethics Committee (HREC REF: 338/2014). The eTHR trauma database was approved previously by the same research ethics committee (HREC REF: R041/2014). As there has been no database approved for the General Surgery component of the study, individual consent was taken from participating patients. Please refer to the attached patient information sheet used for this purpose, Appendix 3. The researcher explained to all patients, that the drive for the study was to make sure GSH is offering patients, within the Cape Metro West health district, the best possible care. The following was explained to each patient who consented to contribute to this research:

- It is necessary to collect data from their hospital folder regarding their health, the operation they had and their post-operative course
- They could expect a telephonic follow-up one month after surgery and contact details would be needed for this purpose
- Their surgical care and follow-up appointments remained unchanged
- No extra visits would be needed
- There were no risks associated with this study
- There were no direct benefits to them for taking part in this study
- There was also no payment for taking part in this study
- The data we collected remained confidential by using a computer generated identifier
- The findings of the study may be published in a scientific journal or discussed at meetings, but no individual participants would be identified.

For the multicentre component of the study, senior management at each participating hospital was consulted and ethics approved by the Western Cape Provincial Health

Research Committee (WC-2015RP23-614). The findings of this research, was communicated back to all hospital managers and clinicians involved.

Chapter 4

The derivation dataset for an Essentials Programme

The following prospective dataset was developed; guided by the rationale of the ACS-NSQIP Essentials Programme (158). This derivation dataset was then used to derive and validate prediction rules for general surgery at Groote Schuur Hospital.

4.1 Materials and methods

a. Patient population

This was defined as a systematic sample of major vascular and general surgery operations performed under general, spinal, or epidural anaesthesia (final case selection occurred through a list of appropriate Current Procedural Terminology [CPT] codes).

Inclusion and exclusion criteria

The first 40 consecutive operations of major general and vascular surgery cases, on adults older than 12 years, during an 8-day cycle for 12 cycles, during a 3-month period from 1st April to 30th June 2014, were included. The following operations were excluded:

- Transplant surgery
- Trauma surgery
- Minor surgery;
  Abscesses, Incision and drainage cases, Lumpectomies, Lipomas, Fissures
- Re-look laparotomies
- Patients under 13 years
- Hernia repair- after the 3rd case
- Appendicectomies- after the 3rd case.

b. Data collection

Methods of data capture

A dedicated clinical auditor captured the data. Under the author's supervision, a standard set of variables was collected on consecutive general surgery patients, who met the inclusion criteria of the ACS-NSQIP Essentials Programme, as described above.

A pilot period of 4 cycles was performed in March 2016, prior to the study period, and was not included in the final dataset. During this pilot period, the main focus was on data variable definitions, and refining the final variable list for inclusion, as well as, to identify and solve any logistical issues of case identification and data capture. After the 1-month (4 cycle) pilot period, the clinical auditor functioned independently.

A random 8-day cycle sampling technique described by the ACS was used. Consecutive eligible patients were selected, starting with a different day of the week, in each 8-day cycle, with 40 patients per cycle. This ensured that patients from all days of the week were included, as recommended. Consecutive patients for inclusion were identified from the main theatre register, which captures all operations in the hospital. Patients were then followed-up in the ward post-operatively; data extraction took place, after the patient granted informed consent. If a patient was discharged prior to in-hospital follow-up, telephonic contact was made, and the records reviewed. The data collection sheet was downloaded from the ACS website. This datasheet was modified for local use, where relevant, after reviewing the first 4 pilot cycles. An on-line dataset with the final variables for inclusion was designed using REDCap. The clinical auditor used a mobile iPad, so that data extraction took place at the bedside and was entered directly into the

REDCap database. Only the author and the clinical auditor had access to the REDCap database. Once entered in the database, a 30-day follow-up from the date of surgery was undertaken by telephonic interview and chart review.

Variables captured

The downloaded ACS-NSQIP Essentials Programme worksheet is included in Appendix 1. This worksheet used for the ACS-NSQIP was made up of 105 variables, including 70 pre-operative risk factors, 11 variables about the operation, and 24 post-operative outcome variables.

The pre-operative risk factors included demographic data, some general health variables, general lifestyle variables, and major pulmonary, cardiac, hepatobiliary, renal, vascular, nervous system and nutritional/immune co-morbidities. Pre-operative laboratory values closest to the time of the operation were also collected.

Operative variables included CPT codes, anaesthesia type, ASA class, post-graduate year of the training surgeon, wound class, and operative and anaesthesia times. Operative complexity is no longer included in the NSQIP, as the dataset is large enough to code on individual CPT codes. In our cohort, local expert consensus coded the operative complexity into an ordinal scale 1-5, indicating increasing complexity. A summary of the changes made to the NSQIP pre-operative and post-operative variables for local adaptation are shown in Table 4.1.

Table 4.1 Pre-operative and intra-operative variable local adaptations to ACS-NSQIP

Variable	NSQIP Category	Local adaptation
Race	1. White 2. Black/ African American 3. American Indian/ Alaska native 4. Native Hawaiian/ Other pacific islander 5. Asian 6. Unknown	1. Black 2. Mixed ancestry 3. Indian 4. White 5. Asian 6. Unknown
Ethnicity	1. Hispanic yes/no	Omitted
Preferred language	1. English 2. Spanish	1. English 2. Afrikaans 3. Xhosa 4. Zulu 5. Sotho 6. Venda 7. Other
Transfer/ Origin status	1. Not transferred, admitted directly from home 2. Acute Care Hospital (inpatient status only) 3. Nursing home/ Chronic Care Facility/ Intermediate Care Unit 4. Transfer from other 5. Transfer from outside ED 6. Unknown	1. Victoria 2. New Somerset 3. Mitchell's Plain District 4. Self-referral 5. General practitioner 6. Clinic/ day hospital 7. Groote Schuur Casualty 8. Other

Surgical specialty	1. General surgery 2. Vascular 3. Thoracic 4. Cardiac 5. Orthopaedics 6. Neurosurgery 7. Urology 8. Otolaryngology 9. Plastics 10. Gynaecology	1. Vascular 2. Hepatobiliary 3. Acute Care 4. Colorectal 5. Surgical oncology 6. Unknown
Operation coding complexity	No longer included	Local expert consensus coded operation complexity into an ordinal scale 1-5
HIV Status	Not included	1. Positive 2. Negative 3. Unknown If positive, CD4 status; 1. CD4>350 2. CD4<350 3. Unknown If positive, ARV status; 1. Yes 2. No 3. Unknown
TB Status	Not included	1. Abdominal TB 2. Pulmonary TB 3. Disseminated TB 4. Unknown
Charlson's comorbidity index	Not used	Included as commonly referred to in surgical outcomes literature as a validated comorbidity score

c. Classification of endpoints

The post-operative outcomes included in the Essential worksheet, included vital status at 30 days, 21 different post-operative complications, unplanned return to operating room and length of stay. The Essentials NSQIP was originally designed to cover many different types of operations from many different surgical subspecialties, and therefore more generic types of variables were collected rather than more disease- or operation-specific variables. This made the Essentials NSQIP suitable for local adaptation. A summary of the endpoints included, as well as, any local adaptations made is shown in table 4.2.

Table 4.2 Classification of endpoints used in ACS-NSQIP and local adaptations

Variable	**NSQIP Category**	**Local adaptation**
	Yes/ no	Unchanged
Post-operative occurrences		
Wound occurrences	Superficial incisional SSI*	Date of occurrence changed to binary outcome yes/no
	Deep incisional SSI *	
	Organ/ space SSI *	
	Wound disruption	
Respiratory occurrences	Pneumonia	Date of occurrence changed to binary outcome yes/no. Free text column excluded.
	Unplanned ventilation	
	Pulmonary embolus	
	On ventilator >48hrs	
Urinary tract occurrences	Progressive renal insufficiency	Date of occurrence changed to binary outcome yes/no. Free text column excluded.
	Acute renal failure	
	Urinary tract infection	
CNS occurrences	Stroke/ CVA	Date of occurrence changed to binary outcome yes/no. Free text column excluded.
Cardiac occurrences	Cardiac arrest requiring CPR	Date of occurrence changed to

	Myocardial infarction	binary outcome yes/no. Free text column excluded.
Other occurrences	Bleeding requiring transfusion Deep vein thrombosis requiring therapy Sepsis Septic shock	Date of occurrence changed to binary outcome yes/no. Free text column excluded.
Discharge destination	1. Chronic care facility 2. Unskilled facility 3. Facility which was home 4. Home 5. Separate acute care 6. Rehabilitation 7. Expired 8. Unknown	1. Home 2. Rehabilitation hospital 3. Referring hospital 4. Expired 5. Still in hospital at 30-days 6. Unknown
Readmission	Readmission for any reason within 30 days of the principle procedure? Yes/no.	Unplanned readmission only including reason for readmission
Post-operative death	Postop death w/in 30 days Postop death >30 days	Postop death w/in30 days only
Unplanned reoperation	Unplanned return to the operating room for a surgical procedure w/in a 30-day post-operative period? Yes/no.	Unchanged including number of reoperations
Follow-up within 30 days	Were you able to follow the case for the full 30 days?	Unchanged including length of follow-up if <30 and method used
Complaint	Not included	Free text for patient complaint included

SSI Surgical site infection*

These endpoints were collected and verified using the following mechanisms:

(30-Day is defined from day of operation)

1. 30-Day telephonic interview
2. 30-Day folder review

3. 30-Day NHLS Laboratory results review
4. Monthly M&M meetings.

All data was entered and updated using the web-based REDCap program (174).

d. Analysis

The analysis of this derivation dataset was descriptive. The operations, which were included, and their associated variables, including the degree of missing data, were described according to the type of data and their degree of normality. Unadjusted rates of morbidity, mortality and resource utilisation were calculated and tabulated. Adequate sample size ensured that confidence intervals for any estimates were adequately tight. The endpoint of 30-day mortality occurred at a rate of 0.0767. With a sample size of 373 major operations in this test dataset, the 95% confidence interval around this estimate was 0.0518 - 0.1121.

4.2 Results

a. Population characteristics

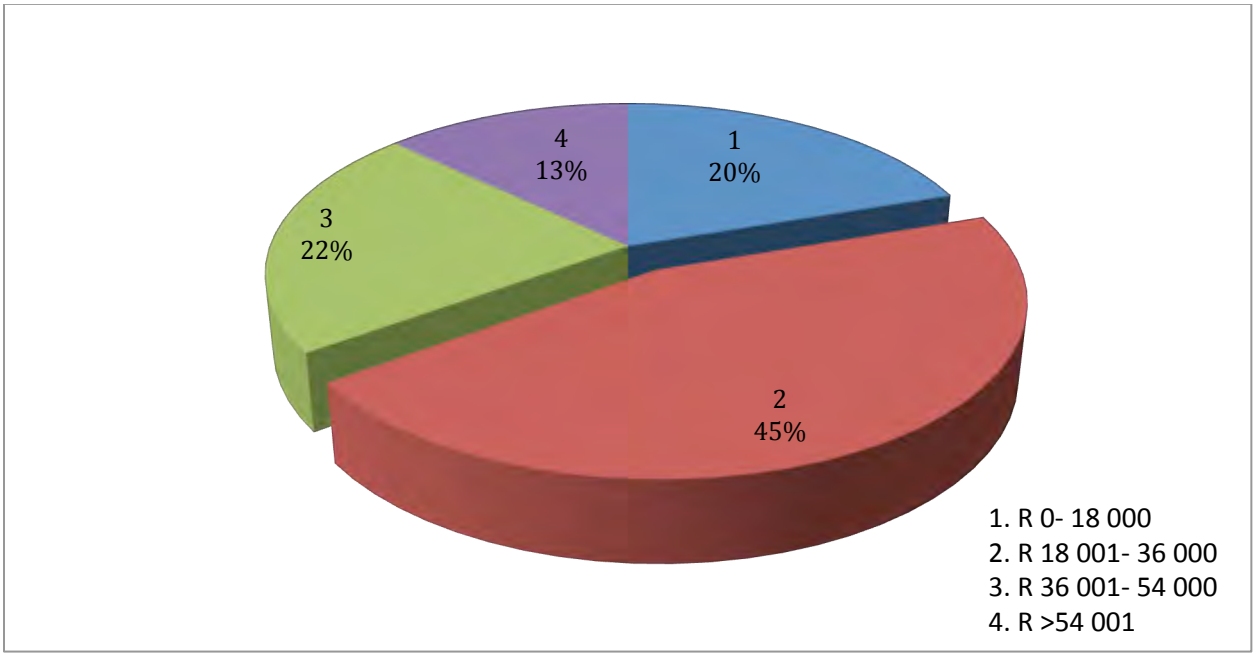
A total of 373 patients were included in the Essentials derivation dataset. The mean age of the cohort was 49.9 years (range 13 – 91), 57.3% were female, 20.4% were vascular patients and 79.6% were general surgery patients. The demographic, pre-operative risk assessment, laboratory and operative (including surgical profile) characteristics of the Essentials derivation dataset are presented in tables 4.3-4.6, respectively.

Table 4.3 Demographic characteristics of the Essentials derivation dataset

Demographic variable	Category	N (%)
Age	<18	9 (2.41)
	18-29	111 (29.76)
	30-39	41 (10.99)
	40-49	50 (13.40)
	50-59	68 (18.23)
	>=60	80 (21.45)
	Missing	14 (3.75)
Gender	Male	159 (42.63)
	Female	214 (57.37)
	Missing	0 (0)
Language	English	120 (32.17)
	Afrikaans	97 (26.01)
	Xhosa	43 (11.53)
	Other	4 (1.07)
	Missing	109 (29.22)
Race	Black	70 (18.77)
	Mixed ancestry	219 (58.71)
	White	37 (9.92)
	Indian	32 (8.58)
	Asian	5 (1.34)
	Missing	10 (2.68)
Referral	Victoria	36 (9.65)
	New Somerset	16 (4.29)
	Mitchell's Plain	40 (10.72)
	Self Referral	106 (28.42)
	General practitioner	7 (1.88)
	Clinic/ day hospital	47 (12.60)
	GSH casualty	43 (11.53)
	Other	57 (15.28)
	Missing	21 (5.63)

The greatest proportion of the cohort was in the >60 years age category (21.5%). English was the most frequently identified preferred language (32.2%) and the greatest proportion of patients was in the mixed ancestry racial category (58.7%). Self-referral was the most frequent method of referral (28.4%) and of the referring secondary hospitals in the Metro, Mitchell's Plain Hospital referred the most patients (10.7%). Residential suburbs were used to categorise patients into quartiles of median household income in Figure 4.1, as guided by the methodology employed by Stats SA.

Figure 4.1 Essentials derivation dataset in quartiles of median annual house income (ZAR)



The greatest proportion of the cohort lived in suburbs where the annual household income was between R18 001 – R36 000. Only 13% of the patients came from areas where the annual household income was greater than R54 001.

The results of the pre-operative risk-assessments are tabulated in table 4.4. The most prevalent co-morbidities in the cohort included hypertension (35.9%), diabetes mellitus (non-insulin dependent 15.8%, insulin dependent 5.4%), chronic obstructive pulmonary disease (6.3%), cancer (6.2%), acute renal failure (4.3%), HIV (3.7%), tuberculosis (2.6%) and congestive heart failure (2.1%). The majority of the patients scored a Charlson's co-morbidity score of 0 (59.5%), and none of the patients scored greater than four. Ventilatory support was required in 0.5% and dialysis in 1.6% perioperatively. Pre-operative blood transfusions were administered to 2.4% of the cohort. The HIV status was unknown in 86.3% and TB status was missing in 32.7%. There were no patients with pre-operative ascites. Excluding TB status, pre-operative risk assessment data was complete for >90% of the cohort.

Table 4.4 Pre-operative risk-assessment characteristics of the Essentials derivation dataset

Pre-operative risk assessment	**Category**	**N (%)**
BMI	Underweight (BMI <18.5)	18 (4.83)
	Normal (18.5≤ BMI< 25)	54 E14.48)
	Overweight (25 ≤ BMI<35)	28 (7.51)
	Obese (BMI ≥35)	37 (9.92)
	Missing	236 (63.27)
Diabetic	Insulin	20 (5.36)
	Non-insulin	59 (15.82)
	Nil	259 (69.44)
	Missing	35 (9.38)
Smoking within the year	Yes	145 (38.87)
	No	193 (51.74)
	Missing	35 (9.38)
Dyspnoea	Moderate exertion	123 (32.98)

	At rest	3 (0.80)
	None	218 (58.54)
	Missing	29 (7.77)
Functional status	Independent	289 (77.48)
	Partially dependent	53 (14.21)
	Dependent	1 (0.27)
	Unknown	1 (0.27)
	Missing	29 (7.77)
Ventilator dependent w/in 48hrs	Yes	2 (0.54)
	No	347 (93.03)
	Missing	24 (6.43)
Chronic Obstructive Pulmonary Disease	Yes	24 (6.43)
	No	325 (87.13)
	Missing	24 (6.43)
Ascites	Yes	0 (0)
	No	350 (93.83)
	Missing	23 (6.17)
Congestive Heart Failure	Yes	8 (2.14)
	No	340 (91.15)
	Missing	25 (6.70)
Hypertension	Yes	134 (35.92)
	No	216 (57.91)
	Missing	23 (6.17)
Acute Renal Failure	Yes	16 (4.29)
	No	335 (89.81)
	Missing	22 (5.90)
Dialysis	Yes	6 (1.61)
	No	344 (92.23)
	Missing	23 (6.17)
Cancer	Yes	23 (6.17)
	No	326 (87.40)
	Missing	24 (6.43)
Open wound	Yes	37 (9.92)
	No	313 (83.91)
	Missing	23 (6.17)
Steroids	Yes	12 (3.22)
	No	338 (90.62)
	Missing	23 (6.17)
>10% weightless	Yes	25 (6.70)
	No	324 (86.86)
	Missing	24 (6.43)
Bleeding disorder	Yes	4 (1.07)
	No	346 (92.76)
	Missing	23 (6.17)
Pre-operative *blood transfusion*	Yes	9 (2.41)

	No	341 (91.42)
	Missing	23 (6.17)
Sepsis	SIRS	62 (16.62)
	Sepsis	46 (12.33)
	Septic Shock	9 (2.41)
	No	234 (62.73)
	Missing	22 (5.90)
TB status	Abdominal TB	1 (0.27)
	Pulmonary TB	8 (2.14)
	Disseminated TB	1 (0.27)
	Unknown	131 (35.12)
	No	110 (29.49)
	Missing	122 (32.71)
HIV status	Positive	13 (3.69)
	Negative	35 (9.94)
	Unknown	304 (86.36)
Charlson's Comorbidity Score	0	222 (59.52)
	1	71 (19.03)
	2	45 (12.06
	3	9 (2.41)
	4	4 (1.07)
	5	0 (0)
	6	0 (0)
	Missing	22 (5.90)

The pre-operative laboratory variables are presented in Table 4.5. Six of the thirteen laboratory variables were missing or not recorded in over 60% of cases. The following laboratory variables were >80% complete- Sodium, Urea, Creatinine, Haemoglobin, WCC, Haematocrit and Platelets.

Table 4.5 Laboratory characteristics of the Essentials derivation dataset

Laboratory parameter	Total N (%)	Missing N (%)	Mean* / median	95% CI*/ IQR
Sodium	336 (90.01)	37 (9.99)	136.89*	117-161*
Ur	325 (87.13)	48 (12.18)	71	35-609
Creatinine	336 (90.08)	37 (9.92)	71	59-90
Albumin	103 (27.61)	270 (72.39)	38.4	4.7-53
Bilirubin	119 (31.91)	254 (68.09)	7	2-70
Alkaline phosphatase	117 (31.37)	256 (68.63)	120.97*	1-702*
*Aspartate aminotransferase^*	49 (13.14)	324 (86.86)	Omitted	Omitted
Haemoglobin	330 (88.47)	43 (11.53)	12.94*	5.1-18.4*
WCC	334 (89.54)	39 (10.46)	9.12*	3.51-77*
Haematocrit	314 (84.18)	59 (15.82)	0.39	0.35-0.44
Platelets	330 (88.47)	43 (11.53)	363.26*	3.86-1388*
PTT	66 (17.69)	307 (82.31)	32.05	25.9-52.3
INR	153 (41.02)	220 (58.98)	1.03*	0.83-19.1*

^*Aspartate aminotransferase was omitted, as it was incorrectly recorded and was misclassified.*

The operative characteristics are presented in table 4.6. Emergency procedures made up as much as 36.7% of the cohort. The vast majority was under general anaesthesia (95.7%). There were no ASA category 5 cases and only 2.7% were classified ASA category 4. In 22.8% of the anaesthetic records, ASA grading was not completed. Only 9.11% of operations were classified as contaminated/ dirty.

Table 4.6 Operative characteristics of the Essentials derivation dataset

Operative variable	Category	N (%)
Status	Elective	236 (63.27)
	Emergency	137 (36.73)
	Missing	0 (0)
Anesthetic type	General	357 (95.71)
	Spinal	13 (3.49)
	Epidural	2 (0.54)
	Missing	1 (0.27)
ASA	1	71 (19.03)
	2	125 (33.51)
	3	93 (24.93)
	4	10 (2.68)
	5	0 (0)
	Missing	74 (19.84)
Specialty	Vascular	76 (20.38)
	Hepatobiliary	32 (8.58)
	Acute care	136 (36.46)
	Colorectal	48 (12.87)
	Surgical oncology	61 (16.35)
	Unknown	20 (5.36)
	Missing	0 (0)
Surgeon qualification	Sub-specialist	77 (20.53)
	Fellow	53 (14.13)
	Consultant	56 (14.93)
	Registrar	152 (40.53)
	Medical officer	15 (4.00)
	Missing	22 (5.87)
Wound classification	Clean	205 (54.96%)
	Clean-contaminated	134 (35.92)
	Contaminated	15 (4.02)
	Dirty	19 (5.09)
	Missing	0 (0)
Operation complexity coding	1	42 (11.26)
	2	62 (16.62)
	3	192 (51.47)
	4	63 (16.89)
	5	14 (3.75)
	Missing	0 (0)

Operative data including the surgical specialty and qualification of the primary operating surgeon are presented graphically in figures 4.2 and 4.3, respectively. When the records with missing primary surgeon data were removed, surgical registrars were, most commonly the primary operating surgeons in the cohort (41.9%), followed by subspecialists (21.5%), and jointly by fellows and consultants (16.0%). The Acute Care Surgical specialty contributed the most cases (36.4 %), followed by the Vascular (20.4%), Surgical Oncology (16.4%), Colorectal (12.9%) and Hepatobiliary (8.6%) specialties.

Figure 4.2 Operative activity in the Essentials derivation dataset by most senior surgeon

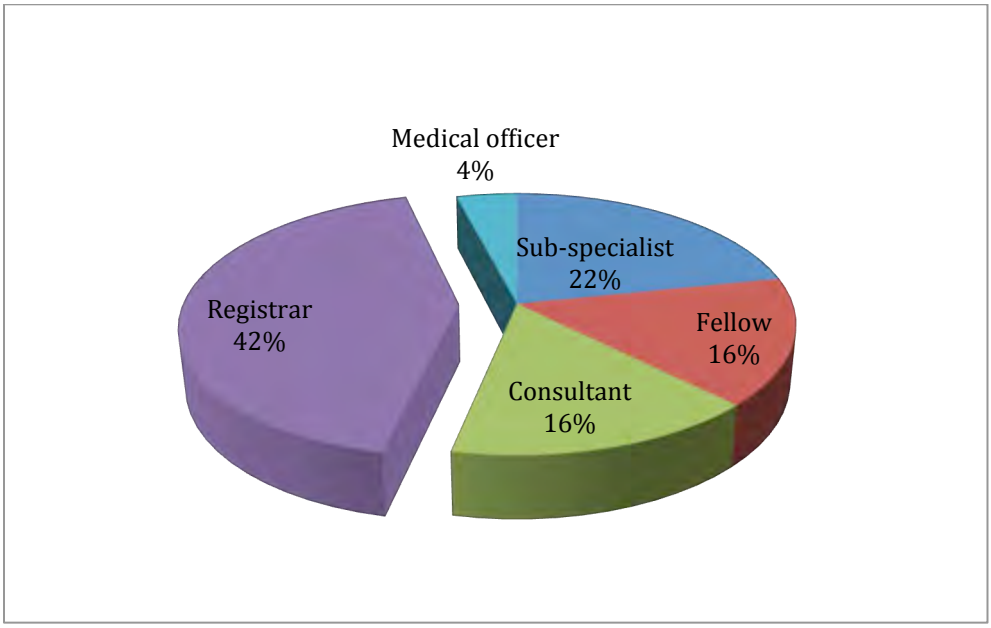
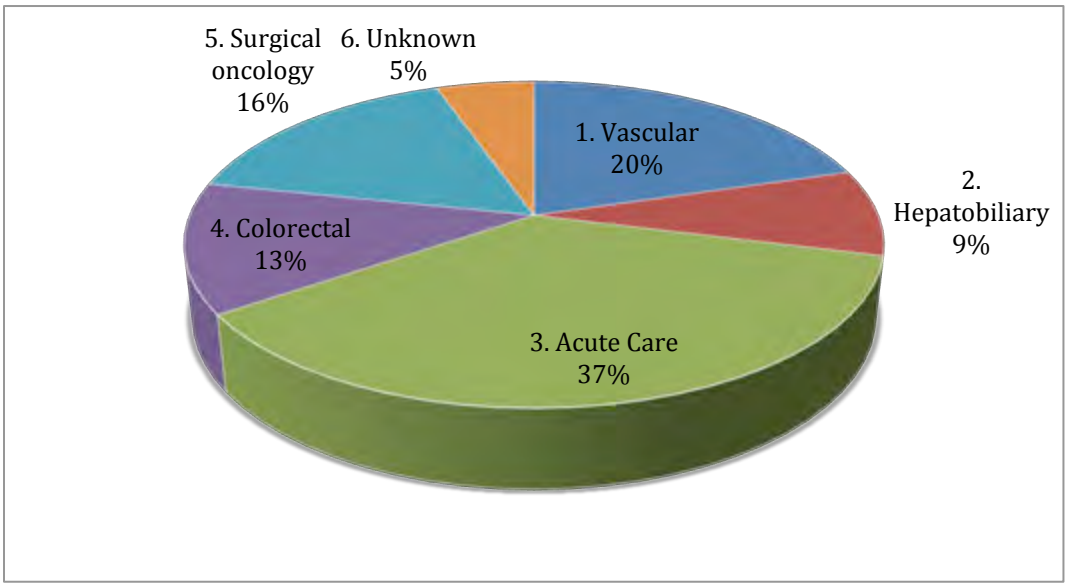


Figure 4.3 Essentials derivation dataset by surgical specialty



All operations were coded by CPT codes to ensure appropriate inclusion into the dataset. The most common CPT codes were 49000 (exploratory laparotomy -25.3%), 19303 (simple mastectomy -10.1%), 47562 (laparoscopic cholecystectomy -7.5%), 44950 (open appendectomy -4.3%), 49505 (open inguinal hernia -2.9%) and 60210 (partial thyroidectomy -2.9%). Inclusion of both hernias and appendectomies were limited to three cases per cycle by the inclusion criteria defined by the ACS-NSQIP.

b. Occurrence of endpoints

The occurrences of endpoints are presented in table 4.7. The commonest endpoint was Surgical Site Infection (20.1%). The majority of the patients (82.3%) were discharged directly home, and only 3.2% were referred back to a referral hospital. Discharge information was missing in 5.1%. Twenty one patients died post-operatively according to the records. However, after folder, telephonic and M&M review, a further 3

deaths were identified increasing the death rate from 5.63% (95% CI 3.53-8.52) to 7.67% (95% CI 4.97–11.2).

Table 4.7 Occurrence of endpoints in the Essentials derivation dataset

Outcome occurrence	Category	N (%)
Wound occurrence	SSI*	48 (12.87)
	Deep SSI*	21 (5.63)
	Organ space SSI*	6 (1.61)
	Wound disruption	0 (0)
Respiratory	Pneumonia	8 (2.14)
	Unplanned intubation	2 (0.54)
	Pulmonary embolus	1 (0.27)
	On ventilator >48 hrs	8 (2.14)
Urinary tract	Progressive renal insufficiency	10 (2.68)
	Acute renal failure	5 (1.34)
	Urinary tract infection	6 (1.61)
CNS	Stroke/ CVA	1 (0.27)
Cardiac	Cardiac arrest requiring CPR^	8 (2.14)
	Myocardial infarction	1 (0.27)
Other	Bleeding requiring transfusion	10 (2.68)
	Deep vein thrombosis	3 (0.80)
	Sepsis	8 (2.14)
	Septic shock	7 (1.88)
	Other	15 (4.02)
Discharge destination	Home	307 (82.31)
	Rehabilitation hospital	10 (2.68)
	Referring hospital	12 (3.22)
	Expired	21 (5.63)
	Still in hospital at 30 days	4 (1.07)
	Missing	19 (5.09)

* *SSI Surgical Site Infections*
^ CPR Cardiopulmonary resuscitation

After an extensive 30-day follow-up, the ACS-NSQIP outcome measures were determined as presented in table 4.8 with 95% CI. A 30-day telephonic follow-up rate of 71.82% of the cohort was achieved.

Table 4.8 NSQIP endpoint measures in the Essentials derivation dataset

Endpoints NSQIP	**Rate %**	**95% CI**
Readmission rate	11.85	8.57-15.85
Post-operative mortality rate	7.67	4.97-11.12
Unplanned reoperation rate	10.25	7.16-14.09
Rate of 30-day follow-up	71.82	66.92-76.35
Complaint reported at 30 days	29.22	23.58-35.37

The data from this dataset was then used to derive and validate a prediction rule for general and vascular surgery at GSH as described in Chapter 6.

Chapter 5

The derivation dataset for a Procedure-targeted Programme

A separate dataset was developed prospectively for a Procedure-targeted programme. The following prospective dataset was based on the ACS-NSQIP procedure-targeted programme (158). This derivation dataset was then used to derive and validate prediction rules for a targeted procedure within the Cape Metro West health district.

As described by Birkmeyer *et al* (23), measuring quality of care for procedures, which are both common and relatively high risk are best assessed directly using risk-adjusted measures of morbidity and mortality. This is why emergency exploratory laparotomy has been chosen as the targeted procedure to audit in the Cape Metro West health district.

5.1 Materials and methods

a. Patient population

All adult patients (>12 years) undergoing emergency exploratory laparotomy surgery within the Cape Metro West health district during the 3-month period between 1st February and 30th April were included in the study. Obstetric, Gynaecology and Trauma patients were excluded.

b. Data collection

Two surgical clinicians at each of the hospitals (GSH, NSH, MPH and VWH) were responsible for the prospective data capture. All patients suitable for inclusion were identified pre-operatively and informed consent was obtained. A dataset was developed

using REDCap and clinicians updated data entry on a daily basis using their mobile phones or iPads in the wards post-operatively. Patients were then followed-up in the ward post-operatively until hospital discharge.

Variables

Using the experience gained from the development of the Essentials dataset, a refined set of variables was collected for a Procedure-targeted dataset. These variables were chosen because they were either: 1) identified as being easily available during the Essentials dataset development, 2) considered to be locally relevant (e.g. HIV status), 3) included in the efficient model by Chang *et al* referenced in the literature review (171), 4) relevant for such a Procedure-targeted database or 5) considered to be standard variables, which added practical value or face validity to the database. A less labour-intensive set of data variables were needed to ensure success of the development of this multi-centre, clinician-entered database.

A total of 77 variables were collected in the Procedure-targeted dataset, compared to 119 in the Essentials dataset. The data dictionary for the Procedure-targeted dataset is included in Appendix 2, and a summary of the changes made to the pre-operative and intraoperative variables collected in the Essentials dataset are tabulated below.

Table 5.1 Exclusion, additions and amendments made to the Essentials dataset for the Procedure-targeted dataset development

Variable Exclusion	**Rationale**
Cycle and case number	Random 8 day sample cycling not used
Operation, CPT code and elective status	Only emergency laparotomies included
Language	No benefit identified
Referral	District surgical hospitals - all referrals likely to be from primary care units in the area
Anesthetic, type	100% expected to occur under general anesthetic
Status	All emergency
Pre-operative *sodium, GGT/ AST, ALP, Hb, WCC, bilirubin*	No hypothesised benefit in this context
Dyspnoea	Impression of collinearity with ASA, functional status and smoking status
Suburb	Broadly captured by hospital
Named surgeon and assistant	Keeping emphasis away from individualizing blame
TB	Little value gained in the essentials dataset
Ventilation	Not intentionally*
Variable addition/ amendment	**Rationale**
Hospital	Multi-centre database
Arterial blood gas results	Hypothesised to be relevant in an emergency procedure setting
Surgeon and anaesthetic qualification	Variable in more limited-resource settings and secondary level facilities.
Operation complexity	Exploratory laparotomy includes a heterogeneous group of operations. Operations were grouped into Negative, foregut, midgut, hindgut and vascular procedures.
Incision type	To assess use and outcome of laparoscopic versus open approaches
Procedure performed	Compatible with the notion of a procedure-targeted approach
WHO Checklist completion rate, DVT and antibiotic prophylaxis	Process variable considered to be important
Variable amendment	**Rationale**
Comorbidities	Refined into Charlson's comorbidity index. Converted 14 comorbidity variables into a 6 point validated index.

**The variable ventilator status was excluded unintentionally.*

c. Classification of endpoints

The endpoints collected for the Procedure-targeted dataset were identical to those in the Essentials dataset, but the ascertainment thereof was different. In the Procedure-targeted dataset, the clinicians on their daily ward rounds verified these endpoints, when the dataset was updated. In this manner, patients were followed-up daily until hospital discharge. As there was less variability in the manner of endpoint determination in the Essentials dataset than in the procedure-targeted dataset, and follow-up was only until hospital discharge, only 22 data variables compared to 49 were needed post-operatively in this dataset. All hospitals in the Metro contribute to the combined M&M meeting, where complications were further verified.

The GSH clinical auditor (using the following mechanisms) followed up a subset of patients for the full 30-Days:

1. 30-Day telephonic interview
2. 30-Day folder review
3. 30-Day NHLS Laboratory results review.

This rationale for this step was to review whether or not the full 30-Day follow-up was really necessary.

d. Analysis

The analysis of these derivation datasets was descriptive. The operations that were included and their associated variables, including the degree of missing data, were described according to the type of data and their degree of normality. Unadjusted rates of morbidity, mortality and resource utilisation were calculated and tabulated. Adequate sample size ensured that confidence intervals for any estimates were adequately tight.

The endpoint of 30-day mortality occurred at a rate of 0.0767. With a sample size of 320 major operations in this test dataset, the 95% confidence interval around this estimate was 0.0361 - 0.0912.

5.2 Results

a. Population characteristics

A total of 320 patients were included in the Procedure-targeted derivation dataset. The demographics, pre-operative risk assessment, laboratory and operative (including surgical profile) characteristics of the Cape Metro Procedure-targeted dataset are presented in tables 5.2-5.5, respectively.

Table 5.2 Demographic characteristics of the Procedure-targeted derivation dataset

Demographic variable	Category	Groote Schuur N (%)	Mitchell's Plain N (%)	New Somerset N (%)	Victoria N (%)	Total N (%)
Patients		109 (34.06)	56 (17.50)	101 (31.56)	54 (16.88)	320 (100)
Age category	<18	5 (4.59)	13 (23.21)	13 (12.87)	12 (22.22)	43 (13.44)
	18-29	19 (17.43)	17 (30.36)	30 (29.70)	12 (22.22)	78 (24.38)
	30-39	17 (15.60)	11 (19.64)	30 (29.70)	7 (12.96)	65 (20.31)
	40-49	23 (21.10)	7 (12.50)	13 (12.87)	11 (20.37)	54 (16.88)
	50-59	20 (18.35)	2 (3.57)	8 (7.92)	7 (12.96)	37 (11.56)
	>60	25 (22.94)	6 (10.71)	7 (6.93)	5 (9.26)	43 (13.44)
	Missing	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Gender	Male	55 (50.46)	23 (41.07)	44 (43.56)	15 (27.78)	137 (42.81)
	Female	54 (49.54)	33 (58.93)	57 (56.44)	39 (72.22)	183 (57.19)
	Missing	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Race	Black	29 (26.61)	22 (39.29)	41 (40.59)	10 (18.52)	102 (31.87)
	White	11 (10.09)	0 (0)	13 (12.87)	3 (5.56)	27 (8.52)
	Mixed ancestry	65 (59.63)	34 (60.71)	41 (40.59)	41 (75.93)	181 (57.10)
	Indian	4 (3.67)	0 (0)	3 (2.97)	0 (0)	7 (2.19)
	Asian	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Missing	0 (0)	0 (0)	3 (2.97)	0 (0)	3 (0.94)

The greatest proportion of patients for all participating hospitals were in the 18-29 age group category (range 22.22% - 30.36%), except for Groote Schuur where the greatest proportion of patients was in the >60 years (22.94%). There was a slightly higher predominance of females in the cohort (57.19%). There was consistency in the racial distribution amongst hospitals; 57.10% of the cohort was of mixed ancestry and 32.18% were Black. Asian, Indian and White were the minority racial groups at all participating hospitals. There was less than three percent missing data for all demographic variables.

Table 5.3 Pre-operative risk-assessment characteristics of the Procedure-targeted derivation dataset

Pre-operative risk assessment	Category	Groote Schuur N(%)	Mitchell's Plain N (%)	New Somerset N (%)	Victoria N (%)	Total N (%)
Diabetic	Insulin	5 (4.59)	0 (0)	1 (0.99)	2 (3.70)	6 (1.88)
	Non-insulin	10 (9.17)	0 (0)	1 (0.99)	2 (3.70)	13 (4.06)
	Nil	93 (85.32)	56 (100)	98 (97.03)	52 (96.30)	299 (93.44)
	Missing	1 (0.92)	0 (0)	1 (0.99)	0 (0)	2 (0.62)
Smoking within the year	Yes	45 (41.28)	18 (32.14)	25 (24.75)	19 (35.19)	107 (33.44)
	No	61 (55.96)	38 (67.86)	71 (70.30)	32 (59.26)	202 (63.12)
	Missing	3 (2.75)	0 (0)	5 (4.95)	11 (3.44)	11 (3.44)
Functional status	Independent	95 (87.16)	54 (96.43)	78 (77.23)	47 (87.04)	274 (85.62)
	Partially dependent	10 (9.17)	0 (0)	0 (0)	0 (0)	10 (3.12)
	Dependent	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	Unknown	2 (1.83)	2 (3.57)	22 (21.78)	6 (11.11)	32 (10.0)
	Missing	1 (0.92)	0 (0)	1 (0.99)	3 (0.93)	1 (0.31)
Pre-operative Sepsis	None	17 (16.16)	8 (14.28)	15 (15.38)	10 (18.52)	50 (15.63)
	SIRS	70 (64.65)	39 (70.12)	72 (71.79)	34 (62.96)	215 (67.19)
	Sepsis	15 (14.40)	7 (12.4)	13 (12.82)	7 (2.06)	42 (13.12)
	Septic shock	7 (4.79)	2 (3.2)	0 (0)	3 (5.56)	12 (3.75)
	Missing	0 (0)	0 (0)	1 (0.99)	0 (0)	1 (0.31)

Charlson's comorbidity index	0	53 (48.62)	46 (82.14)	88 (87.13)	49 (90.74)	236 (73.75)
	1	24 (22.02)	5 (8.93)	10 (9.90)	3 (5.56)	42 (13.12)
	2	23 (21.10)	2 (3.57)	2 (1.98)	2 (3.70)	29 (9.06)
	3	2 (1.83)	1 (1.79)	0 (0)	0 (0)	3 (0.94)
	4	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	5	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	6	6 (5.50)	2 (3.57)	1 (0.99)	0 (0)	9 (2.81)
	Missing	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
HIV status	Positive	10 (9.17)	5 (8.93)	13 (12.87)	1 (1.85)	29 (9.06)
	Negative	16 (14.68)	10 (17.86)	3 (2.97)	14 (25.93)	43 (13.44)
	Unknown	83 (76.15)	41 (73.21)	84 (83.17)	39 (72.22)	247 (77.19)
	Missing	0 (0)	0 (0)	1 (0.99)	0 (0)	1 (0.31)

Groote Schuur Hospital did have the highest proportion of patients with diabetes mellitus (13.76%), a highest proportion of patients with dependent or partially dependent functional status scores (10.09%), the highest proportion of septic or septic shocked patients (19.19%) and highest mean Charlson's co-morbidity index (1.07: 95% CI 0.78 – 1.36) suggesting, to some degree, the highest risk patients are being appropriately referred to the tertiary referral hospital. HIV status was poorly documented with unknown status being >70% for all participating hospitals. The highest confirmed HIV prevalence was at New Somerset (12.87%). Laboratory findings are presented in Table 5.4. Rates of missing data were high and ranged 15-93%.

Table 5.4 Laboratory characteristics of the Procedure-targeted derivation dataset

Laboratory parameter	**Total N (%)**	**Missing N (%)**	**Mean*/ median**	**95%CI*/ IQR**
Urea	258 (80.63)	62 (19.37)	4.8*	3.4 - 7.3*
Creatinine	269 (84.06)	51 (15.94)	74	62 - 96
Albumin	47 (14.69)	273 (85.31)	32.11*	29.10 - 35.11*
Haematocrit	157 (49.06)	163 (50.94)	0.39*	0.33 - 0.47*
INR^	24 (7.50)	296 (92.50)	1.26	1.15 - 1.38
Ph.	84 (26.25)	236 (73.75)	7.41	7.38 - 7.47
Bicarbonate	82 (25.63)	238 (74.37)	23.71*	20.54 - 36.81*
Base excess	83 (25.94)	237 74.06)	-0.82	-3.62 - 1.81
lactate	69 (21.56)	251 (78.44)	1.63	1.0 - 2.2

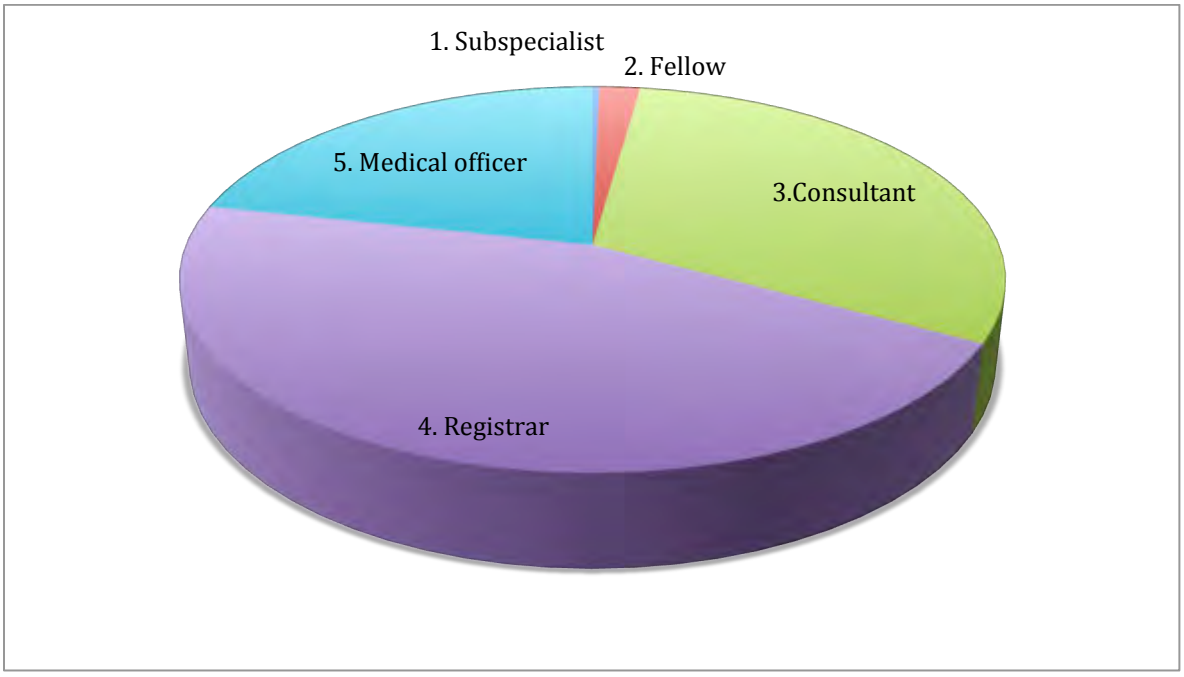
As seen in the Essentials derivation dataset, the majority of patients in this dataset were classified in either ASA category 1 or 2 (74.06%). When dichotomized into ASA ≥ 3, Groote Schuur Hospital had the highest proportion (26.60%). Only Groote Schuur Hospital had surgical specialty representation. The Acute Care Surgery specialty contributed the most patients (70.64%), followed by Colorectal (15.60%), Hepatobiliary (10.09%) and Vascular (1.83%) subspecialties. Approximately a quarter of the operations were classified as contaminated/ dirty (23.44%). The classification system for operative complexity was comprehensively completed with the majority of cases involving midgut structures (54.06%). There was a negative laparotomy rate of 15%. Two operations were vascular cases (ruptured AAA's).

Table 5.5 Operative characteristics of the Procedure-targeted derivation dataset

Operative variable	Category	Groote Schuur N(%)	Mitchell's Plain N (%)	New Somerset N(%)	Victoria N(%)	Total N(%)
ASA	1	31 (28.44)	37 (66.07)	66 (65.35)	31 (57.41)	165 (51.56)
	2	35 (32.11)	6 (10.71)	24 (23.76)	7 (12.96)	72 (22.50)
	3	14 (12.84)	6 (10.71)	5 (4.95)	4 (7.41)	29 (9.06)
	4	14 (12.84)	2 (3.57)	2 (1.98)	0 (0)	18 (5.62)
	5	1 (0.92)	2 (3.57)	0 (0)	1 (1.85)	4 (1.25)
	Missing	14 (12.84)	3 (5.36)	4 (3.96)	11 (20.37)	32 (10.0)
Speciality	Vascular	2 (1.83)	0 (0)	0 (0)	0 (0)	2 (0.62)
	Hepatobiliary	11 (10.09)	0 (0)	0 (0)	0 (0)	12 (3.75)
	Acute care	77 (70.64)	56 (100)	101 (100)	54 (100)	286 (89.38)
	Colorectal	18 (15.60)	0 (0)	0 (0)	0 (0)	18 (5.62)
	Surgical oncology	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unknown	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Missing	2 (1.83)	0 (0)	0 (0)	0 (0)	2 (0.62)
Surgeon qualification	Sub-specialist	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	Fellow	6 (5.50)	0 (0)	0 (0)	0 (0)	6 (1.88)
	Consultant	25 (22.94)	22 (39.29)	36 (35.64)	14 (25.93)	97 (30.31)
	Registrar	73 (66.97)	18 (32.14)	46 (45.54)	6 (11.11)	143 (44.69)
	Medical officer	2 (1.83)	16 (28.57)	17 (16.83)	33 (61.11)	68 (21.25)
	Missing	2 (1.83)	0 (0)	2 (1.98)	1 (1.85)	5 (1.56)
Wound classification	Clean	36 (33.03)	3 (5.36)	72 (71.29)	26 (48.15)	137 (42.81)
	Clean-contaminated	49 (44.95)	13 (23.21)	26 (25.74)	15 (27.78)	103 (32.19)
	Contaminated	10 (9.17)	21 (37.50)	1 (0.99)	12 (22.22)	44 (13.75)
	Dirty	12 (11.01)	19 (33.93)	0 (0)	0 (0)	31 (9.69)
	Missing	2 (1.83)	0 (0)	2 (1.98)	1 (1.85)	5 (1.56)
Operative complexity	Foregut	10 (9.17)	10 (17.86)	9 (8.91)	7 (12.96)	36 (11.25)
	Hindgut	17 (15.60)	5 (8.93)	8 (7.92)	5 (9.26)	35 (10.94)
	Midgut	43 (39.45)	35 (62.50)	62 (61.39)	33 (61.11)	173 (54.06)
	Negative	9 (8.26)	3 (5.36)	5 (4.95)	4 (7.41)	48 (15.0)
	Other	26 (23.85)	3 (5.36)	5 (4.95)	15 (14.85)	4 (7.41)
	Vascular	2 (1.83)	0 (0)	0 (0)	0 (0)	2 (0.62)

The qualification of the primary operating surgeon is represented graphically in figure 5.1.

Figure 5.1 Qualification of the primary operating surgeon in the Procedure-targeted derivation dataset



b. Occurrence of endpoints

The occurrences of endpoints stratified by hospital are presented in table 5.6. The most common occurrences were Surgical Site Infections (6.25%), followed by sepsis (4.06%), and acute renal failure (3.12%). The majority of the patients (89.69%) were discharged directly home with only 1.88% still in hospital after 30 days. Discharge information was missing in 0.31%.

Table 5.6 Occurrence of endpoints in the Procedure-targeted dataset by hospital

Outcome occurrences	Category	Groote Schuur N(%)	Mitchell's Plain N (%)	New Somerset N(%)	Victoria N(%)	Total N(%)
Wound occurrence	SSI*	6 (5.50)	1 (1.79)	0 (0)	2 (3.70)	9 (2.81)
	Deep SSI*	3 (2.75)	3 (5.36)	0 (0)	0 (0)	6 (1.88)
	Organ space SSI*	2 (1.83)	1 (1.79)	0 (0)	0 (0)	3 (0.94)
	Wound disruption	2 (1.83)	0 (0)	0 (0)	0 (0)	2 (0.62)
Respiratory	Pneumonia	0 (0)	0 (0)	1 (0.99)	0 (0)	1 (0.31)
	Unplanned intubation	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	Pulmonary embolus	0 (0)	0 (0)	1 (0.99)	0 (0)	1 (0.31)
	On ventilator >48 hrs	7 (6.42)	0 (0)	1 (0.99)	0 (0)	8 (2.50)
Urinary tract	Progressive renal insufficiency	1 (0.92)	0 (0)	1 (0.99)	0 (0)	2 (0.62)
	Acute renal failure	9 (8.26)	0 (0)	1 (0.99)	0 (0)	10 (3.12)
	Urinary tract infection	3 (2.75)	0 (0)	1 (0)	0 (0)	3 (0.94)
CNS	Stroke/ CVA	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Cardiac	Cardiac arrest requiring CPR	2 (1.83)	0 (0)	2 (1.98)	0 (0)	4 (1.25)
	Myocardial infarction	1 (0.92)	0 (0)	1 (0.99)	0 (0)	2 (0.62)
Other	Bleeding requiring transfusion	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)
	Deep vein thrombosis	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Sepsis	12 (11.01)	0 (0)	0 (0)	1 (1.85)	13 (4.06)
	Septic shock	3 (2.75)	0 (0)	1 (0.99)	0 (0)	4 (1.25)
	Other	17 (5.31)	1 (1.79)	1 (0.99)	0 (0)	19 (5.93)
Discharge destination	Home	89 (81.65)	53 (94.64)	94 (93.07)	51 (94.44)	287 (89.69)
	Rehabilitation hospital	0 (0)	1 (1.79)	1 (0.99)	0 (0)	2 (0.62)
	Referring hospital	2 (1.83)	0 (0)	0 (0)	0 (0)	2 (0.62)
	Expired	15 (13.76)	1 (1.79)	3 (2.97)	3 (5.56)	21 (6.56)
	Still in hospital at 30 days	4 (3.67)	1 (1.79)	1 (0.99)	0 (0)	6 (1.88)
	Missing	1 (0.92)	0 (0)	0 (0)	0 (0)	1 (0.31)

SSI Surgical Site Infections*

The ACS-NSQIP endpoint measures with 95% CI are presented in table 5.7. As all patients were only followed-up until hospital discharge, 30 day follow-up measures were not included in the procedure-targeted dataset.

Table 5.7 NSQIP endpoint measures in the Procedure-targeted dataset

Endpoints NSQIP	Rate (%)	95% CI
Readmission rate	11.31	7.74 - 16.25
Post-operative mortality rate	6.23	3.99 - 9.58
Unplanned reoperation rate	10.41	6.99 - 15.22

A sample of 90 patients from the Groote Schuur cohort was also followed up at 30 days from the date of surgery. Of these folders, 95.6% were available for retrospective review at 30-day follow-up and 4.4% were missing. Blood, urine and pus swab cultures were performed in 24.4%, 24.4% and 12.4% of the cases, respectively. Histology was sent at 78.19% of operations. Only 46.7% of the 90 cases were reached telephonically, of which, 23.8% reported a complication and 26.2% reported a complaint. All reported complications were confirmed according to our records. The agreement of the occurrence of endpoints as per the in-hospital follow-up compared to the 30-day follow-up is presented in table 5.8. Kappa statistics for all endpoint measures were greater than 0.81, translating to almost perfect agreement.

Table 5.8 Interobserver agreement between in-hospital and 30-Day endpoint measures in the Procedure-targeted dataset

Endpoint measure	Agreement %	Kappa Statistic	95% CI
Post-operative complication rate	93.90	0.87	0.76 - 0.98
Readmission rate	97.53	0.86	0.67 - 1.0
Post-operative mortality rate	97.67	0.87	0.71 - 1.0
Unplanned reoperation rate	98.78	0.94	0.83 - 1.0
Admission to ICU	100	1	1

The data from this dataset was then used to derive and validate a prediction rule for emergency laparotomy within the Cape Metro health district as described in the following chapter.

Chapter 6

Derivation and validation of the prediction rules developed for General Surgery

As was the case with the ACS QI programs, prediction rules for trauma and general surgery were developed and validated separately, following the objective framework and generic principles described in Chapter 3: Methods. This chapter describes the derivation and validation of the Essentials and Procedure-targeted prediction rules using the derivation datasets described in Chapters 4 and 5, respectively.

6.1 Materials and methods

a. Outcomes for prediction

The primary outcome for both general surgery programme prediction rules was a major in-hospital complication, defined according to the following list of ACS-NSQIP occurrences as shown in table 6.1:

Table 6.1 Definition of a major complication

NSQIP occurrence	Major complication
Wound occurrences	Deep incisional SSI
	Organ/ space SSI
	Wound disruption
Respiratory occurrences	Pneumonia
	Pulmonary embolus
Urinary tract occurrences	Progressive renal insufficiency
CNS occurrences	Stroke/ CVA
Cardiac occurrences	Cardiac arrest requiring CPR
	Myocardial infarction
Other occurrences	Septic shock
	Mortality (in-hospital or 30-day)

These occurrences were chosen, as they were all likely to require additional procedures or management, which would change the clinical course of a hospital stay. Furthermore, the above complications were considered to be at minimal risk of classification bias, and had clear definitions provided by the ACS-NSQIP Essentials programme. Major complications were chosen above in-hospital mortality, because with an estimated in-patient mortality of 0.06, only about 28 events could be anticipated. If the 10-1 rule was then adhered to, a maximum of 3 predictors would be included in the models and this would be impractical. However, once a prediction rule had been developed to predict a major in-hospital complication, and the variables were condensed into a single prediction rule, then this could be used in a univariate logistic regression model to predict in-hospital mortality.

The secondary outcome measure included resource utilisation, namely:

- Length of stay (LOS) of greater than 2 weeks
- Post operative ICU admission.

b. Validation dataset

The 2012 sample of the ACS-NSQIP Essentials programme, which consists of 320,816 patients was used for the validation dataset. Over 350 hospitals in the United States of America, contributed to the ACS-NSQIP in 2012. The data was collected by surgical clinical reviewers at these participating hospitals, as described in the literature review, according to a methodology which had been replicated in this study to develop the Essentials and Procedure-targeted programmes. The complete validation dataset was used to validate the prediction rule developed from the Essentials programme. Only

emergency abdominal general surgery operations in the ACS-NSQIP dataset were included in the dataset to validate the Procedure-targeted predictions. This validation dataset included 41,633 general surgery patients who underwent an emergency abdominal operation. Using the ACS-NSQIP dataset for validation assessed both the methodological and geographical transportability of the prediction rules.

c. Identification of predictors

The univariate relationship between variables collected in both derivation datasets and the occurrence of either a major complication, death, LOS> 2 weeks or an admission to ICU, was tested using $\chi 2$ or Fisher's exact, where appropriate. Significance was set at $p<0{\cdot}1$ to err on the side of inclusion.

Missing data

Only variables, which were >80% complete in the derivation datasets were considered for inclusion in the prediction rules. For those variables, which were >80% complete, but still had >5% missing data, multiple imputation methods were used to complete the datasets. Outcome measures, which were not complete, were entered into the regression as missing in both datasets.

d. Building prediction rules

Primary prediction rules were developed in both datasets to predict the primary outcome of a major complication. Secondary prediction rules were also developed in both datasets for the secondary outcome of LOS> 2 weeks. A prediction rule, specifically for ICU admission, was not built in, as a patient who had been identified at high risk for death or major complication by the primary prediction rules, should theoretically be allocated an ICU

admission post-operatively. A validated primary prediction rule was applied to the derivation datasets using the outcome of ICU admission as a method to review current ICU admission practices.

In the LOS models for both derivation datasets, the prescence or absence of a post-operative complication was included as an additional predictor variable, because complications are likely to influence LOS. Furthermore, these models were only dervived from patients who were alive at discharge, as including patients who die in hospital would skew the LOS data.

Preference for all prediction rules was given to variables identified in the efficient model proposed by Chang *et al* (171). Furthermore, age category and ASA class were included, regardless of the significance level for face validity and clinical utility. For those remaining, the significance level for entry into the logistic regression models was $p<0.1$, in order to err on the side of inclusion with model construction. A forward selection algorithm was then used, whereby each variable was screened individually and added into the multivariable logistic regression model in order of statistical significance set at $p<0.05$.

The outcomes data for 373 patients who experienced a total of 69 major complications were included in the Essentials derivation dataset. Thus, applying the 10:1 rule, we considered up to seven candidate predictors for the model predicting the primary outcome in the Essentials dataset. The outcomes data for 320 patients who experienced a total of 44 major complications were included in the Procedure-targeted derivation dataset. Thus, applying the 10:1 rule we considered up to four candidate predictors for the model predicting the primary outcome in the Procedure-targeted dataset.

e. Validating the prediction rules

Comparisons of derivation and validation datasets were tabulated against variables included in the prediction rules, as well as, the unadjusted outcome measures. The primary and secondary prediction rules developed in both derivation datasets were then applied to the appropriate validation dataset. The prediction rules were validated on their discriminative ability (ROC) and ability to calibrate (GOF) for the outcome measures of major complication and in-hospital death for the primary prediction rules, as well as, LOS>2 weeks for the secondary prediction rules. An a priori level of discrimination of 0.7 and calibration of $p>0.05$ has been set in the hypothesis. The univariate association of the primary prediction rule and post operative ICU admission was also assessed in both derivation datasets.

f. Constructing a scoring system

In keeping with the concept of parsimony, the variables included in both primary prediction rules were all dichotomised into binary predictors at intervals established after reviewing the beta-co-efficients of the risk estimates of each category within a variable. Logistic regression analysis was then repeated using only binary predictors and the precision and calibration of the two models were compared.

The beta-co-efficients of the binary predictors were then divided and rounded to the nearest integer to create a scoring system. Operations were grouped into low or high risk of a major complication at logical points. These scores could easily be calculated pre-operatively for appropriate management decisions like allocation to ICU post-operatively. The scoring system was then applied to the outcome measure of in-hospital mortality. The precision and calibaration of the scoring systems in both derivation datasets were further assessed.

g. Power calculation

For the post-hoc assessment of power, to test the measure of association for a single predictor in a model to predict the primary outcome of a major in-hospital complication, which occurs at a rate of 0.15, a threshold odds ratio of 1.5 was chosen. Using the SAS university proc power logistic option and the Shieh-O'Brien approximation, a sample size of 341 would translate to a study powered to 90%, as shown in the SAS university output below (Table 6.2).

Table 6.2 Power Calculation for a general surgery model predicting major in-hospital complication with a single predictor

Method	Shieh-O'Brien approximation
Alpha	0.05
Response Probability	0.15
Test Predictor	ASA
Odds Ratio for Test Predictor	1.5
Unit for Test Pred Odds Ratio	1
Nominal Power	0.9
Total Number of Bins	5

Computed N Total	
Actual Power	**N Total**
0.901	341

All analyses were performed using either SAS University edition or STATA 14.

6.2 Prediction rules for the Essentials Programme at GSH

There were a total of 86 major complications, including 24 deaths, in 69 patients in the Essentials programme. Since the primary outcome was the presence or absence of a major complication, a total of 69 major complications were noted. The occurrence of the primary and secondary outcome measures are shown in table 6.3.

Table 6.3 Occurrence of outcome measures for prediction in the Essentials derivation dataset

Outcome for prediction	Category	N (%)
Major complication	Yes	69 (18.50)
	No	260 (69.71)
	Missing	44 (11.80)
Death	Yes	24 (6.43)
	No	289 (77.48)
	Missing	60 (16.09)
ICU admission	Yes	53 (14.21)
	No	320 (85.79)
	Missing	0 (0)
LOS>14 days	Yes	75 (20.11)
	No	255 (68.36)
	Missing	43 (11.53)

In the Essentials derivation dataset, a post-operative major complication occurred at a rate of 20.97% (95% CI 16.7 – 25.78), death at a rate of 7.67 (95% CI 4.97 – 11.2), ICU admission at a rate of 14.21% (95% CI 110.83 – 18.17) and LOS> 14 days at a rate of 22.73% (95% CI18.32 – 27.63).

a. Identification of Essentials predictors

The univariate association between the categorical variables collected in the Essentials derivation dataset and the endpoints for prediction are presented in table 6.4. Significance was set at p<0.1 and significant associations are highlighted in red.

Table 6.4 The univariate association between categorical predictors and the outcomes for prediction in the derivation dataset

Categorical predictor variable	Major complication p-value	Death p-value	LOS> 14days p-value	ICU post-operatively p-value
Age category	0.431	0.335	0.968	0.852
Gender	0.006	0.086	0.525	0.034
Race	0.53	0.852	0.285	0.144
Emergency status	<0.0001	<0.0001	0.546	0.286
ASA	<0.0001	<0.0001	0.004	<0.0001
Wound classification	<0.0001	<0.0001	0.015	0.076
Surgical specialty	<0.0001	0.058	<0.0001	<0.0001
Surgeon qualification	0.198	0.496	0.242	0.715
Operation complexity coding	0.14	0.251	<0.0001	<0.0001
Diabetic status	<0.0001	0.469	0.075	0.048
Smoking within the year	0.926	0.055	0.159	0.206
Dyspnoea status	0.002	<0.0001	0.335	<0.0001
Functional status	0.008	0.034	0.007	0.423
Ventilator dependent w/in 48hrs	0.487	0.13	0.777	0.232
Charlson's comorbidity index	0.001	0.001	0.003	0.298
COPD* status	0.029	0.001	0.505	0.25
CHF^ status	0.015	0.655	0.024	0.207
HPT^{+} status	0.456	0.275	0.399	0.101
ARF$^{!}$ status	0.01	<0.0001	0.32	0.007
Dialysis status	0.066	0.244	0.011	0.14
Cancer status	0.098	0.03	0.301	0.355
Open wound status	0.13	0.169	0.006	0.387
Steroid status	0.434	0.458	0.52	0.398
>10% weight loss	0.334	0.05	0.535	0.002
Bleeding disorder	0.47	0.756	0.037	0.337
Pre-operative blood transfusion	0.004	0.008	0.078	0.003

Sepsis status	<0.0001	<0.0001	0.024	0.009
HIV status	0.116	0.066	0.341	0.117
Sodium	<0.0001	0.07	0.002	0.189
Urea	<0.0001	0.09	0.088	0.174
Creatinine	0.004	0.0001	0.945	0.512
White Cell Count	0.995	0.38	0.278	0.771
Haematocrit	0.001	0.245	0.275	0.446
Platelets	0.95	0.457	0.002	0.901
Major complication			0.001	
Any complication			0.002	

COPD Chronic Obstructive Pulmonary Disease*
CHF^ Congestive heart Failure
HPT+ Hypertension-
ARF! Acute renal failure
LOS>14§ Length of stay greater than 14 days

Twenty categorical predictors (58.8%) had a significant univariate association to the primary outcome measure of major complication at the 0.1 level. These included gender ($p = 0.006$), emergency status ($p<0.0001$), ASA class ($p<0.0001$), wound classification ($p<0.0001$), surgical specialty ($p<0.0001$), diabetic status ($p<0.0001$), dyspnea status ($p = 0.002$), functional status ($p = 0.008$), Charlson's co-morbidity index ($p = 0.001$), COPD status ($p = 0.029$), CHF status ($p = 0.015$), dialysis status ($p = 0.066$), Cancer status ($p = 0.098$), ARF status ($p = 0.001$), pre-operative blood transfusion ($p = 0.004$), sepsis status ($p<0.0001$), sodium ($p<0.0001$), urea ($p<0.0001$) creatinine ($p = 0.004$) and haematocrit ($p = 0.001$).

There was a trend that those predictors, which were significantly associated with the primary outcome measure, also tended towards a significant association with the remaining outcome measures, including a LOS of greater than 14 days. LOS of greater than 14 days for those patients who were alive at discharge was associated with both a major ($p = 0.001$) and any complication ($p = 0.002$).

b. Building the Essentials prediction rules

Categorical predictors, which were >80% complete were then considered for inclusion in the multivariate models. Thirteen categorical predictors were therefore considered for inclusion in the multivariate model for the primary prediction rule and eighteen predictors for the secondary prediction rule. Multiple imputation methods were then used for any of these variables, which were greater than 5% incomplete. After considering 13 variables individually for inclusion in the primary prediction rule, the final multivariate model with the best performance to predict a major complication in the essentials validation dataset is presented in the following table.

Table 6.5 Multivariate model predicting a major complication following general or vascular surgery at GSH

Predictor (reference)	Odds ratio	95% CI	Co-efficient	95% CI	P-Value
Emergency status (Emergency)	3.05	1.52 -6.18	1.11	0.41 - 1.8	0.002
ASA Score (ASA 1)	1.86	1.28 – 2.93	0.63	0.17 - 1.0	0.006
Age	1.02	0.8 – 1.18	0.002	-0.01 - 0.02	0.802
Wound classification (Clean)	1.52	1.08 – 2.29	0.42	0.04 - 0.81	0.030
Pre-operative blood transfusion > 4 units	4.93	1.22 – 24.92	1.59	0.04 - 3.14	0.044
Constant			-4.31	-5.66 - -2.96	<0.0001

The final five variable model included 4 independent predictors ($p<0.05$), as well as age ($p=0.802$), which was included for face validity. These co-efficients were the log odds for an adverse complication. Therefore, by exponentiating the co-efficients we were able to derive the odds ratios. An emergency case had a 3.05 fold increased odds of a major complication ($p<0.002$). For each increase in ASA category there was a 1.86 fold increased odds of a major complication ($p = 0.006$), and each increase in wound

classification category was associated with a 1.52-increased odds of a major complication (p=0.03). Pre-operative blood transfusion of four or more units was associated with 4.93-increased odds of a major complication (p=0.044). The final model had a ROC of 0.7755 (95% CI 0.70829 – 0.84263) and a GOF statistic of 0.25 (p = 0.6139). Therefore, the prediction rule developed for a major complication following general or vascular surgery at GSH was based on the following formula:

log(P/(1-P)) = -4.310848 + 1.11655*(Emergency status) + 0.639019*(ASA) + 0.0026781*(Age) +0.4208917*(Wound classification) +1.59471*(Preop blood transfusion > 4).

When the prediction rule was used to predict an in-hospital post-operative death, the model had a ROC of 0.8846 (95% CI 0.82807 – 0.94116) and a GOF statistic of 1.72 (p = 0.1896). When used to predict admission to ICU post-operatively however, the ROC was 0.5471 (95% CI 0.45114 – 0.64301) and a GOF statistic of 0.52 (p = 0.4714). After considering 18 variables individually for inclusion in the secondary prediction rule, the final multivariate model with the best performance to predict a LOS>14 in the Essentials validation dataset is presented in the following table.

Table 6.6 Multivariate model predicting a Length of Stay greater than 14 days following general or vascular surgery at GSH

LOS>14 Model	Odds ratio	95% CI	Co-efficient	95% CI	P-Value
ASA score (1)	1.67	1.02 – 2.09	0.51	0.04 - 0.98	0.031
Age	1.03	0.93 – 1.34	-0.01	-0.03 - 0.01	0.205
Wound contamination (Clean)	1.52	1.13 – 2.31	0.42	0.05 - 0.79	0.026
Specialty (Vascular)	0.62	0.47 – 0.57	-0.47	-0.74 - -0.21	<0.0001
Presence of major complication	2.21	1.09 – 4.73	0.79	0.05 - 1.53	0.036
Constant			-1.18	-2.57 - -0.21	0.04

The final five variable model included four independent predictors ($p<0.05$) as well as age ($p=0.205$), which was included for face validity. These co-efficients were the log odds for a LOS>14 days, and therefore by exponentiating the co-efficients we were able to derive the odds ratios. For each increase in ASA category there is a 1.67 fold increased odds of a LOS>14days ($p = 0.031$) and each increase in wound classification category was associated with a 1.52 increased odds of a LOS>14days ($p=0.026$). A general surgery case was associated a 0.62 decreased odds of a LOS>14 days relative to a vascular case ($p<0.0001$). A major complication (death excluded) was associated with a 2.21 increased odds of a LOS>14 days ($p = 0.036$). The final model had a ROC of 0.7573 (95% CI 0.68871 – 0.8258) and a GOF statistic of 2.95 ($p = 0.086$). Therefore the prediction rule developed for a LOS>14 days following general or vascular surgery at GSH was based on the following formula:

log(P/(1-P)) = -1.180884 + 0.5167747*(ASA) -0.0134933*(Age) +0.4206053*(Wound classification) -0.4787236*(General surgery) +0.7947534(Major complication).

c. Validation of the Essentials prediction rules

A comparison of the Essentials derivation and validation datasets was performed on the variables included in the primary and secondary prediction rules. The result of this comparison is presented in Table 6.7.

Table 6.7 Comparison of Essentials Derivation and Validation datasets on variables included in the prediction rules

Predictor variable		GSH Essentials N (%)	ACS Essentials NSQIP N (%)
Total		373	320,830
Age	<65	276 (73.10)	212, 030 (66.09)
	65-74	63 (16.89)	60, 442 (18.84)
	75-84	18 (4.84)	36, 530 (11.39)
	>=85	16 (4.39)	11, 828 (3.69)
	Missing	0 (0)	0 (0)
Sex	Female	159 (42.63)	1,327,462 (57.29)
	Male	214 (57.37)	989.431 (42.71)
Missing		0 (0)	0 (0)
Emergency case	Yes	137 (36.73)	45, 585 (14.21)
	No	236 (62.93)	275, 231 (85.79)
	Missing	0 (0)	14 (0)
ASA Class	1	71 (19.03)	29, 389 (9.16)
	2	125 (33.51)	137, 576 (42.88)
	3	93 (24.93)	129, 638 (40.41)
	4	10 (2.68)	22, 244 (6.93)
	5	0 (0)	867 (0.27)
	Missing	74 (19.84)	1, 116 (0.35)
Wound class	Clean	205 (54.96)	156, 649 (48.82)
	Clean/Contaminated	134 (35.92)	109, 701 (34.19)
	Contaminated	15 (4.02)	31, 334 (9.77)
	Dirty/Infected	19 (5.09)	23, 154 (7.22)
	Missing	0 (0)	0 (0)
Specialty	General surgery	297 (79.62%)	277, 915 (86.63)
	Vascular surgery	76 (20.38%)	42, 900 (13.37)
	Missing	0 (0)	0 (0)
Major complication	Present	69 (18.50)	17, 995 (5.61)
	Absent	260 (69.71)	302, 820 (94.39)
	Missing	69 (18.50)	0 (0)

The ACS-NSQIP Essentials validation cohort had a higher proportion of patients older than 65 (33.9% compared to 26.1%), in ASA categories 4 or 5 (7.2% compared to 2.7%) and with contaminated or dirty wounds (17.0% compared to 9.1%). The GSH

Essentials derivation dataset combined a higher proportion of emergency cases (36.7% compared to 14.2%), vascular cases (20.4% compared to 13.4%) and had a higher rate of major complications (18.5% compared to 5.6%). Table 6.8 presents the comparison of the unadjusted outcomes between the Essentials derivation and validation datasets.

Table 6.8 Comparison of unadjusted outcomes for prediction between the Essentials Derivation and Validation datasets

Unadjusted outcome measure	GSH Essentials Derivation rate (95% CI)	ACS-NSQIP Essentials Validation rate (95% CI)
30-day major complication rate	20.97 (16.7 – 25.78)	5.61 (5.53 - 5.69)
In-hospital mortality rate	5.63 (3.53 - 8.52)	1.08 (1.04 - 1.11)
30-day mortality rate	7.67 (4.97 – 11.2).	1.28 (1.25 - 1.32)
LOS>14 days	22.73 (18.32 – 27.63)	5.58 (5.51 - 5.66)

Figures 6.1 and 6.2 show the area under the ROC curves generated from the ACS-NSQIP validation dataset after applying the co-efficients of the GSH Essentials primary prediction rule to predict in hospital mortality and major complication, respectively.

Figure 6.1 Validation Area under ROC curve of GSH Essentials primary prediction rule applied to the ACS-NSQIP to predict in-hospital death

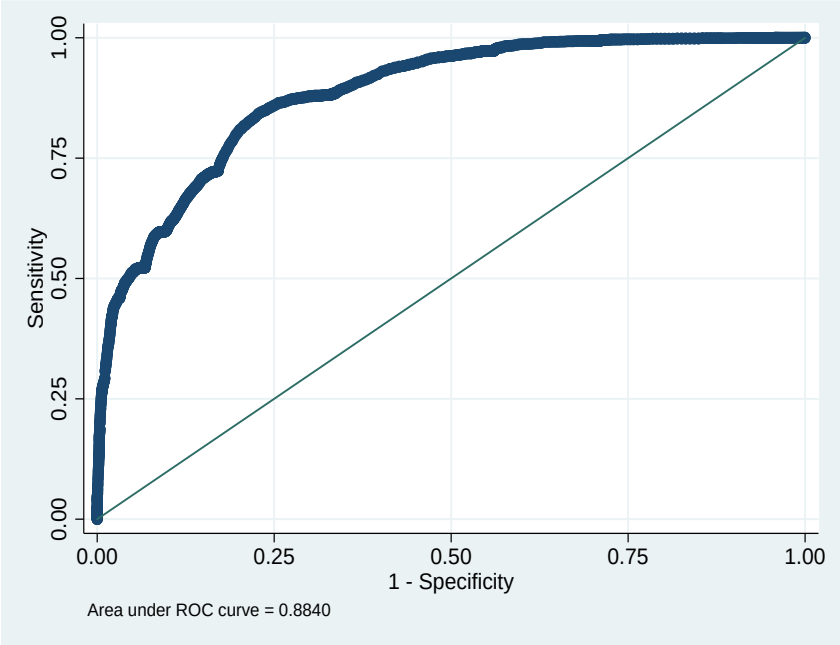
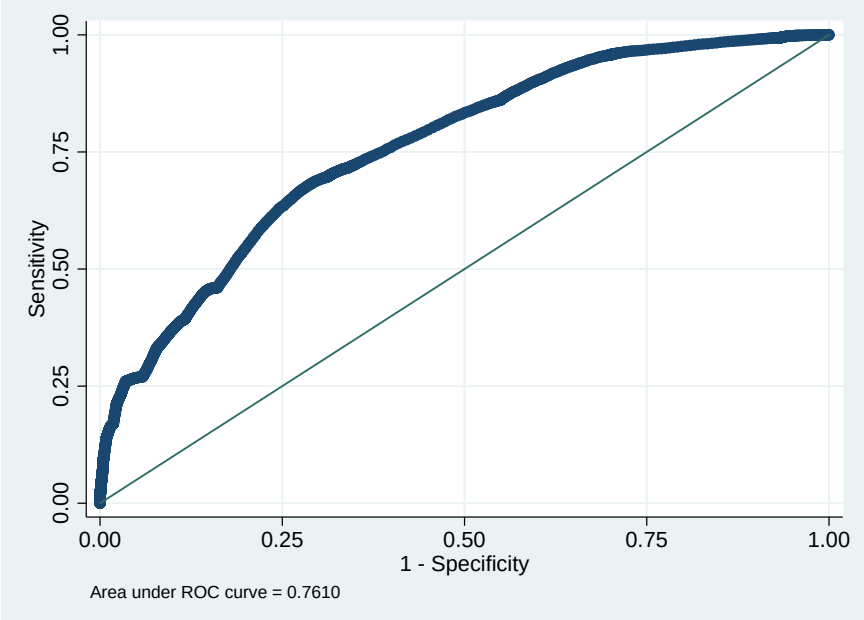


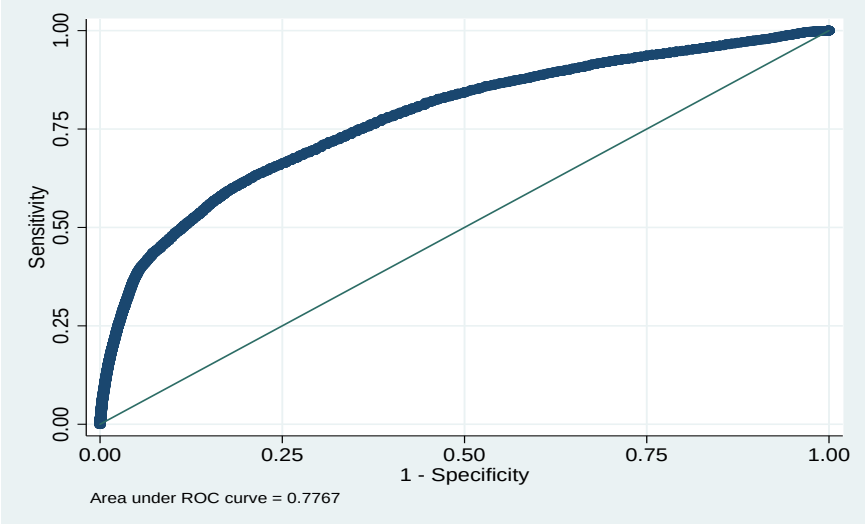
Figure 6.2 Validation Area under ROC curve of GSH Essentials primary prediction rule applied to the ACS-NSQIP to predict a major complication



For the outcome of in-hospital death, the GSH primary prediction rule had an excellent discrimination of 0.884 (95% CI 0.8779 – 0.8901), but the calibration suggested a discrepancy between observed and expected in-hospital deaths, generating a GOF statistic of 240.53 (p <0.001). Similarly, for the outcome of a major complication there was an excellent discrimination of 0.761 (95% CI 0.757 – 0.764), but the calibration suggested a discrepancy between observed and expected major complications with a GOF statistic of 943.31 (p<0.0001).

Figure 6.3 demonstrates the area under the ROC curve generated from the ACS-NSQIP validation dataset after applying the co-efficients of the GSH Essentials secondary prediction rule to predict a length of stay of greater than 14 days.

Figure 6.3 Validation Area under ROC curve of GSH Essentials secondary prediction rule applied to the ACS NSQIP to predict a length of stay greater than 14 days



The secondary prediction rule applied to the ACS NSQIP had a ROC of 0.7038 (95% CI 0.7019 – 0.7058) and GOF of 299.37 (p<0.0001), again, suggesting a

discrepancy between observed and expected despite the good discrimination of the model. Table 6.9 summarises the performance of the primary and secondary prediction rules in both the derivation and validation datasets.

Table 6.9 Summary performance statistics of primary and secondary Essentials prediction rules in the derivation and validation datasets

Outcome	Cohort	Discrimination	95% CI	Calibration statistic	P-Value
Major complication	Derivation	0.775	0.708 – 0.842	0.25	0.613
	Validation	0.761	0.757 – 0.764	943.31	<0.001
In hospital death	Derivation	0.884	0.828 – 0.941	1.72	0.189
	Validation	0.884	0.87795 – 0.890	240.53	<0.001
LOS>14	Derivation	0.757	0.68871 – 0.821	2.95	0.085
	Validation	0.703	0.7019 – 0.705	299.37	<0.001
ICU admission	Derivation	0.547	0.45114 – 0.643	0.52	0.4714

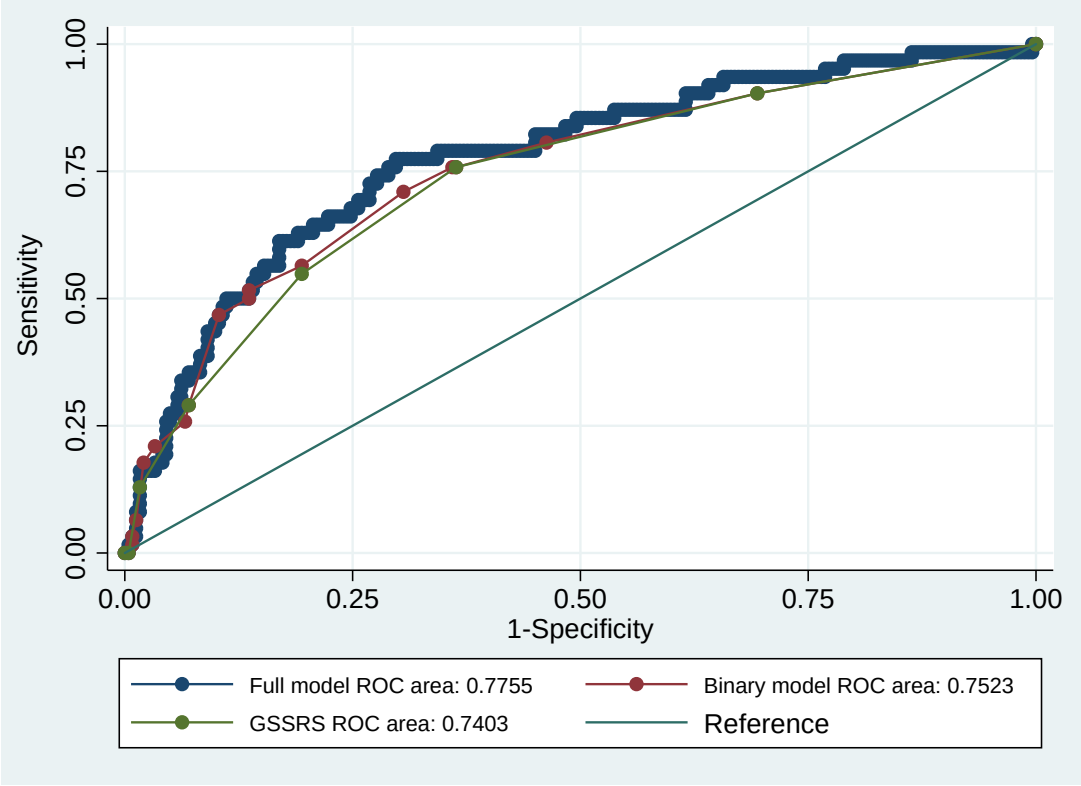
d. An Essentials scoring system

To develop a scoring system, the variables included in the validated Essential's primary prediction rules were all dichotomised into binary predictors at intervals established after reviewing the beta-co-efficients of the risk estimates of each category within a variable. Logistic regression analysis was then repeated using only binary predictors for the outcome of a major complication. The beta-co-efficients of the binary predictors were then divided and shrunk down to the nearest integer to create a scoring system from 0 to 7 called the Groote Schuur Surgery Risk Score (GSSRS) which was generated by the following formula:

GSSRS= 1*(Wound class> clean-contaminated)+1*(ASA>2)+1*(Age>60 years)+2*(emergency status)+2*(Pre-operative blood transfusion>4 packed RBC's).

The comparative discriminatory ability of the full, binary and GSSRS models are presented in figure 6.4. The areas under each ROC curve were no different between the models (p=0.567).

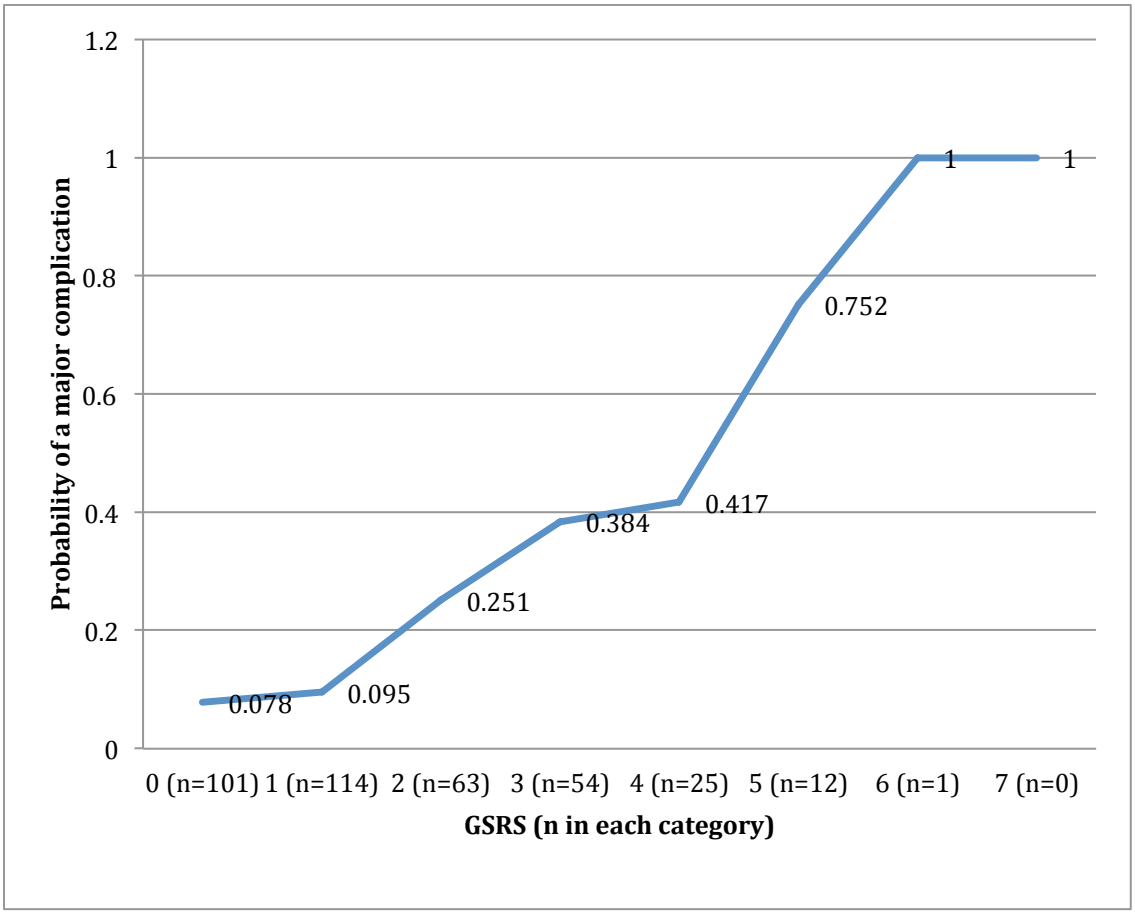
Figure 6.4 Test of discrimination for 3 models predicting a major complication following general or vascular surgery at GSH



p=0.567.

The median GSSRS in the derivation dataset was 1 (IQR 0-3). A GSRS Score of >3 had a sensitivity of 76.82% and specificity of 75.0% to detect a major complication in the Essentials derivation dataset and an increase in GSSRS was associated with a stepwise increase in the occurrence of a major complication as shown in figure 6.5.

Figure 6.5 Probability of a major complication by Groote Schuur Surgery Risk Score (GSSRS)



e. Risk-adjusted benchmarking analysis for the Essentials Programme

Finally, using the GSH prediction rules O/E ratios for major complications and LOS>14 days were generated in order to perform a risk-adjusted benchmarking analysis of GSH and the ACS-NSQIP. The result of this analysis is presented in table 6.10.

Table 6.10 Observed versus expected outcomes in the ACS-NSQIP validation dataset according to the GSH Essentials prediction rules

Outcome	Observed	Expected	O/E	95% CI
Major complication	21,050	48, 928	0.43	0.42-0.44
LOS>14	17,911	149, 233	0.12	0.11-0.12

A patient undergoing a general surgery or vascular operation at GSH is twice as likely to experience a major complication or ten times as likely to spend longer than 2 weeks in hospital if operated on at GSH compared to an average performing hospital in the ACS-NSQIP consortium.

6.3 Prediction rules for a Procedure-targeted Programme in the Cape Metro West

In the Procedure-targeted dataset a major post-operative major complication occurred at a rate of 14.15% (95% CI 10.47 – 18.52), death at a rate of 6.56% (95% CI 4.11 – 9.86), ICU admission at a rate of 12.69% (95% CI 9.23 – 16.89) and LOS> 14 days at a rate of 8.83 % (95% CI 5.95 – 12.52). For GSH hospital, a major post-operative complication occurred at a rate of 27.36% (95% CI 19.61 – 36.76), death at a rate of 13.76% (95% CI 8.4 – 21.7), ICU admission at a rate of 24.3% (95% CI 17.1-33.5) and LOS> 14 days at a rate of 9.34% (95% CI 5.05 – 16.65). The occurrence of these outcome measures is presented in table 6.11. All unadjusted endpoints occurred more frequently at GSH ($p<0.05$).

Table 6.11 Occurrence of outcome measures for prediction in the procedure-targeted dataset

Outcomes for prediction	Category	N (%)
Major complication	Yes	44 (13.75)
	No	267 (83.44)
	Missing	9 (2.81)
Death	Yes	21 (6.56)
	No	299 (93.44)
	Missing	0 (0)
ICU admission	Yes	40 (12.50)
	No	275 (85.94)
	Missing	5 (1.56)
LOS>14 days	Yes	28 (8.75)
	No	289 (90.31)
	Missing	3 (0.94)

a. Identification of Procedure-targeted predictors

The univariate association between the categorical variables collected in the procedure-targeted dataset and the endpoints for prediction are presented in table 6.12. Significance was set at $p<0.1$ and significant associations are highlighted in red.

Table 6.12 The univariate association between categorical predictors and the outcomes for prediction in the Procedure-targeted derivation dataset

Categorical predictor	Major complication	Death	Any complication	LOS> 14	ICU post-operatively
Age category	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Gender	0.521	0.651	0.451	0.037	0.415
Race	0.338	0.336	0.377	0.684	0.249
Diabetic	0.01	0.028	0.037	0.034	0.078

Smoking within the year	0.296	0.22	0.054	0.316	0.096
Functional status	0.018	0.007	0.012	0.035	0.273
Sepsis status	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Charlson's comorbidity index	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HIV Status	0.62	0.863	0.581	0.76	0.051
BMI	0.5	0.37	0.471	0.368	0.49
Ur	0.001	<0.0001	0.003	0.493	0.015
Creatinine	0.002	<0.0001	0.002	0.173	0.006
Albumin	0.264	0.381	0.132	0.415	0.545
Haematocrit	0.315	0.176	0.178	0.189	0.359
INR	0.242	0.364	0.403	0.511	0.465
Ph	0.203	0.051	0.19	0.705	0.74
Bicarb	0.072	0.064	0.175	0.435	0.675
Base excess	0.366	0.264	0.441	0.274	0.223
lactate	0.042	0.043	0.406	0.8	0.002
ASA	<0.0001	<0.0001	<0.0001	0.006	<0.0001
Speciality	<0.0001	0.0001	<0.0001	0.003	<0.0001
Surgeon qualification	<0.0001	0.016	0.001	0.003	<0.0001
Anaesthetic qualification	0.001	0.065	<0.0001	0.079	<0.0001
Wound classification	0.249	0.253	0.124	0.93	0.565
Perforation of hollow viscous	0.081	0.385	0.021	0.264	<0.0001
Operative complexity	<0.0001	<0.0001	<0.0001	0.07	<0.0001
Incision type	<0.0001	0.003	0.001	0.131	<0.0001
Bowel resection performed	<0.0001	0.024	<0.0001	<0.0001	<0.0001
Stoma sited	<0.0001	0.064	0.003	<0.0001	<0.0001
Pre-operative CT performed	<0.0001	0.034	0.001	0.001	<0.0001
Pre-operative blood transfusion	0.002	0.001	0.005	0.049	<0.0001
WHO Checklist completion rate	0.508	0.2	0.42	0.472	0.865
DVT Prophylaxis documented	0.004	0.584	0.005	0.242	0.002
Antibiotic prophylaxis documented	0.436	0.226	0.448	0.417	0.532
Major complication				<0.0001	
Any complication				<0.0001	

In the Procedure-targeted dataset a greater proportion of variables collected were associated with the primary outcome measure (64.71%), compared to the Essentials derivation dataset (58.82%). These twenty one predictors included age category

(p<0.0001), ASA class (p<0.0001), surgical specialty (p<0.0001), surgeon qualification (p<0.0001), anaesthetic qualification (p<0.001), perforation of hollow viscous (p=0.081), operative complexity (p<0.0001), incision type (p<0.0001), whether or not a resection was performed (p<0.0001), whether or not a stoma was sited (p<0.0001), whether or not a pre-operative CT scan was performed (p<0.0001), pre-operative blood transfusion (p=0.002), documentation of DVT prophylaxis (p=0.004), diabetic status (p = 0.01), functional status (p = 0.018), pre-operative sepsis (p<0.0001), Charlson's co-morbidity index (p<0.0001), urea (p = 0.001), creatinine (p = 0.002), bicarbonate (p=0.072) and lactate (p=0.042). As seen in the Essentials dataset, there was a trend for those variables identified as significant for the primary outcome in the univariate screen to also be significant for the secondary outcome i.e. LOS of greater than 14 days for those patients who were alive at discharge was associated with both the presence of a major (p <0.0001) and any complication (p <0.0001).

b. Building the Procedure-targeted prediction rules

All categorical predictors identified as significant in the univariate screen, which were more than 80% complete were then considered for inclusion in the multivariate models. Nineteen categorical predictors were considered for inclusion in the multivariate model for the primary prediction rule and fourteen predictors for the secondary prediction rule. Multiple imputation methods were used for any of these variables, which were more than 5% incomplete. After considering nineteen variables individually for inclusion in the primary prediction rule, the final multivariate model with the best performance to predict a major complication in the procedure-targeted derivation dataset is presented in Table 6.13.

Table 6.13 Multivariate model predicting a major complication following an exploratory laparotomy in the Cape Metro

Major complication Model (reference level)	Odds ratio	95% CI	Co-efficient	95% CI	P-Value
Age	1.05	1.01 – 1.67	0.05	0.02 - 0.07	<0.0001
ASA Score (1)	1.39	1.06 – 2.05	0.31	0.09 - 0.52	0.005
Pre-operative sepsis (None)	1.92	1.34 – 2.67	0.65	0.27 - 1.03	0.001
Urea	6.06	4.56 – 7.89	1.80	0.93 – 2.66	<0.0001
Constant			-6.54	-8.29 - -4.79	<0.0001

The final four variable model included four independent predictors of a major complication. These co-efficients were the log odds for a major complication, and by exponentiating the co-efficients we were able to derive the odds ratios. For every one year increase in age there was a 1.052-fold increased odds of a major complication ($p<0.001$), and each increase in ASA category was associated with a 1.395-fold increased odds of a major complication ($p=0.005$). Each increase in pre-operative sepsis class was associated with 1.918-fold increased odds of a major complication ($p=0.001$). A pre-operative blood urea level of greater than 7.1 mmol/l was associated with 6.06-fold increased odds of a major complication ($p<0.001$). All the included variables were best fitted by assuming a linear association with the primary outcome. The final model had a ROC of 0.867 (95% CI 0.8152 – 0.9183) and a GOF statistic 3.38 ($p = 0.066$). Therefore the prediction rule developed for a major complication following an emergency laparotomy in the Western Cape Metro was based on the following formula:

log(P/(1-P)) = -6.544383 + 0.0511935*(Age) + 0.3073609*(ASA) + 0.6515268*(Pre-operative sepsis status) + 1.801279*(Preop Urea >7.1).

When the prediction rule was used to predict an in-hospital post-operative death, the model had a ROC of 0.8869 (95% CI 0.83453 – 0.93925) and a GOF statistic of 1.91 (p = 0.1668). When used to predict admission to ICU post-operatively, the ROC was 0.8755 (95% CI 0.82791 – 0.92319), but the GOF was 4.82 (p = 0.0281).

After considering fourteen variables individually for inclusion in the secondary prediction rule, the final multivariate model with the best performance to predict a LOS of greater than 14 days in the procedure-targeted derivation dataset was presented in Table 6.14.

Table 6.14 Multivariate model predicting a Length of Stay greater than 14 days following an emergency laparotomy in the Cape Metro

LOS> 14 Model (reference level)	**Odds ratio**	**95% CI**	**Co-efficient**	**95% CI**	**P-Value**
Age	1.04	0.75 – 1.55	0.002	-0.031 - 0.037	0.869
Pre-operative sepsis (no sepsis)	2.20	1.45 – 3.54	0.78	0.29 - 1.28	0.002
Any complication	16.03	12.06 – 35.78	2.77	1.53 - 4.01	<0.0001
Functional status (independent)	1.75	1.06 – 3.45	0.55	0.04 - 1.07	0.033
Constant			-6.06	-8.23 - -3.89	<0.0001

The final four variable model included three independent predictors as well as age (p=0.869), which was included for face validity. These co-efficients were the log odds for a LOS of greater than 14 days (LOS>14), and by exponentiating the co-efficients we were able to derive the odds ratios. For each increase in pre-operative sepsis class, there was a 2.2-fold increased odds of a LOS>14 (p = 0.002), and every increase in functional status class was associated with a 1.75-fold increased odds of a LOS>14 (p=0.026). The

occurrence of any post-operative complication was associated with 16.03-fold increased odds of a LOS>14 (p<0.001). The final LOS model had a ROC of 0.876 (95% CI 0.778 – 0.975) and a GOF of 0.661 (p = 0.416). Therefore, the prediction rule developed for a LOS greater than 14 days following an emergency exploratory laparotomy in the Cape Metro was based on the following formula:

log(P/(1-P)) = -6.064299+ 0.002936*(Age) +0.7892301*(Pre-operative sepsis class) +0.5592842*(Functional status class) +2.774222*(Presence of any complication).

c. Validation of the Procedure-targeted prediction rules

A comparison of the Procedure-targeted derivation and validation datasets was performed on the variables included in the primary and secondary prediction rules. The results of this comparison are presented in Table 6.15.

Table 6.15 Comparison of Procedure-targeted Derivation and Validation datasets by variables included in the prediction rules

Variable	Category	Cape Metro Procedure-targeted N (%)	ACS NSQIP Validation N (%)
Total patients		320	41, 633
Age	<18	43 (13.44)	0 (0)
	18-29	78 (24.38)	8, 352 (20.06)
	30-39	65 (20.31)	6, 464 (15.53)
	40-49	54 (16.88)	6, 038 (14.50)
	50-59	37 (11.56)	6, 885 (16.54)
	>59	43 (13.44)	13, 894 (33.37)
	Missing	0 (0)	0 (0)
ASA Score	1	165 (51.56)	8, 306 (19.95)
	2	72 (22.50)	17, 123 (41.13)
	3	29 (9.06)	10, 867 (26.10)
	4	18 (5.62)	4, 778 (11.48)
	5	4 (1.25)	525 (1.26)
	Missing	32 (10.0)	34 (0.08)
Pre-operative sepsis	None	17 (16.16)	26, 843 (64.48)
	SIRS	70 (64.65)	7, 370 (17.70)

	Sepsis	15 (14.40)	5, 736 (13.78)
	Septic shock	7 (4.79)	1, 684 (4.04)
	Missing	0 (0)	0 (0)
Pre-operative urea	Abnormal	73 (22.81)	8, 059 (19.36)
	Normal	185 (57.81)	33, 574 (80.64)
	Missing	62 (19.37)	0 (0)
Functional status	Independent	274 (85.62)	39, 601 (95.12)
	Partially dependent	10 (3.12)	1, 342 (3.22)
	Dependent	1 (0.31)	457 (1.10)
	Missing	33 (10.31)	233 (0.56)
Any complication	Present	59 (18.44)	7, 417 (17.82)
	Absent	249 (77.81)	34, 216 (82.18)
	Missing	12 (3.75)	0 (0)

The ACS-NSQIP validation dataset had a higher proportion of patients in the greater than 59 years (33.37% compared to 13.44%) and a higher proportion of patients in ASA Class 4 or 5 (12.74% compared to 6.87%). A higher proportion of patients were assessed with any grade of sepsis in the derivation dataset compared to the validation dataset (83.84% compared to 35.52%). Pre-operative urea, functional status and rates of any complication were relatively comparable. Rates of missing data were still higher in the derivation dataset. Table 6.16 presents the comparison of the unadjusted outcomes between the Essentials derivation and validation datasets.

Table 6.16 Comparison of unadjusted outcomes for prediction between the Procedure-targeted Derivation and Validation dataset

Unadjusted outcome measure	**Cape Metro Derivation rate (95% CI)**	**ACS-NSQIP Procedure-targeted rate (95% CI)**
Major complication rate	14.15 (10.47 – 18.52)	19.74 (19.36 – 20.13)
In-hospital mortality rate	6.56 (4.12 – 9.86)	3.78 (3.61 – 3.97)
LOS>14 days	8.83 (5.95 – 12.51)	10.47 (10.17 – 10.76)

Figures 6.6 and 6.7 show the area under the ROC curves generated from an emergency general surgery subset in the ACS-NSQIP validation dataset, after applying

the co-efficients of the Cape Metro procedure-targeted primary prediction rule to predict in- hospital mortality and major complication, respectively.

Figure 6.6 Validation Area under ROC curve of the Cape Metro Procedure-targeted primary prediction rule applied to the ACS-NSQIP to predict in-hospital death

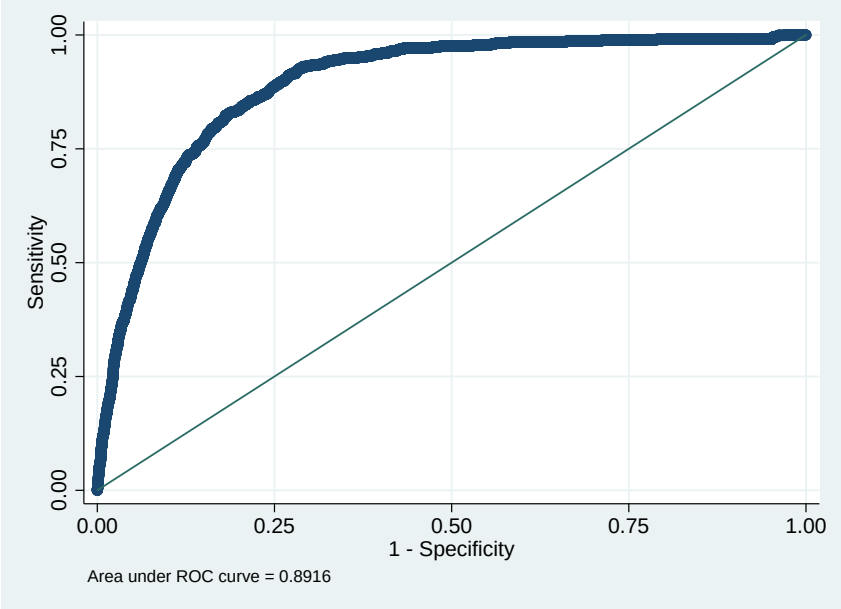
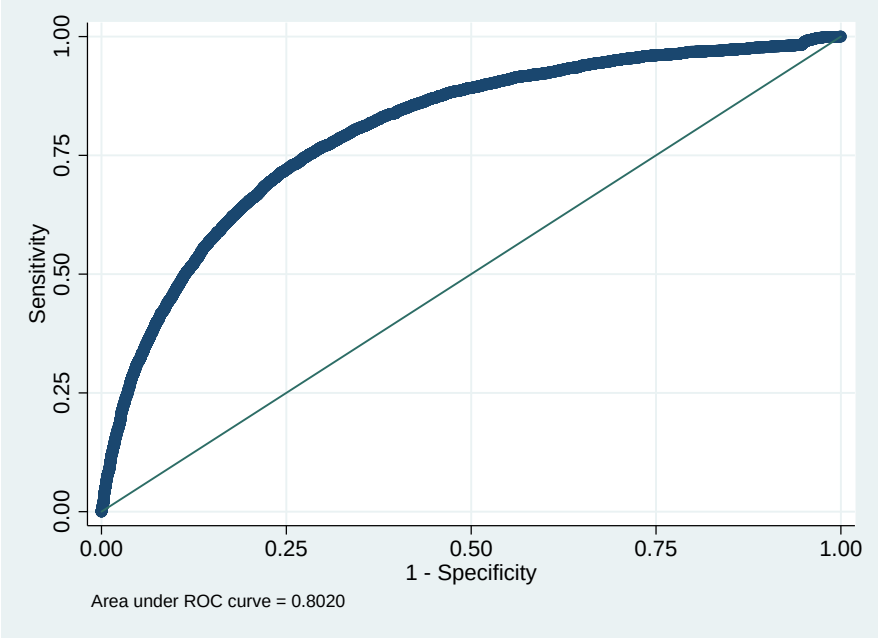
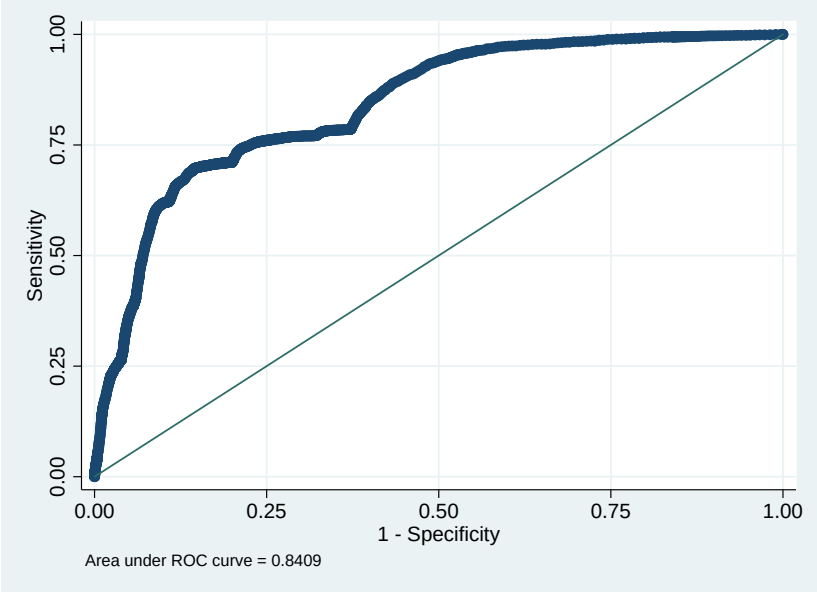


Figure 6.7 Validation Area under ROC curve of Cape Metro Procedure-targeted prediction rule applied to the ACS-NSQIP to predict a major complication



For the outcome of in hospital death, the discrimination of the Cape Metro West primary prediction was excellent (0.892; 95% CI 0.886 – 0.903), but calibration suggested a discrepancy between observed and expected in hospital deaths, generating a GOF statistic of 25.09 ($p < 0.001$). Similarly, for the outcome of a major complication, the discrimination was excellent (0.820; 95% CI 0.795 – 0.808), but the calibration suggested a discrepancy between observed and expected major complications, generating a GOF statistic 189.76 ($p<0.0001$). Figure 6.8 demonstrates the area under the ROC curve generated from the emergency general surgery subset of the ACS-NSQIP validation dataset, after applying the co-efficients of the Cape Metro Procedure-targeted secondary prediction rule to predict a length of stay of greater than 14 days.

Figure 6.8 Validation Area under ROC curve of The Cape Metro Procedure-targeted secondary prediction rule applied to the ACS-NSQIP to predict a length of stay greater than 14 days



The secondary prediction rule applied to the emergency general surgery ACS-NSQIP subset had a ROC of 0.841 (95% CI 0.834 – 0.847) and GOF of 318.68 (p<0.001) again suggesting a discrepancy between observed and expected despite the good discrimination of the model. Table 6.17 summarises the performance of the Cape Metro primary and secondary prediction rules in both the derivation and validation datasets.

Table 6.17 Summary performance statistics of primary and secondary Cape Metro Procedure-targeted prediction rules in the derivation and validation datasets

Outcome	Cohort	Discrimination	95% CI	Calibration statistic	P-Value
Major complication	Derivation	0.866	0.815 - 0.918	3.38	0.066
	Validation	0.820	0.795 - 0.808	189.76	<0.001
In hospital death	Derivation	0.886	0.834 - 0.939	1.91	0.166
	Validation	0.891	0.886 - 0.903	25.09	<0.001
LOS>14*	Derivation	0.876	0.778 - 0.975	0.66	0.416
	Validation	0.840	0.834 – 0.841	318.68	<0.001
ICU admission	Derivation	0.875	0.827 - 0.923	4.82	0.028

LOS>14 Length of stay greater than 14 days*

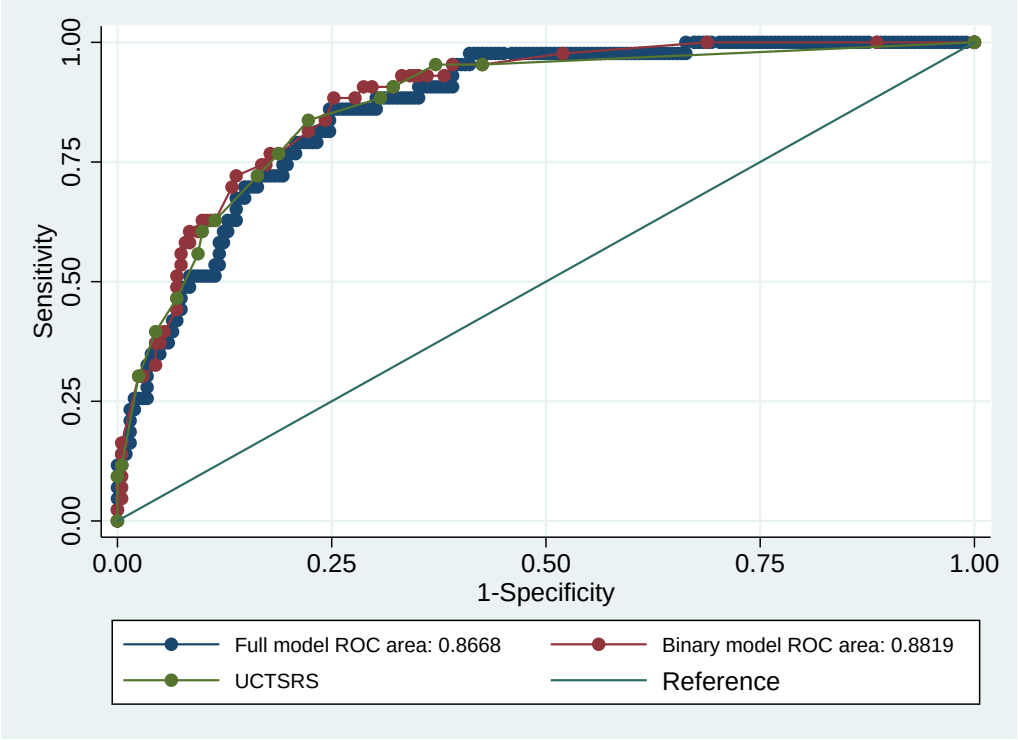
d. A Procedure-targeted scoring system

Similar to the final step in the Essentials programme, in order to develop a scoring system, the variables included in the validated Cape Metro primary prediction rules were all dichotomised into binary predictors at intervals established after reviewing the beta-co-efficients of the risk estimates of each category within a variable. Logistic regression analysis was then repeated using only binary predictors for the outcome of a major complication. The beta-co-efficients of the binary predictors were then divided and shrunk down to the nearest integer to create a scoring system from 0 to 7 and was called University of Cape Town Surgical Risk Score (UCTSRS), which was generated by the following formula:

UCTSRS= 3*(Sepsis or septic shock) + 2 (ASA >2) + 1 * (Age>60) + 1* (Urea>7.1mmol/l)

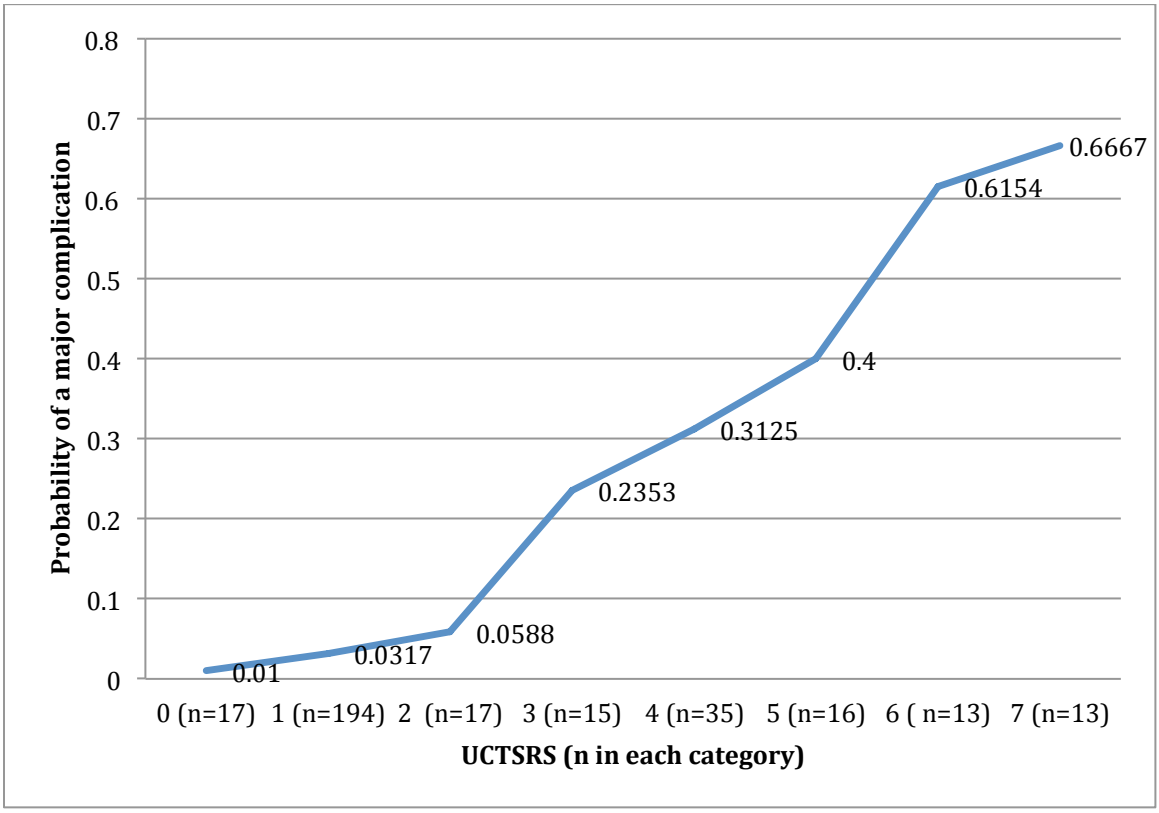
The comparative discriminatory ability of the full, binary and UCTSRS models are presented in figure 6.9. The areas under each ROC curve were no different between the models (p=0.329).

Figure 6.9 Test of discrimination for 3 models predicting a major complication following an emergency laparotomy in the Cape Metro



The median UCTSRS in the derivation dataset was 1 (IQR 1-3). A UCTSRS Score of >2 had a sensitivity of 82.77% and specificity of 65.91 % to detect a major complication in the Procedure-targeted derivation. An increase in UCTSRS was associated with a stepwise increase in a major complication, as shown in figure 6.10.

Figure 6.10 Probability of a major complication as predicted by University of Cape Town Surgical Risk Score (UCTSRS)



e. Risk-adjusted benchmarking analysis for the Procedure-targeted Programme

Using the Cape Metro prediction rules, the O/E ratios for in-hospital deaths, and patients who spent 14 days in hospital, were generated from the emergency general surgery validation ACS-NSQIP dataset. These O/E ratios are presented in Table 6.18.

Table 6.18 Observed versus expected outcomes in the ACS-NSQIP validation dataset according to the Procedure-targeted prediction rules

Outcome	Observed	Expected	O/E	95% CI
Major complication	8,219	9106	0.9026	0.8832-0.9223
LOS>14	3,510	4357	0.8056	0.7792-0.8327

A patient undergoing an emergency exploratory laparotomy at a hospital in the Cape Metro West district of South Africa was 10% more likely to experience a major complication, or 20% more likely to spend longer than 2 weeks in hospital, compared to an average performing hospital in the ACS-NSQIP consortium.

Chapter 7

A novel approach to global benchmarking of risk adjusted surgical outcomes:

The universal risk calculator

The Essentials derivation dataset required that the GSH departments of General Surgery and Quality Assurance employ a full-time clinical reviewer. This may not be a sustainable solution in the long term, nor is it a realistic prospect for other surgical units operating in the Cape Metro or other LMIC settings. Hence, the current lack of surgical outcomes data in the developing world. The possibility of reliable external risk-adjusted benchmarking of GSH general and vascular surgery departments, with other centres in similar settings is therefore not currently realistic.

However, following the acquisition of very large surgical outcomes databases by the ACS, investigators have been able to use prediction rules developed by multiple logistic regression modeling. These rules identify independent predictors, as well as calculate individualised probabilities of an adverse event following major surgery. In addition, risk calculators based on these models have been created, and are predominantly utilised today as a decision aide and an informed consent tool within the context of, "informed informed consent."

However, beyond patient-level decision making, the author proposes that the availability of these risk calculators, now makes global benchmarking of surgical outcomes feasible. They offer an alternative access to the logistic regression model to

institutions, which do not have direct access to the data upon which a risk calculator is based. In this manner, these calculators can provide the 'E', which is necessary for the O/E ratio to perform risk-adjusted benchmarking of surgical outcomes. These calculators are therefore, not only useful for patient-level decision making, but can also have important implications for hospital, regional or even global-level quality benchmarking. This concept was piloted in this study, by using the data collected for the Essentials derivation dataset. The Universal risk calculator developed by Bilimoria *et al,* (available on-line at *www.riskcalculator.facs*) uses 21 pre-operative factors that are all included in the ACS Essentials QI programme, and therefore, in the GSH Essentials derivation dataset (Table 7.1).

Table 7.1 Pre-operative and intraoperative variables required by the ACS Universal calculator

Patient demographics
Age
Sex
Height
Weight
Procedure related
Name of procedure (converted to CPT code by the risk calculator)
Emergency case (Yes/ No)
ASA Class (1-5)
Wound class (Clean/ Clean-contaminated/ Contaminated/ Dirty-infected)
Pre-operative risk assessment
Steroid use for chronic condition (Yes/ No)
Ascites within 30 days prior to surgery (Yes/ No)
Systemic sepsis within 48 hours prior to surgery (None/ SIRS/ Sepsis/ Septic shock)
Ventilator dependent (Yes/ No)
Disseminated cancer (Yes/ No)
Diabetes (None/ Oral medication/ Insulin medication)
Hypertension requiring medication (Yes/ No)

Previous cardiac event (Yes/ No)
Congestive heart failure in 30 days prior to surgery (Yes/ No)
Dyspnea (None/ With moderate exertion/ At rest)
Current smoker within 1 year (Yes/ No)
History of severe COPD (Yes/ No)
Dialysis (Yes/ No)
Acute renal failure (Yes/ No)

Adapted from ACS universal calculator available at www.riskcalculator.facs.org

7.1 Materials and Methods

After individually inputting the data from each patient into the online calculator, the built-in regression models in the calculator were able to predict an individualised probability for each of the ACS-NSQIP outcome measures included in the Essentials derivation dataset as well as LOS. This step was performed for every patient included in the Essentials derivation dataset. Missing data was left unchecked in the calculator. The individual patient probabilities, calculated by the calculator, were then aggregated to obtain the expected number of adverse events (E) for the GSH Essentials derivation cohort. After calculating the E for the cohort for a specific outcome, this was compared to what was actually observed during the study period (O), and O/E ratios for each outcome of interest were generated. Confidence intervals (CI) were then calculated using a Poisson CI calculator.

The LOS data was right-skewed and non-parametric. The expected and observed total LOS was compared using the Kruskal Wallis Rank test, using only LOS data for patients who were alive at the time of discharge. This data was also converted to time to event data, and Kaplan Meier curves comparing estimated and observed LOS were then plotted.

a. Interpretation

If the O/E ratio of the cohort for any outcome measure was above 1, and the lower bound of the confidence interval was>1, then the number of adverse events at GSH were significantly larger than would be expected on the basis of its patient characteristics, benchmarked against the ACS-NSQIP consortium. If the O/E ratio of the cohort for any outcome measure was below 1 and the upper bound of the confidence interval was<1, then the surgical unit had experienced a significantly smaller number of adverse events than would have been expected on the basis of its patient characteristics, benchmarked against the NSQIP consortium. A key goal of any QI initiative is to provide readily interpretable feedback to the participating institution to identify targets for improvement. Resultant O/E ratios for each complication and 95% confidence intervals were plotted on caterpillar plots.

7.2 Results

Data describing the 373 patients who made up the Essentials derivation dataset, which were required by the ACS Universal calculator were then individually inserted into the on-line risk calculator. Individual probabilities of each outcome of interest were summed and compared to the observed outcomes. The result of this analysis is presented in Table 7.2.

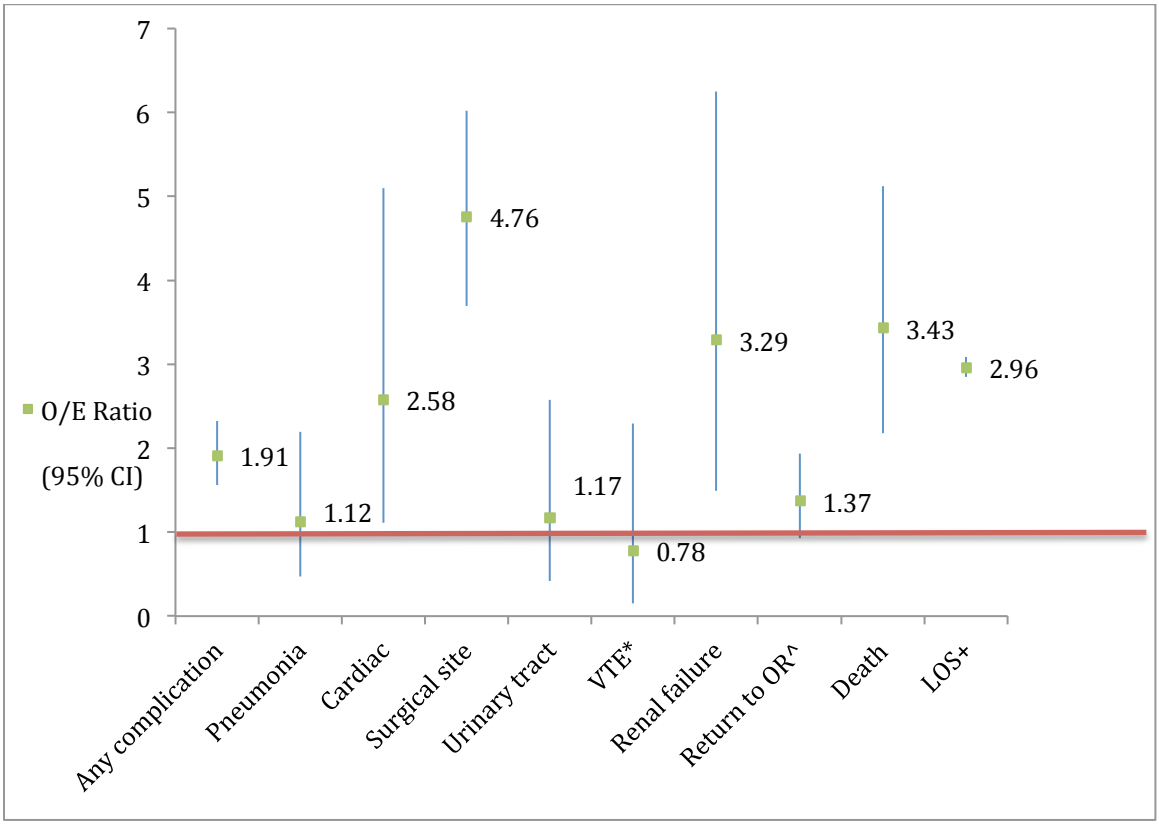
Table 7.2 O/E Ratios for the GSH Essentials derivation dataset benchmarked against the ACS NSQIP consortium

Outcome	Observed	Expected*	O/E Ratio	95% CI
Any complication	111	58	1.91	1.57 - 2.31
Pneumonia	8	7	1.12	0.41 - 2.25
Cardiac complication	8	3	2.58	1.15 - 5.24
Surgical site infection	70	14	4.76	3.64 - 5.89
Urinary tract infection	6	5	1.17	0.44 - 2.61
Venous thromboembolism	3	4	0.78	0.15 - 2.19
Renal failure	9	2	3.29	2.06 - 8.54
Return to OR	33	24	1.38	0.95 - 1.93
Death	24	7	3.43	2.19 - 5.11

**Rounded to the nearest integer*

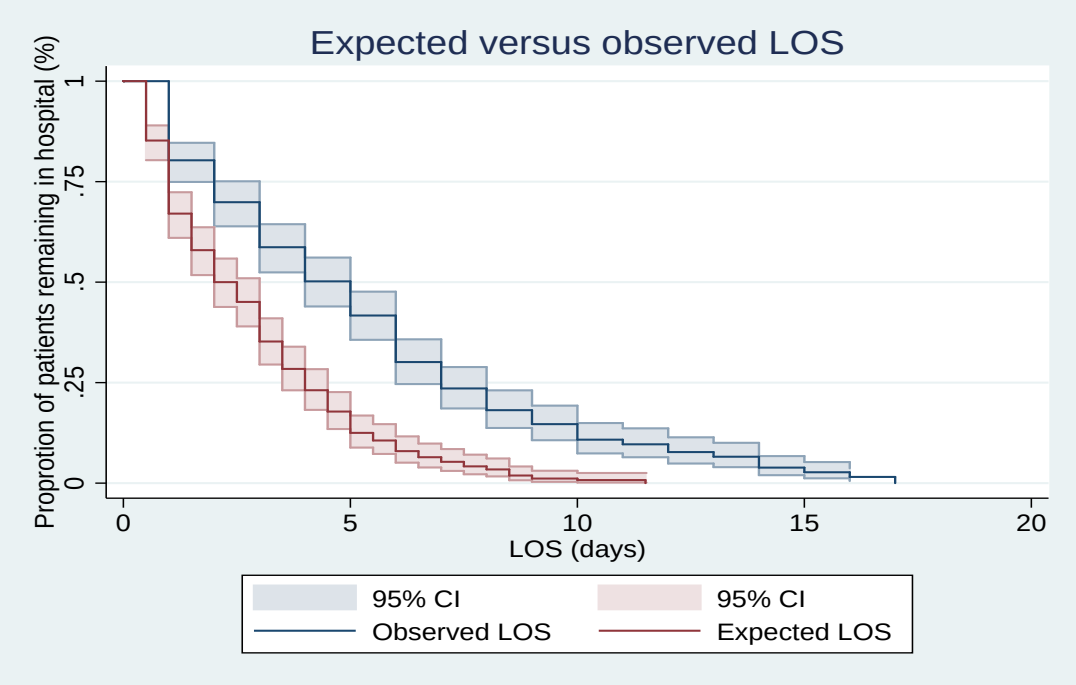
O/E ratios were greater than one for every complication audited except for venous thromboembolism, which had an O/E ratio of 0.78 (95% CI 0.15 – 2.19). Ratios that reached statistical significance included cardiac complications (O/E= 2.58; 95% CI 1.15 – 5.24), surgical site infections (O/E = 4.76; 3.64 - 5.89), renal failure (O/E = 3.29; 95 % CI 2.06 – 8.54) and death (O/E=3.43; 95% CI 2.19 – 5.11). The overall complication O/E was 1.91 (95% CI 1.57 – 2.31). O/E ratios are summarised in the caterpillar plot presented in Figure 7.1.

Figure 7.1 Caterpillar plot benchmarking Groote Schuur against the ACS-NSQIP consortium



The median (IQR) for expected and observed LOS was 3 (1,5) and 7 (3,12), respectively. According to the risk calculator, the cohort would have been expected to spend a total of 1,098 days in hospital and the observed total LOS was 3,253 days. The O/E for LOS was 2.96 (95% CI 2.86 – 3.07). The Kruskal-Wallis rank sum for the observed LOS was 144869.50 and for the expected was 89400.50 ($p<0.001$). Kaplan Meier curves with 95% CIs comparing time to discharge for patients without complications are presented in figure 7.2.

Figure 7.2 Observed versus expected length of stay (LOS) in patients without complications



Model validation

The performance of models predicting 30-day mortality, 30-day morbidity and LOS>30 days using the variables included in the risk calculator were excellent. The ROC for the model predicting 30-day mortality was 0.928 (95% CI 0.843 – 1.0), for 30-day morbidity was 0.825 (95% CI 0.661 – 0.988) and LOS>30 days was 0.912 (95% CI 0.839 – 0.986). The GOF was 0.63 ($p = 0.42$), 0.25 ($p = 0.62$) and 0.97 ($p = 0.32$), respectively.

These statistics presented in Chapters 5-7 surpass the a priori cut-off of discrimination and calibration set in the hypothesis. These studies provide evidence to support the hypothesis that emerging m-Health technology provides a solution to mitigate the lack of reliable surgical outcomes research and surgical quality improvement programmes for general and vascular surgery in LMICs.

Chapter 8

The derivation dataset for a Trauma Quality Improvement Programme

The following prospective dataset was developed as guided by the analytical rationale of the ACS TQIP (160).

8.1 Materials and Methods

a. Patient population

All patients admitted to the GSH Trauma Unit during the 15-month period, April 2014- July 2015, were included in the study. The inclusion and exclusion criteria for the trauma surgery derivation dataset are presented in table 8.1.

Table 8.1 Inclusion and exclusion criteria for Trauma Surgery derivation dataset

Inclusion criteria
Age older than 12 years
Entered into eTHR*
Primary mechanism of injury classified as either blunt or penetrating
Blunt is defined as an injury classified by the admitting physician sustained after a fall, motor vehicle collision or after being struck by or against
Penetrating is defined as an injury classified by the admitting physician sustained after a bite, firearm, knife or sharp object assault
ED^ discharge disposition and hospital discharge disposition must be known
Exclusion criteria
Dead on arrival
Referred to another subspecialty (Orthopaedics, Neurosurgery etc).
Disposition treated and discharged or transferred to a step-down facility
Thermal injuries

eTHR electronic Trauma Health Registry ED^ Emergency Department*

b. Data collection

In 2013, study collaborators from the University of British Columbia designed

and implemented a practical, user friendly, mobile electronic Trauma Health Record (eTHR), for point of care data collection by front-line clinicians (172). eTHR was designed to populate standard clinical reports, while wirelessly populating an electronic trauma registry in real time, with standardised data. Within 18 months after implementation of the eTHR, electronically generated records replaced all the previous hand-written records in the GSH Trauma Unit. This data source was used for generating the derivation dataset. The first 3 months of data was discarded as pilot data, and only the data from 1st April 2014 to 7th July 2015 was used.

Variables

eTHR has the ability to collect 644 variables per admission. However, after reviewing both the guidelines for the ACS-TQIP programme and other literature, the variables included in the derivation dataset were divided as: variables collected on admission, variables describing detailed anatomical injury scoring, variables describing operative intervention and classification of endpoints.

(i) Variables collected at admission

The variables, which were generated prospectively by the admitting clinicians using eTHR, were retrospectively extracted for each patient included in the derivation dataset. These are shown in table 8.2.

Table 8.2 Data variables collected on admission for the Trauma Surgery derivation dataset

Variable	Data type; category list
Unique patient identifier	Numerical
Age	Continuous
Race	Categorical; 1. Black 2. Mixed ancestry 3. Indian 4. White 5. Other
Sex	Binary; male/female
Injury date	dd/mm/yyyy
Injury time	00h00
Date of admission	dd/mm/yyyy
Time of admission	00h00
Complaint	Categorical; 1. Fall 2. Motor vehicle collision 3. Struck by or against 4. Bite 5. Firearm 6. Knife 7. Sharp object 8. Other
Mechanism	Categorical; 1. Blunt 2. Penetrating
Reason	Categorical; 1. Community assault 2. Gang violence 3. Interpersonal violence 4. Other
Intentional	Binary; yes/ no
Injury location (suburb)	Free text
Referral	Categorical; 1. Scene of injury 2. Transfer from referral hospital (named)

Transport	Categorical; 1. Air 2. Ambulance 3. Police 4. Vehicle 5. Walked 6. Other
Charlson's comorbidity index	Ordinal; 0. Nil known 1. Myocardial infarction 1. Congestive heart failure 1. Peripheral disease (includes aortic aneurysm >= 6 cm 1. Cerebrovascular disease: CVA with mild or no residua or TIA 1. Dementia 1. Chronic pulmonary disease 1. Connective tissue disease 1. Peptic ulcer disease 1. Mild liver disease 1. Diabetes without end-organ damage 1. Hemiplegia 1. Moderate or severe renal disease 1. Diabetes with end-organ damage 1. Tumor without metastasis 1. Leukemia 1. Lymphoma 3. Moderate or severe liver disease 4. Metastatic solid tumor 4. AIDS
HIV status	Categorical; 1. HIV negative 2. HIV positive 3. HIV status unknown
CD4 status (if HIV status confirmed positive)	Categorical; 1. >350 2. <350 4. Unknown
Diabetic status	Binary; non-diabetic/ diabetic
Smoking status	Binary; current smoker/ non-smoker
Illicit drug use	Binary; yes/ no
Drug type	Categorical; 1. Alcohol 2. Cocaine 3. Marijuana 4. Narcotics

	5. Hallucinogens 6. Tik 1. Other
Admission vitals on arrival ***Temperature*** ***Pulse*** ***GCS (motor, verbal and eye response)*** ***Blood pressure*** ***Respiratory rate***	Continuous
Airway	Binary; maintained/ required intubation
Level of consciousness	Categorical; 1. Alert 2. Verbal 3. Painful stimulus Unresponsive
Laboratory values on arrival ***pH*** ***Lactate*** ***Haemoglobin***	Continuous

(ii) Variables describing detailed anatomical injury scoring

Abbreviated Injury scales (AIS) have been the most commonly used anatomical injury scoring systems for descriptive analyses, benchmarking and quality improvement initiatives in the trauma literature to date (176). The AIS classification system is a consensus-derived, anatomically based, seven-digit injury scoring system. The first six digits refer to a unique numerical identifier, which designates the injured body region (out of nine regions), the type of anatomic structure, and the specific anatomic structure. The seventh digit refers to an ordinal injury severity scale with categories ranging from one ('minor injury') to six ('maximal injury'). From the AIS scores, an ISS value, a pragmatic quantitative summary measure of the overall severity of anatomic and functional damage is generated. This is calculated by summing the squares of the highest AIS severity codes in each of the three (out of the nine) most severely injured ISS body regions (177).

In order to generate a detailed anatomical injury scoring system in the Trauma Unit, we implemented a protocol whereby all injuries were to be classified on discharge, according to the Abbreviated Injury Scale 2005 Update 2008. This was done by means of drop down menus, which included AIS injury descriptions, which were built into the eTHR application under the nine anatomical regions, which make up the scoring system. The application was then programmed to calculate the ISS using these clinician-entered AIS scores by summing the squares of the three most severely injured ISS body regions.

Due to the complexity and cost related to the collection of AIS and ISS, hospital-based trauma registries in limited-resource settings use the number of serious injuries categorised as nil, one or multiple as a marker for anatomical injury scoring. This data variable has also been added to eTHR as well as this derivation dataset.

(iii) Variables describing operative intervention

In addition to the data collected in the derivation dataset, table 8.3 describes the data variables collected for patients who were operatively managed during the study period.

Table 8.3 Additional data variables collected for operatively managed patients

Variable	**Data type; category list**
Operation date	Dd/mm/yyyy
Operation commencement time	00h00
Operation completion time	00h00
WHO Checklist utilized	Binary; yes/no
Triage color	Ordinal; 1. Purple 2. Green 3. Yellow 4. Orange 5. Red
Damage control	Binary; yes/ no
Role of registrar	Ordinal; 1. Independent

	2. Supervised 3. Assisted
Operation	Categorical; 1. Neck Dissection 2. Vascular Dissection 3. Removal of Foley Catheter 4. Antero-Lateral Thoracotomy 5. Postero-Lateral Thoracotomy 6. Emergency room Thoracotomy 7. Sternotomy> 8. Pericardial Window 9. VATS 10. Exploratory Laparotomy 11. Exploratory Laparoscopy 12. Relook Laparotomy 13. VAC change 14. Open abdomen closure 15. Stoma closure 16. Septic Wound Management 17. Rigid Sigmoidoscopy 18. Fasciotomy 19. Fasciotomy closure 20. Amputation 21. Skin graft 22. Wound Debridement 23. Burn wound management 24. Hernia repair
Procedure category	Categorical; 1. Cardiac 2. Extremity amputation 3. Neck 4. Thoracic 5. Gastrointestinal 6. Genitourinary 7. Vascular
Cardiac procedure sub-category	Categorical; 1. Pericardial window with negative findings 2. Pericardial window with drainage 3. Repair of cardiac injury
Neck procedure sub-category	Categorical; 1. Neck dissection with negative findings 2. Vascular repair 3. Tracheal repair
Thoracic procedure sub-category	Categorical; 1. Lung tractotomy 2. Repair of diaphragmatic injury 3. Segmental resection of lung 4. Suture of oesophageal wound 5. Wide drainage of oesophageal wound
Gastrointestinal procedure sub-category	Categorical; 1. Primary repair of colonic injury 2. Colostomy creation 3. Resection of colonic injury with primary anastomosis 4. Resection of small bowel with primary anastomosis")

	5. Primary repair of small bowel 6. Primary repair of stomach 7. Abdominal VAC application 8. Exploratory laparotomy with intra-abdominal collection drainage 9. Resection of colonic injury without primary anastomosis 10. Exploratory laparotomy with negative findings 11. Simple packing of liver 12. Distal Pancreatectomy 13. Repair of diaphragmatic injury 14. Splenectomy 15. Wide drainage 16. Colonic ligation 17. Exploratory laparotomy with abdominal packing 18. Resection, debridement, and primary repair of duodenum 19. Small bowel ligation 20. Primary repair of rectal injury 21. Repair of mesenteric injury 22. Suture repair of liver 23. Resectional debridement of liver 24. Simple packing of spleen 25. Partial gastrectomy with small bowel anastomosis 26. Duodenal exclusion 27. Ileostomy creation
Genitourinary procedure sub-category	Categorical; 1. Renorrhaphy 2. Nephrectomy 3. Ureteric stent insertion 4. Primary repair of bladder 5. Cystostomy
Vascular procedure sub-category	Categorical; 1. Repair of aortic injury 2. Repair of injury of major vessel in upper extremity 3. Repair of injury of major vessel in lower extremity

The role of the registrar in the surgery was defined as ‘independent’ if no consultant surgeon was present in theatre, ‘supervised’ if a consultant surgeon was present, but not for the entire procedure and ‘assisted’ if a consultant surgeon was present for the entire operation.

(iii) Classification of endpoints

The endpoints collected for the derivation Trauma Surgery dataset included the

presence or absence of the following adverse events: death, missed injuries, abdominal compartment syndrome, acute lung injury/respiratory distress, acute renal failure, bleeding requiring transfusion, cardiac arrest requiring CPR with return of spontaneous circulation, catheter related blood stream infections, cellulitis, coagulopathy, decubitis ulcer, delayed haemo/pneumothorax, extremity compartment syndrome, myocardial infarction, pneumonia, pulmonary embolus, retained haemothorax, stroke/CVA, systemic sepsis, unplanned intubation and unplanned return to the ICU. Additional endpoints collected for the operatively managed patients included: Surgical site infection (deep or organ space), wound disruption and unplanned reoperation.

These were similar complications to those audited by ACS-TQIP but were refined to only include those complications with clear diagnostic endpoints. To avoid misclassification bias, superficial surgical site infection, deep vein thrombosis and urinary tract infections were excluded. These adverse events were also classified according to the Clavien Dindo grading system, which was included in the eTHR (see table 8.4) (178).

Table 8.4 Clavien-Dindo Grading Classification used for outcome grading

Grade		Definition
1		Any deviation of the clinical course without the need for pharmacological treatment or surgical, radiological or endoscopic interventions.
11		Requiring pharmacological treatment other than such allowed for Grade 1 complications.
III		Requiring surgical, endoscopic or radiological intervention
	III a	Intervention not under general anaesthesia
	III b	Intervention under general anesthesia
IV		Life-threatening complication (including CNS complications)* requiring ICU management
	IV a	Single organ dysfunction (including dialysis)
	IV b	Multiorgan dysfunction
V		Death of a patient

Adapted from the Clavien-Dindo Classification of Surgical Complications (178)

Clinicians entered these endpoints into eTHR and classified the grading during daily ward rounds when eTHR was updated, as well as at the time of discharge. This ensured that patients were followed-up daily until hospital discharge. Furthermore, each mortality in the Trauma Unit was referred to the Unit Manager who collected the patient records for discussion at the monthly morbidity and mortality (M&M) meeting. The primary investigator compiled the reports and verified that any death and morbidity notification had been correctly entered. Finally, all in-hospital trauma deaths were discussed at the M&M meeting, and were classified into either preventable or non-preventable deaths. Outcome measures describing resource use were also audited including admission to the Intensive Care Unit (ICU) and length of stay greater than 30 days (LOS>30).

c. Analysis

The primary analysis of the Trauma Surgery derivation dataset was descriptive. Variables collected at admission were categorised and presented under the subheadings of patient reserve characteristics, pre-hospital circumstances of the injury event and patient characteristics describing the physiological injury sustained. The results of the detailed anatomical injury scoring using AIS 2005 update 2008 on eTHR, was then described for the derivation dataset under the nine anatomical body regions. An incidence rate ratio for each AIS grade was calculated. A description of the operations performed in the study period were classified by the primary operating surgeon into procedural categories (gastrointestinal, vascular, cardiac, genitourinary, thoracic, extremity amputations and neck) and further into sub-categories on eTHR. Outcome measures were described according to the defined endpoints as well as the Clavien-Dindo Grading classification. Throughout the descriptive analysis, rates of missing data were documented.

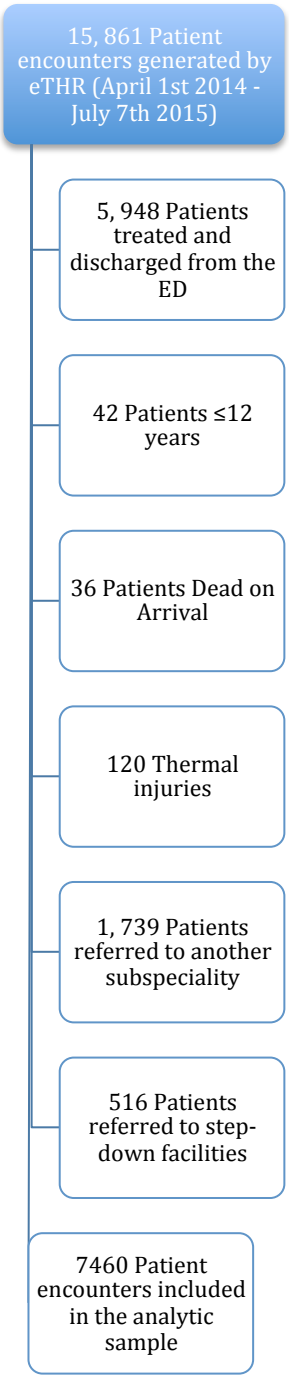
Adequate sample size ensured that confidence intervals for any descriptive estimates were adequately tight. The primary endpoint of in-hospital mortality was expected to occur at the lowest rate of all endpoints. With an estimated sample size of 5000 admissions during the 15-month study period, the 95% confidence interval around a mortality rate of 0.06 is 0.053-0.066.

8.1 Results

a. Population characteristics

A total of 15, 861 patient encounters were logged into the newly implemented eTHR during the study period 1st April, 2014 to 7th July, 2015. A total of 8, 401 patients were excluded for the following reasons: treated and discharged without admission (n=5,948), younger than 13 years of age (n=42), dead on arrival (n=36), transferred to a step-down facility (n=516), suffered a thermal injury (n=120) or referred to another subspeciality (n=1,739). A total of 7, 460 patients were included in the derivation dataset analysis. A flow diagram for this inclusion and exclusion process is presented in Figure 8.1.

Figure 8.1 Analytic sample inclusion process for the Trauma Surgery derivation dataset



i. Patient characteristics on arrival

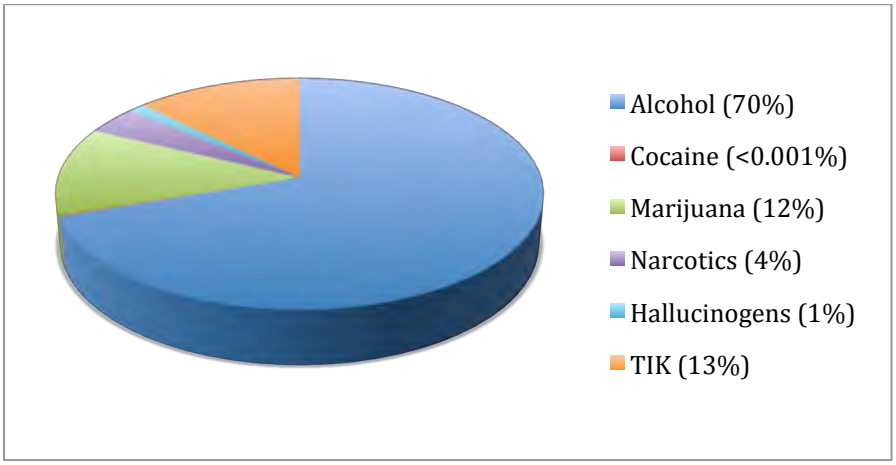
The variables describing patient characteristics on arrival are presented in Tables 8.5-8.7. Table 8.5 presents the data describing the patient's characteristics or 'reserve' as conceptualised by Osler (179).

Table 8.5 Patient reserve characteristics assessed on arrival in the Trauma surgery derivation dataset

Patient characteristics	Category	Derivation (%)
Gender	Male	5,788 (77.59)
	Female	1,659 (22.24)
	Missing	13 (0.17)
Age	<18	662 (8.87)
	18 - 29	3172 (42.52)
	30 - 39	1637 (21.94)
	40 - 49	884 (11.85)
	50 - 55	343 (4.6)
	65 - 69	149 (2.0)
	>70	503 (6.74)
	Missing	110 (1.48)
Race	Black	3411 (45.72)
	Mixed ancestry	3138 (42.07)
	Indian	135 (1.81)
	White	506 (6.79)
	Other	54 (0.73)
	Missing	216 (2.87)
Charlson's comorbidity index	0	7,328 (98.23)
	1	132 (1.77)
	Missing	0 (0)
Diabetic status	Diabetic	189 (2.53)
	Non-diabetic	7,721 (97.47)
	Missing	0 (0)
HIV status	HIV positive	214 (2.87)
	Unknown	4,246 (97.13)
	Missing	0 (0)
Smoking status	Current smoker	220 (2.95)
	Non-smoker	7,240 (97.05)
	Missing	0 (0)
Illicit drug use	Yes	1,829 (24.52)
	No	5,631 (75.48)
	Missing	0 (0)

The majority of the patients were male (77.6%), the mean age of the cohort was 33.3 (range 13-95), and the majority of patients (87.8%) were either Black (45.7%) or of Mixed ancestry (42.1%). Charlson's co-morbidity indices were very low with only 132 patients (1.8%) identified as scoring a one, and the recorded prevalence of diabetes mellitus was only 2.5%. The majority of patients were recorded to not know their HIV status (97.1%). The HIV status was reported as positive in 2.9%. Only 2.9% of the cohort admitted to smoking cigarettes but 24.5% of the cohort admitted to using illicit drugs prior to the traumatic incident or had evidence of their use as assessed by the admitting physician. The types of drugs used by these 1,829 patients were specified and are shown in Figure 8.2. All data fields were greater than 80% complete.

Figure 8.2 Illicit drug use reported by those patients with evidence of drug abuse of at the time of admission in the Trauma Surgery derivation dataset



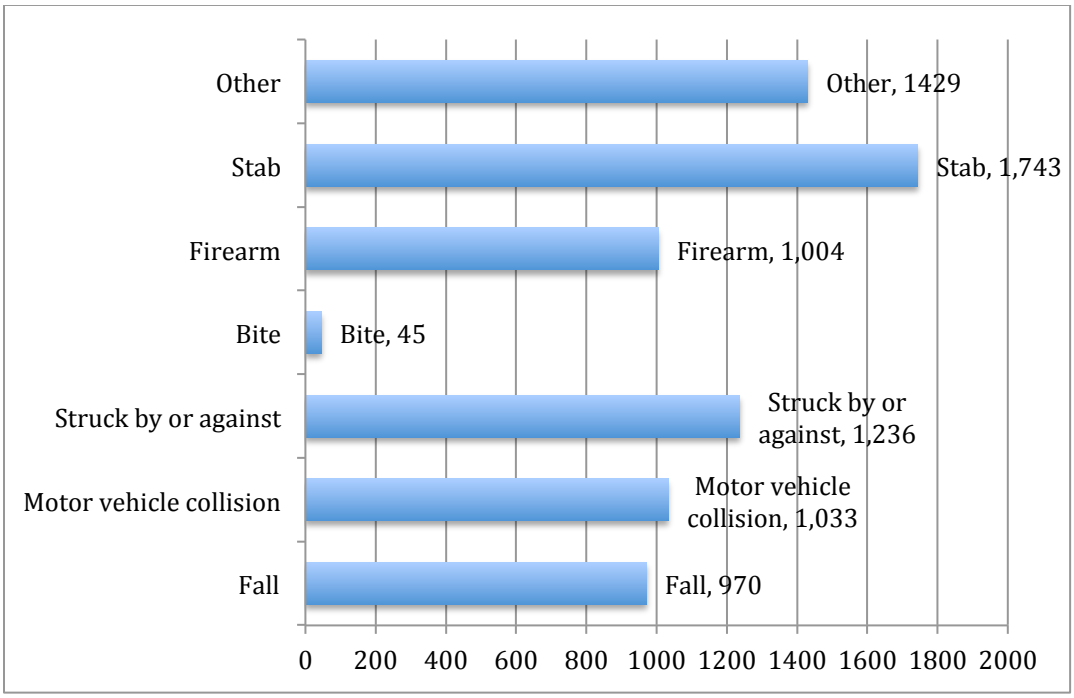
Data regarding the pre-hospital circumstances of the injury, which were included in the Trauma Surgery derivation dataset are presented in Table 8.6.

Table 8.6 Pre-hospital circumstances of the injury events included in the Trauma Surgery derivation dataset

Pre-hospital circumstances	Category	Derivation (%)
Complaint	Fall	970 (13.0)
	Motor vehicle collision	1,033 (13.85)
	Struck by or against	1,236 (16.57)
	Bite	45 (0.6)
	Firearm	1,004 (13.46)
	Stab	1,743 (23.36)
	Other	1429 (19.16)
	Missing	0 (0)
Mechanism	Blunt	4,534 (60.78)
	Penetrating	2,926 (39.22)
	Missing	0 (0)
Reason	Community assault	457 (6.13)
	Gang	604 (8.1)
	Interpersonal	2,250 (30.16)
	Other	349 (4.68)
	Missing	3,800 (50.94)
Intentional	Yes	3,708 (49.71)
	No	3,435 (46.05)
	Missing	317 (4.25)
Referral	Directly from the scene	2495 (33.45)
	False Bay Hospital	62 (0.83)
	GF Jooste Hospital	36 (0.48)
	Gugulethu	482 (6.46)
	Hanover Park	263 (3.53)
	Heideveld	198 (2.65)
	Mitchell's Plain	510 (6.84)
	New Somerset Hospital	201 (2.69)
	Retreat	89 (1.19)
	Vanguard	169 (2.27)
	Victoria Hospital	178 (2.39)
	Other	691 (9.26)
	Missing	2,086 (27.96)
Transport	Air	11 (0.15)
	Ambulance	4249 (56.96)
	Police	80 (1.07)
	Vehicle	1622 (21.74)
	Walked	519 (6.96)
	Other	75 (1.01)
	Missing	904 (12.12)

The mechanism of injury was classified as blunt in 60.78% and penetrating in 39.22%. The mechanism was further classified into complaints as shown in Figure 8.3.

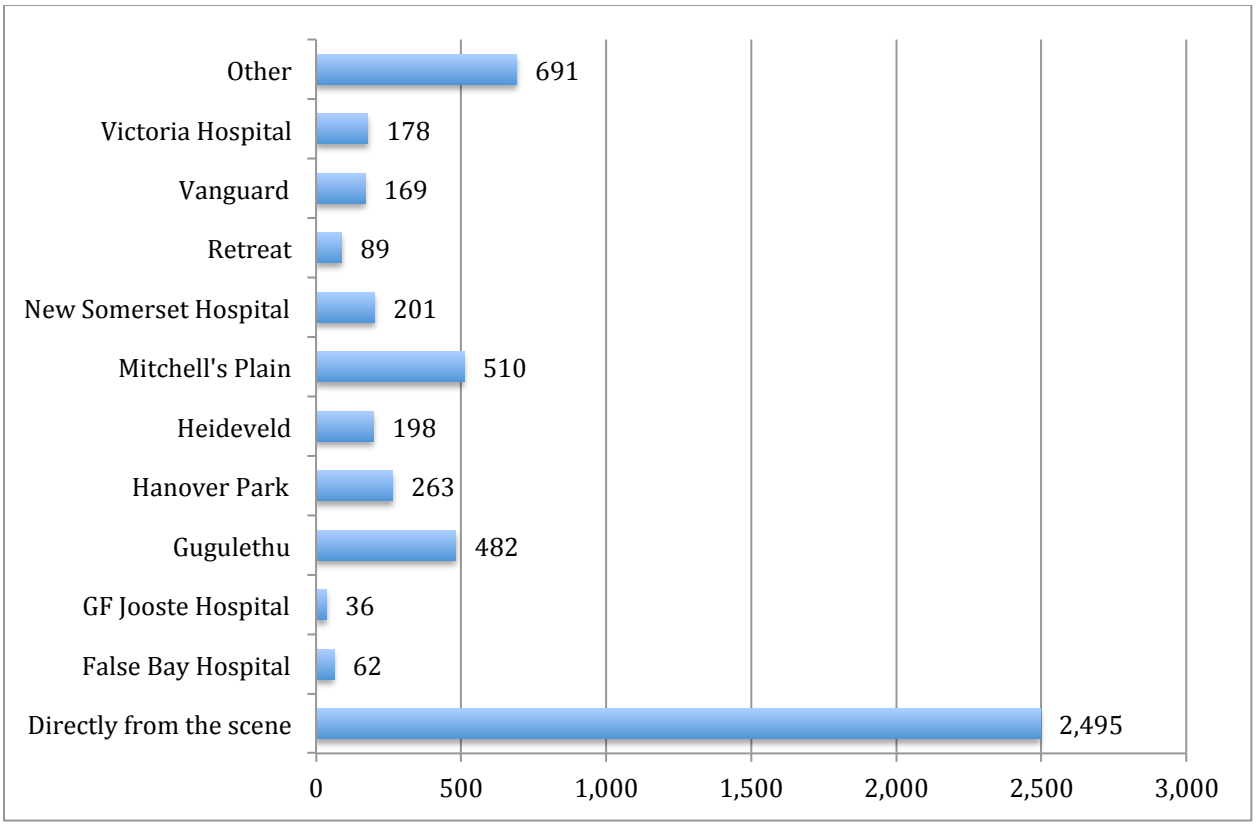
Figure 8.3 Mechanism of injury by presenting complaint



'Struck by or against' was the most common presenting complaint in the blunt category (17%), followed by 'motor vehicle collisions' (14%) and 'falls' (13%). Stab injuries were the most common form of penetrating category (23%), followed by firearm injuries (13%) and bite injuries (1%). A greater proportion of the injuries were classified as intentional (49.71%) compared to 46.05%, which were classified as non-intentional injuries. The commonest reason for intentional injuries was interpersonal violence (30.16%). However, the reason attributed at admission was poorly documented with a missing rate of 50.94%. The majority of the cohort arrived by ambulance (56.96%) and

only 0.15% were airlifted from the scene of injury. The place of transfer (referral) has been summarised in Figure 8.4 after excluding the 27.96% of cases who had missing variables for referral.

Figure 8.4 Place of transfer prior to arrival at GSH



The greatest proportion of patients presented from the scene of injury (46%), followed by Heideveld Day Hospital (13%), jointly by Mitchell's Plain and Vanguard Hospitals (9%), and Hanover Park Day Hospital (5%). Table 8.7 represents the categorical data describing the physiological injury status at admission.

Table 8.7 Patient characteristics describing the physiological injury sustained by the Trauma Surgery derivation dataset on admission

Physiological injury	Characteristics	Derivation N(%)
Hypothermic (<35.5 °C)	Yes	1,189 (15.94)
	No	4,801 (64.36)
	Missing	1,470 (19.71)
Pulse	Normal	4,442 (59.54)
	Tachycardic (>100 b/min)	2,405 (32.24)
	Bradycardic (<50 b/min)	133 (1.78)
	Missing	480 (6.43)
Hypotensive (<90mmHg)	Yes	201 (2.69)
	No	7,259 (97.31)
	Missing	0 (0)
Respiratory rate	<9	19 (0.25)
	9 - 30	6,343 (85.03)
	>30	639 (8.57)
	Missing	459 (6.15)
GCS*	12-15	6,656 (89.22)
	9-11	125 (1.68)
	6-8	128 (1.72)
	<6	551 (7.39)
	Missing	0 (0)
Level of consciousness	Alert	6,177 (82.82)
	Verbal stimulus	306 (4.11)
	Painful stimulus	183 (2.46)
	Unresponsive	142 (2.04)
	Missing	652 (8.57)
Acidotic (Ph< 7.32)	Yes	570 (7.64)
	No	530 (7.1)
	Missing	6,360 (85.25)
Anaemic (Hb<9.0)	Yes	115 (1.54)
	No	1,226 (16.43)
	Missing	6,119 (82.02)
Hyperlactataemia (lactate >2.2)	Yes	560 (7.51)
	No	526 (7.05)
	Missing	6,374 (85.44)
Intubated	Yes	389 (5.21)
	No	7,071 (94.79)
	Missing	0 (0)

GCS Glascow Coma Scale*

The majority of patients had physiological parameters within normal limits on arrival; 64.4% had a normal temperature, 59.5% had a normal pulse rate, 97.3% were normotensive, 85% had a normal respiratory rate and 82.8% had a normal level of consciousness. These data were complete for over 80% of the cohort. With regards to blood gas parameters on arrival, 7.6% of the cohort were acidotic, 1.5% were anaemic and 7.5% of the cohort had lactate levels greater than 2.2. However, these parameters were missing for greater than 80% of the derivation dataset. A small minority of the patients (5.2%) required respiratory support and intubation during transfer to GSH or on arrival.

ii. Detailed anatomical injury scoring

By incorporating the AIS 2005 update 2008 scoring system, into eTHR by means of descriptive drop-down menus, 6,501 patients in the Trauma Surgery derivation dataset (87.1%) had injury severity scores generated post-operatively or at discharge by a member of their care team. However, 12.9% of the cohort was discharged without an AIS or an ISS. Table 8.8 summarises the AIS scores of the cohort by the nine anatomical regions. Included in the table, is the frequency and rate of each AIS injury within the 9 anatomical regions, as well as an overall incidence rate ratio for each AIS for the 15-month period. The 6,501 patients contributed to 8,126 person-years of follow-up.

Table 8.8 Abbreviated Injury Severity Scores by body region described in the Trauma Surgery derivation dataset

Body region	AIS score	Anatomical frequency (%)	Incidence rate ratio (10 000 person-years)
Head	1	251 (39.59)	308.88
	2	97 (15.3)	119.36
	3	119 (18.77)	146.44
	4	102 (16.09)	125.52
	5	58 (9.15)	71.37
	6	7 (1.1)	8.61
Face	1	230 (69.91)	283.04
	2	90 (27.36)	110.75
	3	9 (2.74)	11.07
	4	0 (0)	0
	5	0 (0)	0
	6	0 (0)	0
Neck	1	105 (78.95)	129.21
	2	11 (8.27)	13.54
	3	12 (9.02)	14.77
	4	4 (3.01)	4.92
	5	1 (0.75)	1.23
	6	0 (0)	0
Thorax	1	105 (22.63)	129.21
	2	13 (2.8)	15.99
	3	292 (62.93)	359.34
	4	45 (9.7)	55.37
	5	8 (1.72)	9.84
	6	1 (0.22)	1.23
Spine	1	0 (0)	0
	2	38 (48.72)	46.76
	3	15 (19.23)	18.45
	4	11 (14.1)	13.53
	5	13 (16.67)	15.99
	6	1 (1.28)	1.23
Abdomen	1	51 (13.78)	62.76
	2	82 (22.16)	100.91
	3	119 (32.16)	146.44
	4	96 (25.95)	118.13
	5	22 (5.95)	27.07
	6	0 (0)	0
Upper extremity	1	256 (52.46)	315.03
	2	166 (34.02)	204.28

	3	66 (13.52)	81.22
	4	0 (0)	0
	5	0 (0)	0
	6	0 (0)	0
Lower extremity	1	233 (51.78)	286.73
	2	113 (25.11)	139.03
	3	86 (19.11)	105.83
	4	18 (4.0)	22.15
	5	0 (0)	0
	6	0 (0)	0
External	1	1 (33.33)	1.23
	2	1 (33.33)	1.23
	3	1 (33.33)	1.23
	4	0 (0)	0
	5	0 (0)	0
	6	0 (0)	0

The most commonly described injury, was an AIS grade 3 to the thorax with 292 cases (395.34/ 10,000 person years), followed by 256 AIS grade 1 cases to the upper extremity (315.03/10,000 person years), 251 AIS 1 grade cases to the head (308.88/10,000 person years), 233 AIS grade 1 cases to the lower extremity (286.73/10,000 person years) and 230 AIS grade 1 cases to the face (283.04/ 10,000 person years). The commonest AIS grade 1 and 2 injuries were therefore, to the upper extremity (256 cases and 166 cases, respectively), AIS grade 3 injuries to the thorax (105 cases), and AIS grade 4, 5 and 6 injuries to the head (102 cases, 58 cases and 7 cases, respectively). AIS injuries to the external body region were very rarely described. Table 8.9 lists the results of the ISS calculations.

Table 8.9 The calculated injury severity scores for the 6 501 patients with completed AIS scoring in the Trauma Surgery derivation dataset.

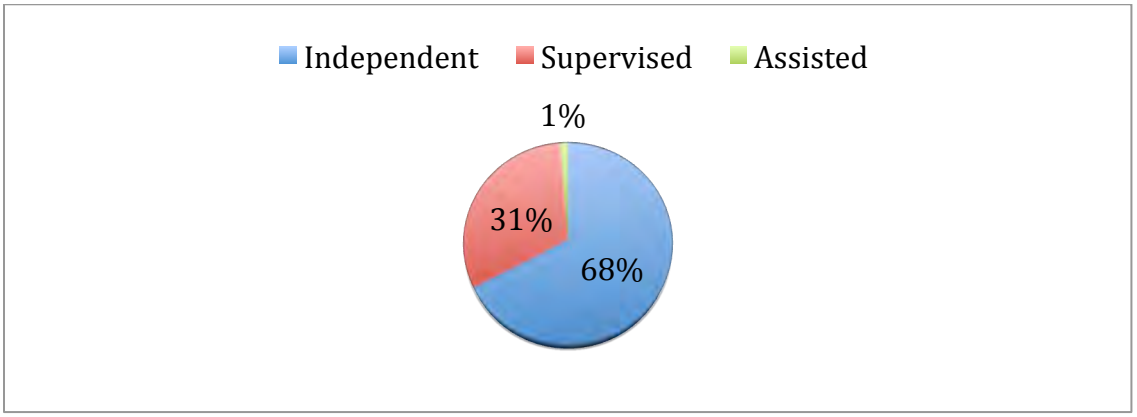
Injury Severity Score	Frequency (%)
0	1,373 (21.12)
1	2,151 (33.09)
2	190 (2.92)
3	41 (0.63)
4	889 (13.67)
5	45 (0.69)
6	7 (0.11)
8	35 (0.54)
9	825 (12.69)
10	64 (0.98)
11	8 (0.12)
12	8 (0.12)
13	77 (1.18)
14	18 (0.28)
16	298 (4.58)
17	37 (0.57)
18	47 (0.72)
19	7 (0.11)
20	33 (0.51)
21	5 (0.08)
22	30 (0.46)
24	5 (0.08)
25	186 (2.86)
26	14 (0.22)
27	13 (0.2)
29	28 (0.43)
30	1 (0.02)
32	11 (0.17)
34	17 (0.26)
35	2 (0.03)
36	1 (0.02)
38	7 (0.11)
41	3 (0.05)
43	3 (0.05)
50	4 (0.06)
59	1 (0.02)
75	17 (0.26)
Missing	959 (12.86)

The median ISS was 5 (IQR 1-9) and according to the ISS grading by Adil Haider *et al*, 976 (35.54%) patients were ISS category 1 (3<ISS<9), 1000 patients were ISS category 2 (8<ISS<16), 462 patients were ISS category 3 (15<ISS<25) and 308 patients were ISS category 4 (ISS ≥25) (170). Of the cohort with an ISS calculated, 27.2% had an ISS of greater than 8 and therefore according to the ACS-TQIP, were severely injured. These patients were included in the severely injured subset validation analysis described in Chapter 9. Nine hundred and fifty patients underwent operative management.

b. Operative procedures: descriptive analysis

950 Patients required operative management under general anaesthetic during the study period and these patients were included in the operative validation dataset. The mean age of the cohort was 28.5 (range 13 – 84) with a mean ISS of 11.9 (range 0 – 75). Seventy six (8.0%) of the cohort required damage control procedures with at least one planned re-operation. Every case had a surgical registrar present and his or her role was independent in 68.2%, supervised in 31.1% and assisted in 0.6% of cases as shown in figure 8.5.

Figure 8.5 Surgical registrars involvement in operations performed



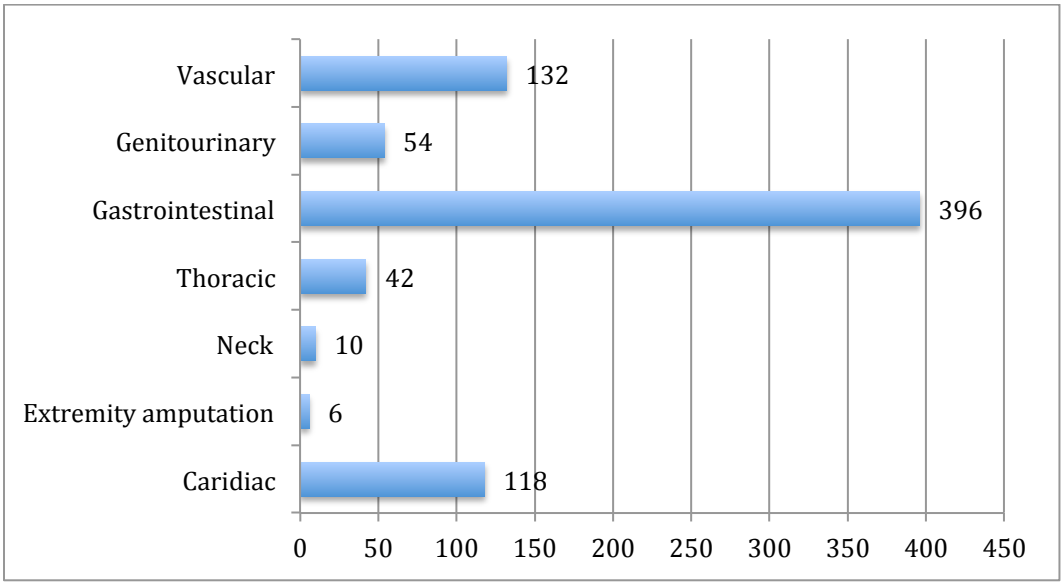
The procedures performed are presented in table 8.10.

Table 8.10 Procedures performed following trauma

Procedure	N (%)
Amputation	6 (0.64)
Antero-Lateral Thoracotomy	15 (1.60)
Burn wound management	2 (0.21)
Emergency room Thoracotomy	5 (0.53)
Exploratory Laparoscopy	19 (2.03)
Exploratory Laparotomy	355 (37.89)
Fasciotomy	10 (1.07)
Fasciotomy closure	7 (0.75)
Hernia repair	1 (0.11)
Neck dissection	6 (0.64)
Open abdomen closure	12 (1.28)
Open abdomen closure	16 (1.71)
Other	13 (1.37)
Pericardial Window	110 (11.74)
Postero-Lateral Thoracotomy	2 (0.21)
Relook Laparotomy	67 (7.15)
Removal of Foley catheter	21 (2.24)
Rigid Sigmoidoscopy	34 (3.63)
Septic Wound Management	42 (4.48)
Skin graft	17 (1.81)
Sternotomy	8 (0.84)
VAC change	49 (5.23)
Vascular dissection	74 (7.90)
VATS	16 (1.71)
Wound Debridement	43 (4.59)
Total	950

The primary operating surgeons classified 758 (79.7%) of the procedures into one of the following categories in eTHR: cardiac, extremity amputation, neck, thoracic, gastrointestinal, genitourinary or vascular at the time of writing the operative record. This information is presented in figure 8.6.

Figure 8.6 Operations by procedure category



Gastrointestinal procedures were most commonly performed (52.2%), followed by vascular (17.4%), cardiac (15.6%), genitourinary (7.1%), thoracic (5.5%), extremity amputations (1.3%) and neck (0.7%) procedures. Within each procedure category, these operations were further subdivided and these data are presented in table 8.11 and the following figures 8.7-8.11.

Table 8.11 Gastrointestinal procedures performed by order of frequency

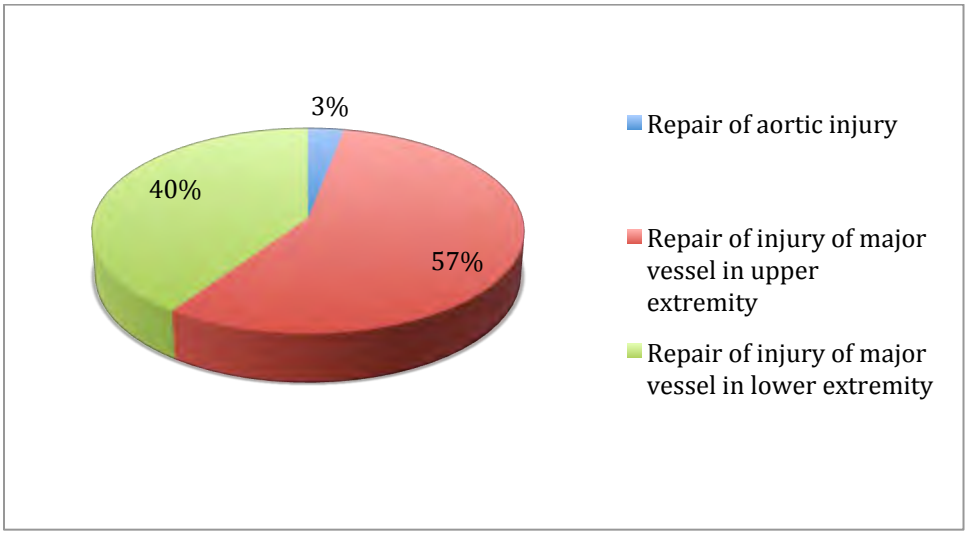
Gastrointestinal procedure	Frequency (%)
Primary repair of colonic injury	81 (18.88)
Colostomy creation	49 (11.42)
Resection of colonic injury with primary anastomosis	47 (10.96)
Resection of small bowel with primary anastomosis	46 (10.72)
Primary repair of small bowel	42 (9.79)
Primary repair of stomach	28 (6.53)
Abdominal VAC application	25 (5.83)
Exploratory laparotomy with intra-abdominal collection drainage	18 (4.2)
Resection of colonic injury without primary anastomosis	10 (2.33)
Exploratory laparotomy with negative findings	9 (2.1)

Simple packing of liver	8 (1.86)
Distal Pancreatectomy	8 (1.86)
Repair of diaphragmatic injury	7 (1.63)
Splenectomy	7 (1.63)
Wide drainage	7 (1.63)
Colonic ligation	7 (1.63)
Exploratory laparotomy with abdominal packing	6 (1.4)
Resection, debridement, and primary repair of duodenum	6 (1.4)
Small bowel ligation	5 (1.17)
Primary repair of rectal injury	3 (0.7)
Repair of mesenteric injury	2 (0.47)
Suture repair of liver	2 (0.47)
Resectional debridement of liver	2 (0.47)
Primary repair of splenic injury	1 (0.23)
Partial gastrectomy with small bowel anastomosis	1 (0.23)
Duodenal exclusion	1 (0.23)
Ileostomy creation	1 (0.23)
Total	429

Of the 396 operations classified as gastrointestinal, 429 gastrointestinal sub-categories were further described. The most frequently performed procedures in the operative validation dataset were primary repair of colonic injury (18.8%), colostomy creation (11.4%), resection of colonic injury with primary anastomosis (10.9%), resection of small bowel with primary anastomosis (10.7%) and primary repair of the small bowel (9.8%). Only one of the following procedures was reportedly performed: primary repair of splenic injury, partial gastrectomy with small bowel anastomosis, duodenal exclusion and ileostomy creation.

Figure 4.5 describes the vascular procedures performed during the study period.

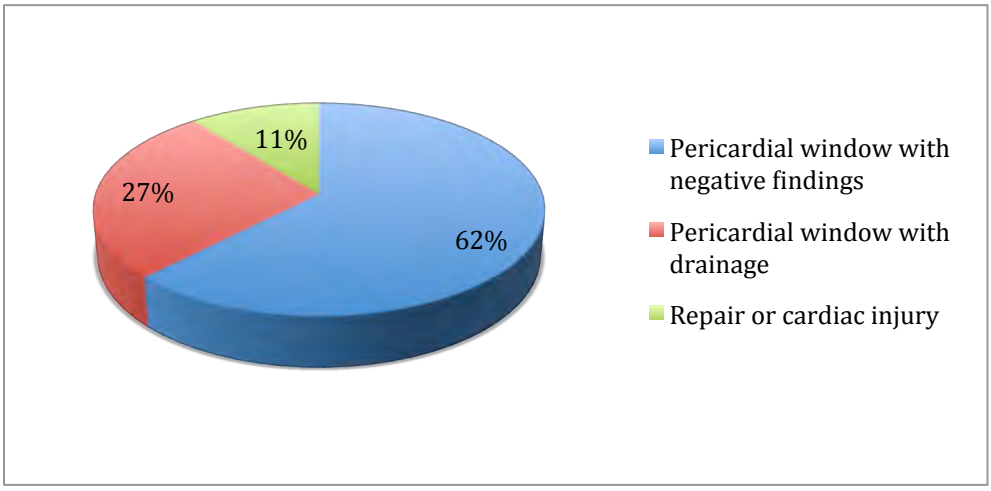
Figure 8.7 Vascular procedures performed



Of the 132 procedures classified as vascular, 99 vascular operations were further sub-classified. Of these, 56 (56.4%) were open repair of an injury to a major vessel in the upper extremity, 40 (39.6%) were open repair of an injury to a major vessel in the lower extremity and 3 (2.9%) were open aortic repairs.

Figure 8.8 describes the cardiac procedures reported during the study period.

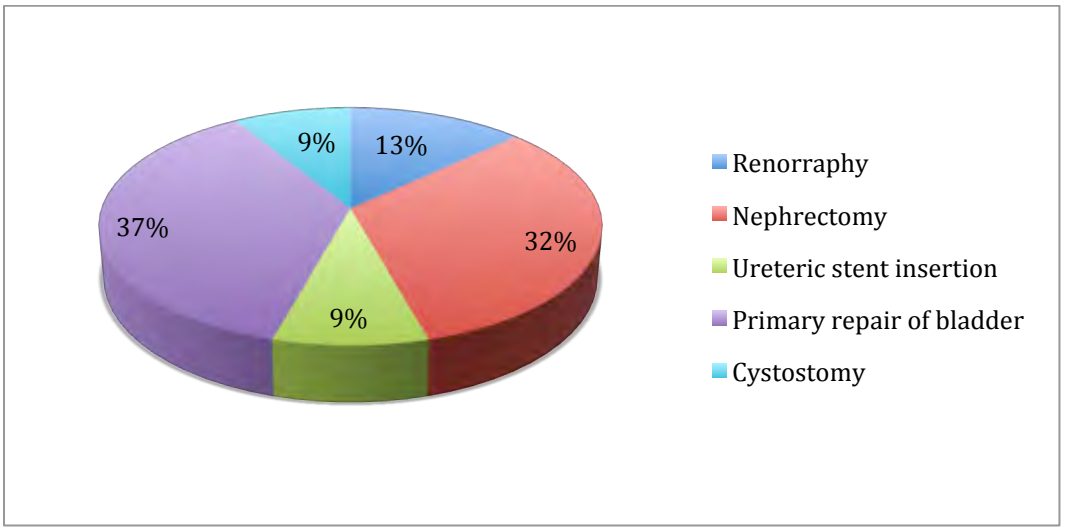
Figure 8.8 Cardiac procedures performed



Of the 118 procedures classified as cardiac, 130 cardiac procedures were further sub-classified. Of these, 81 (62.3%) were negative pericardial sub-xiphoid windows, 35 (26.9%) were positive pericardial sub-xiphoid windows and 14 (10.7%) were direct suture repair of a cardiac injury.

Figure 8.9 describes the genitourinary procedures reported during the study period.

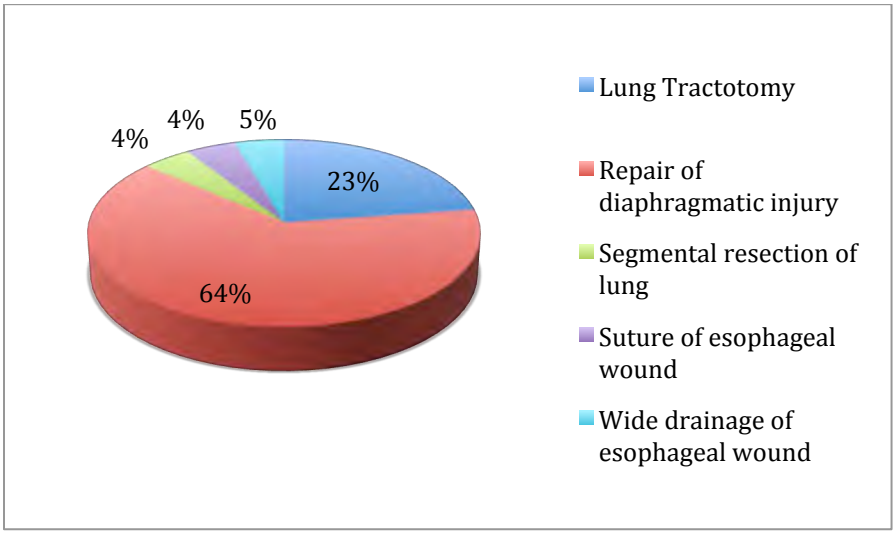
Figure 8.9 Genitourinary procedures performed



Of the 54 procedures classified as genitourinary, 46 procedures genitourinary procedures were further sub-classified. Of these 17 (36.9%) were primary repair of a bladder injury, 15 (32.6%) were nephrectomies, 6 (13.1%) were renorrhaphies, 4 (8.7%) were cystostomies and 4 (8.7%) were ureteric stent insertions.

Figure 8.10 describes the thoracic procedures reported during the study period.

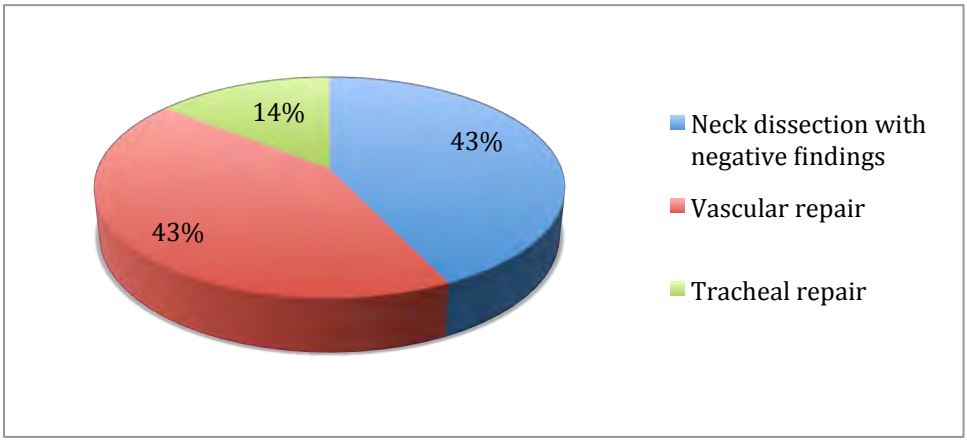
Figure 8.10 Thoracic procedures performed



Of the 42 procedures classified as thoracic, 22 procedures were further sub-classified. Of these 14 (63.6%) were repair of diaphragmatic injuries, 5 (22.7%) were lung tractotomies and the remaining 3 cases were made up of one segmental lung resection, one esophageal wound suture and one wide drainage of an esophageal wound.

Figure 8.11 describes the neck procedures reported during the study period.

Figure 8.11 Neck procedures performed



Of the 10 procedures classified as neck procedures, 7 procedures were further sub-classified. Of these 3 (42.8%) were negative neck dissections, 3 (42.8%) were vascular repairs and 1 (14.2%) was a tracheal repair.

c. Outcome measures

The outcome measures for the Trauma Surgery dataset are presented in Table 8.12. There were 188 in-hospital deaths during the study period giving an in-hospital mortality rate of 2.52% (95% CI 2.18 – 2.90). In addition, there were 205 reported complications, giving an overall in-hospital adverse event rate of 5.23% (95% CI 4.72 – 5.73). During the 15-month study period, 174 patients were admitted to an ICU, giving an ICU admission rate of 3.59% (95% CI 3.11 – 4.16). There were 592 patients who had a hospital stay of greater than 30-days, resulting in a LOS>30 days rate of 7.94% (95% 7.34 – 8.57).

Table 8.12 Endpoint assessments in the Trauma Surgery derivation dataset

Outcome measure	Rate %	95% CI	Missing endpoint frequency (%)
In-hospital mortality	2.52	2.18 - 2.90	0 (0)
In-hospital adverse event	5.23	4.72 - 5.73	0 (0)
ICU admission	3.59	3.11 - 4.16	2,624 (34.17)
LOS>30	7.94	7.34 - 8.57	0 (0)

The recorded adverse events were then further classified using the Clavien Dindo grades as presented in Table 8.13.

Table 8.13 Adverse events classified by Clavien Dindo Grading in the Trauma Surgery derivation dataset.

Clavien Dindo Grade	Frequency (%)
0	7072 (94.73)
1	54 (0.72)
2	84 (1.13)
3a	26 (0.35)
3b	39 (0.52)
4	2 (0.03)
5	188 (2.52)

The majority of the patients (94.73%) had no reported adverse events. Fifty-four patients (0.72%) had a reported adverse event, which did not require an intervention. Eighty four patients (1.13%) had an adverse event requiring pharmacological treatment and 65 patients (0.87%) had a complication requiring surgical intervention. Two patients suffered life-threatening complications and 188 patients died. All these in-hospital deaths were discussed at the monthly in-house morbidity and mortality meeting and after a multi-disciplinary discussion, 42 of the 188 deaths (22.34%) were classified as preventable deaths. The preventable death rate of the entire cohort was 0.563% (95% CI 0.406 – 0.761).

Chapter 9

Derivation and validation of the prediction rules developed for the Trauma Quality Improvement Programme

9.1 Materials and Methods

Osler *et al* conceptualised outcome prediction in trauma surgery into a useful equation (179):

Outcome = anatomical injury + physiological injury + patient reserve

However, due to the heterogeneous nature of traumatic injuries, detailed anatomical injury descriptions such as the Injury Severity Score (ISS) are not available at the clinician's disposal at the time of admission. On the other hand, physiological and patient reserve data are. For this reason, two prediction rules were developed and validated for GSH Trauma Surgery:

1. A triage prediction rule, which is useful for application using data available to a clinician on admission
2. An injury severity prediction rule, which is derived using all three components of Osler's algorithm.

The primary outcome for prediction was in-hospital mortality.

a. Validation datasets

For the purpose of developing a quality improvement programme for trauma surgery at GSH, two cohorts of patients were prioritised viz. severely injured patients and those patients requiring operative intervention. Both prediction rules were generated using the

same derivation dataset described above and validated on two subset cohorts namely:

1. A severely injured validation dataset, which included any patient in the derivation dataset with an ISS of greater than or equal to 9 (as per the ACS-TQIP definition of a severe injury).
2. An operative validation dataset, which included any patient in the derivation dataset who underwent an operation under General Anaesthetic (GA) during the study period.

b. Analysis plan

The data collected on admission in Trauma Surgery derivation and validation datasets were compared using chi-squared proportional probability testing with a significance level of $p<0.05$. Proportional probability testing was used to compare rates of endpoints in severely and non-severely injured patients, as well as in operatively and non-operatively managed patients ($p<0.05$). In addition, a comparison of the anatomical injury scoring was performed between the derivation and operative validation datasets. Depending on the degree of normality of the ISS in both datasets, either a *t*-test or Kruskal-Wallis rank sum test was performed to compare the ISS of the operatively and non-operatively managed patients.

Identification of predictors

The univariate relationship between each categorical predictor variable collected at admission, which was greater than 80% complete, and in-hospital mortality, any in-hospital morbidity, ICU admission and LOS>30 was tested using $\chi 2$ or Fisher's exact tests, where appropriate. Significance was set at $p<0{\cdot}1$ to err on the side of inclusion. Furthermore, the univariate relationship of the ISS and the outcome measures was

performed using the student *t*-test or Wilcoxon Rank Sum test, depending on the degree of normality of the data.

Building the models

Because the majority of trauma occurs in resource limited settings, where current trauma scoring systems are not used routinely, empahsis was placed on the most simple and parsimonious models. This was achieved by giving prefence to data variables which seem to have a low data burden (measured by low rates of missing data), as well as by using categorical or binary data wherever possible, thereby putting less emphasis on the granularity of the data required. Furthermore, datapoints which have been included in well established trauma scoring systems like the Revised Trauma Score, Glasgow Coma Score, Kampala Trauma Score and the Trauma Injury Severity Score, were scrutinised for face validity. The significance level for entry into the model was set at $p = 0.1$, in order to err on the side of inclusion with model construction. A forward selection algorithm was then used, whereby each variable was screened individually and added into the multivariable logistic regression model in order of statistical significance. After each variable addition, the change of the risk estimates and their associated standard errors were reviewed to screen for co-linearity.

Included in the derivation dataset, were the outcomes data for 7,460 patients with an incidence of in-hospital mortality of 2.52% and a total of 188 deaths. Thus, applying the 10:1 rule, it was feasible to consider up to 19 variables in a prediction rule predicting in-hospital death.

Constructing a scoring system

The beta-co-efficients of the GSH triage prediction rule were then divided and rounded to the nearest integer to create a scoring system which could easily be applied by the physician admitting a trauma patient. This was called Groote Schuur Trauma Score (GSTS).

Validating the prediction rules

The methodological and geographical transportability of the newly derived prediction rules were validated. This involved comparing the performance of the GSH derived trauma prediction rules and scoring system against other well validated trauma scores and prediction rules, which have been developed and adopted in diverse settings around the world. This was achieved by performing head-to-head analyses with the GSH derived prediction rules and the following validated trauma-scoring systems: Revised Trauma Score (RTS), Kampala Trauma Score (KTS) and the Trauma and Injury Severity Score (TRISS).

1. *Revised Trauma Score (RTS)*

The RTS, introduced in the early 1980s, is one of the most commonly used physiological trauma scores (180). A weighted RTS ranging from 0 to 7.8408 has been calculated using eTHR with the following built-in formula:

RTS = 0.7326*(Systolic blood pressure) + 0.2908*(Unassisted respiratory rate) + 0.9368*(Glascow Coma Scale)

As the RTS was a score, which can be calculated at the time of admission, its performance to predict in-hospital mortality was assessed against the GSTS.

2. Kampala Trauma Score (KTS)

The Kampala Trauma Score (KTS) was developed and tested in 2000 by Kobusingye and Lett to create an injury severity score for resource-limited settings, which required minimal data collection and recording (181). KTS, which relied on the number of serious injuries, age, systolic blood pressure, respiratory rate, and neurologic status, was shown to be highly predictive of the need for admission or death (182). The total KTS ranges from 5 to 16, with lower scores indicating more severe injury. The scoring rubric for the KTS is shown in Table 9.1. This scoring algorithm was built into eTHR, and a KTS score could, therefore, be obtained from eTHR for every patient included in the derivation dataset, with the necessary completed data.

Table 9.1 Components of the Kampala Trauma Score

Component	**Score**
Age (years)	
5-55	2
<5 or >55	1
Systolic blood pressure (mmHg)	
>89	4
50-89	3
1-49	2
Undetectable	1
Respiratory rate (/min)	
10-29	3
>30	2
<9	1
Neurological status	
Alert	4
Responds to verbal stimuli	3
Responds to painful stimuli	2
Unresponsive	1
Serious injuries	
None	3
1	2
≥2	1
Total score	5-16

The KTS was a combined trauma scoring system, which took into account all three components of Osler's equation for predicting an outcome including physiological injury, patient reserve and anatomical injury. Due to the simplicity of the anatomical injury scoring system incorporated into the KTS it has been adopted as a triage prediction rule and its performance will therefore be compared to the GSTS.

3. The Trauma and Injury Severity Score (TRISS)

TRISS was developed by Champion and Boyd *et al* (161), and combined both anatomical measures, physiological measures of injury severity and patient reserve (ISS, RTS and patients age respectively) in order to predict survival from trauma (Ps), using the following formula:

$$\mathbf{Ps = 1/1+e^{-b}}$$

Where 'b' was calculated as follows: b = b 0 + b 1 (RTS) + b2 (ISS) +B3 (Age>54 years) (161).

The co-efficients b 0 to b 3 (Table 9.2) were derived from multiple-regression analysis of the Major Trauma Outcome Study database. The Age Index was 0 if the patient was below 54 years of age or 1 if 55 years and over. The co-efficients (b 0—b 3) were different for blunt and penetrating trauma. If the patient was less than 15 years old, the blunt co-efficients were used regardless of the actual mechanism of injury.

TRISS has become the standard method for outcome assessment.

Table 9.2 Co-efficients (b) used in determining survival probability in the Trauma and Injury Severity Score

Co-efficient	Blunt	Penetrating
b0	-0.4499	-2.5355
b1	0.8085	0.9934
b2	-0.0835	-0.0651
b3	-1.7430	-1.1360

Using these tabulated co-efficients, and applying them to the derivation dataset it was possible to calculate a TRISS survival probability for each patient with the necessary completed data in the derivation dataset. The performance of the TRISS prediction rule was then compared to the GSH injury scoring prediction rule.

For the purpose of developing a quality improvement programme for trauma surgery at GSH, as mentioned, two cohorts of patients were prioritised viz. severely injured patients and those patients requiring operative intervention. For this reason, the head-to-head comparisons described were performed using: 1) the entire derivation dataset, 2) only severely injured patients (ISS≥9) and 3) only patients requiring operative intervention. In every analysis, the discriminative ability (ROC) and the calibration (GOF) to predict the primary outcome of in-hospital mortality of the GSTS, the GSH injury scoring prediction rule and these validated trauma scoring systems were compared. A comparison of the degree of data burden by comparing rates of missing data was also performed. An a priori level of discrimination of 0.7 and calibration of p greater than 0.05 has been set in the hypothesis.

c. Using co-efficient based injury severity scores to generate risk-adjusted outcomes

Using the co-efficient based scores, namely RTS and TRISS, survival probabilities have been calculated for each patient in the derivation dataset according to the MTOS prediction rules. These resultant probabilities were then subtracted from one to generate an individual probability of death for each subject. These individual probabilities were then aggregated and rounded to the nearest integer to generate an expected outcome (E) for the cohort, which could then be compared to the observed deaths (O) in subjects with complete data for the injury score used (RTS or TRISS). The resultant O/E ratio with accompanying 95% CI's could then be calculated for the entire cohort including severely injured and operatively managed patients, benchmarking the GSH trauma unit against the MTOS cohort.

d. Power calculation

For the post-hoc assessment of power to test the measure of association for a single predictor in a model to predict in-hospital mortality, which occurs at a rate of 0,06, an odds ratio of 1.5 was chosen. Using the SAS university proc power, logistic option and the Shieh-O'Brien approximation, a total of 1,080 patients would be needed to power this study at 90%, as shown in the SAS University output below (Table 9.3).

Table 9.3 Power Calculation for a trauma surgery model predicting in-hospital mortality using ISS as a single predictor.

Method	Shieh-O'Brien approximation
Alpha	0.05
Response Probability	0.06
Test Predictor	ISS
Odds Ratio for Test Predictor	1.5
Unit for Test Pred Odds Ratio	1
Nominal Power	0.9
Total Number of Bins	4

Computed N Total	
Actual Power	**N Total**
0.900	1080

All analyses were performed using either SAS University edition or STATA 14.

9.2 Results

a. Population characteristics

Comparative analysis: Derivation and severely injured validation datasets

Patients with an ISS>8 were included in the severely injured validation dataset, which totaled 1,770 patients. Of these, 1,000 (56.5%) patients had an ISS between 9 and 15, 462 (26.10%) had an ISS between 16 and 24, and 308 patients (17.4%) had an ISS of ≥ 25. Tables 9.4 – 9.7 show the comparative analysis of the derivation and severely injured validation datasets using data taken on admission which was >80% complete.

Table 9.4 Comparison of patient reserve characteristics between Trauma Surgery derivation and severely injured validation datasets

Patient reserve characteristic	Category	Derivation N (%)	Severely injured N (%)	P-value
Gender	Male	5,788 (77.59)	1,587 (89.66)	
	Female	1,659 (22.24)	181 (10.23)	<0.0001
Age	<18	662 (8.87)	174 (9.89)	0.08
	18 - 29	3172 (42.52)	860 (48.89)	<0.0001
	30 - 39	1637 (21.94)	393 (22.34)	0.36
	40 - 49	884 (11.85)	184 (10.46)	0.1
	50 - 55	343 (4.6)	54 (3.07)	0.004
	65 - 69	149 (2.0)	19 (1.08)	0.009
	>60	503 (6.74)	70 (3.98)	<0.0001
Race	Black	3,411 (45.72)	803 (45.35)	0.78
	Mixed ancestry	3,138 (42.07)	748 (42.29)	0.87
	Indian	135 (1.81)	33 (1.87)	0.86
	White	506 (6.79)	15 (0.82)	<0.0001
	Other	54 (0.73)	55 (3.15)	<0.0001
Charlson's comorbidity index	0	7,328 (98.23)	1748 (98.76)	
	1	132 (1.77)	22 (1.24)	0.11
Diabetic status	Diabetic	189 (2.53)	24 (1.36)	
	Non-diabetic	7,721 (97.47)	1,746 (98.64)	0.003
HIV status	HIV positive	214 (2.87)	37 (2.09)	
	Unknown	4,246 (97.13)	1,733 (97.91)	0.07
Smoking status	Current smoker	220 (2.95)	33 (1.86)	
	Non-smoker	7,240 (97.05)	1,737 (98.14)	0.012
Illicit drug use	Yes	1,829 (24.52)	516 (29.15)	
	No	5,631 (75.48)	1,254 (70.85)	0.0001

Patients with an ISS of greater than 8 were more likely to be male (89.6% vs. 77.6%, p<0.0001) and were younger than the total group of patients in the derivation dataset, based on a higher proportion of patients in the 18-29 age category (48.9% vs. 42.4%, p<0.0001) and fewer patients in all age categories greater than 50 years of age. The derivation dataset had a higher proportion of diabetic patients (2.5% vs. 1.4%, p = 0.003). The severely injured cohort had more patients who were classified as non-

smokers (98.1% vs. 97.1%, p = 0.012) but a higher proportion of patients with evidence of illicit drug use on admission (29.2% vs. 24.5%, p = 0.0001).

Table 9.5 Comparison of pre-hospital circumstances of the injury events between the Trauma Surgery derivation and severely injured validation datasets

Prehospital circumstances	Category	Derivation (%)	Severely injured N (%)	P-value
Complaint	Fall	970 (13.0)	103 (5.82)	<0.001
	Motor vehicle collision	1,033 (13.85)	36 (2.03)	<0.001
	Struck by or against	1,236 (16.57)	221 (12.49)	<0.001
	Bite	45 (0.6)	1 (0.06)	<0.001
	Firearm	1,004 (13.46)	503 (28.42)	<0.001
	Stab	1,743 (23.36)	542 (30.62)	<0.001
	Other	1429 (19.16)	174 (9.83)	<0.001
Mechanism	Blunt	4,534 (60.78)	697 (39.38)	
	Penetrating	2,926 (39.22)	1,073 (60.62)	<0.001
Intentional	Yes	3,708 (49.71)	1,203 (70.64)	
	No	3,435 (46.05)	500 (29.36)	<0.001
Transport	Air	11 (0.15)	5 (0.31)	0.15
	Ambulance	4249 (56.96)	1,493 (92.1)	<0.001
	Police	80 (1.07)	5 (0.31)	0.003
	Vehicle	1622 (21.74)	91 (5.61)	<0.001
	Walked	519 (6.96)	21 (1.3)	<0.001
	Other	75 (1.01)	6 (0.37)	0.009

The severely injured validation dataset had a significantly higher overall proportion of penetrating injuries (60.6% vs. 39.2%, p<0.001), including a higher proportion of both firearm injuries (28.4% vs. 13.5%, p<0.001) and stab injuries (30.6% vs. 23.4%, p<0.001), but not bite injuries. Intentional injuries were significantly higher in the seriously injured validation cohort (70.6% vs. 49.7%, p<0.001). Severely injured patients were more likely to be transferred by ambulance (92.1% vs. 56.9%, p <0.001), whereas in the derivation dataset patients were more likely to arrive escorted by the

police (1.1% vs. 0.3%, p = 0.003), by a private vehicle (21.7% vs. 5.6%, p<0.0001) or by walking (6.9% vs. 1.3%, p<0.001).

Table 9.6 A comparison of the patient characteristics describing the physiological injury sustained by the Trauma Surgery derivation and severely injured validation datasets

Physiological injury	Characteristics	Derivation N (%)	Severely injured N (%)	P-value
Hypothermic	Yes	1,189 (15.94)	409 (27.64)	
	No	4,801 (64.36)	1,071 (72.36)	<0.001
Pulse	Normal	4,442 (59.54)	873 (51.93)	<0.001
	Tachycardic	2,405 (32.24)	773 (45.98)	<0.001
	Bradycardic	133 (1.78)	35 (2.08)	0.39
Hypotensive	Yes	201 (2.69)	103 (5.82)	
	No	7,259 (97.31)	1,667 (94.18)	<0.001
Respiratory rate	<9	19 (0.25)	8 (0.45)	0.15
	9 - 30	6,343 (85.03)	1,594 (90.06)	<0.001
	>30	639 (8.57)	168 (9.49)	0.21
GCS	12-15	6,656 (89.22)	1,471 (83.11)	<0.001
	9-11	125 (1.68)	57 (3.22)	<0.001
	6-8	128 (1.72)	73 (4.12)	<0.001
	<6	551 (7.39)	169 (9.55)	<0.001
Airway	Yes	389 (5.21)	181 (10.23)	
	No	7,071 (94.79)	1,589 (89.77)	<0.001

The severely injured patients were more likely to be hypothermic (27.6% vs. 15.9%, p<0.001), tachycardic (45.9% vs. 32.4%, p<0.001), hypotensive (5.8% vs. 2.6%, p<0.001), have a GCS score of less than 12 (p<0.001), and require intubation (10.2% vs. 5.2%, p<0.001) on arrival. There were no statistically significant differences between the proportions of patients with abnormal respiratory rates on arrival. The derivation dataset had a higher proportion of patients in the highest GCS category: GCS 12-15 (89.22% vs 83.11%, p<0.0001). Every other GCS category had a significantly higher proportion of

patients in the severely injured cohort. A comparison of the rates of endpoints in the derivation dataset, as well as those defined as severely (ISS>8) or not severely (ISS<8) injured is presented in table 9.7.

Table 9.7 A comparison of the rates of endpoints in the Trauma Surgery derivation dataset by injury severity

Endpoint measure	Overall Derivation Rate (%)	95% CI	ISS>8 (%)	95% CI	ISS<9 Rate (%)	95% CI	P-value
In-hospital mortality	2.52	2.18 - 2.90	6.16	5.13 - 7.38	0.19	0.09 - 0.36	<0.001
In-hospital adverse event	5.2	4.72 - 5.73	13.22	11.64 - 14.79	1.09	0.81 - 1.39	<0.001
ICU admission	3.59	3.11 - 4.16	10.94	9.18 - 12.71	0.74	0.42 - 1.06	<0.001
LOS>30	7.94	7.34 - 8.57	12.94	11.37 - 14.50	3.24	2.73 - 3.74	<0.001

P-value derived after probability proportion test comparing rates of endpoints in the severely and non-severely injured patients*

The severely injured patients within the derivation dataset had significantly higher rates of all endpoints including in-hospital mortality (6.1% vs 0.2%, p<0.001), in-hospital adverse event (13.2% vs 119%, p<0.001), ICU admission (10.9% vs 0.7, p<0.001) and LOS of greater than 30 days (7.9% vs 3.2%, p<0.001) compared to the non-severely injured patients.

Comparative analysis: Derivation and operative validation datasets

There were 950 patients who required operative management during the study period. These patients were included in the operative validation dataset. Tables 9.8 – 9.11 present the comparative analysis of the derivation and operative validation datasets using data taken on admission which were >80% complete.

Table 9.8 Comparison of patient reserve characteristics between Trauma Surgery derivation and Operative validation datasets

Patient reserve	Category	Derivation N (%)	Operative N (%)	P-value
Gender	Male	5,788 (77.59)	887 (93.37)	
	Female	1,659 (22.24)	63 (6.63)	<0.001
Age	<18	662 (8.87)	105 (11.16)	0.021
	18 - 29	3,172 (42.52)	489 (51.97)	<0.001
	30 - 39	1,637 (21.94)	220 (23.38)	0.314
	40 - 49	884 (11.85)	87 (9.25)	0.018
	50 - 55	343 (4.6)	28 (2.98)	0.022
	65 - 69	149 (2.0)	3 (0.32)	<0.001
	>60	503 (6.74)	9 (0.96)	<0.001
Race	Black	3,411 (45.72)	383 (56.16)	<0.001
	Mixed ancestry	3,138 (42.07)	288 (42.23)	0.925
	Indian	135 (1.81)	5 (0.73)	0.015
	White	506 (6.79)	6 (0.88)	<0.001
	Other	54 (0.73)	0 (0)	0.008
Charlson's comorbidity index	0	7,328 (98.23)	944 (99.37)	
	1	132 (1.77)	6 (0.63)	0.009
Diabetic status	Diabetic	189 (2.53)	8 (0.84)	
	Non-diabetic	7,721 (97.47)	942 (99.16)	0.001
HIV status	HIV positive	214 (2.87)	26 (2.74)	
	Unknown	4,246 (97.13)	924 (97.26)	0.821
Smoking status	Current smoker	220 (2.95)	17 (1.79)	
	Non-smoker	7,240 (97.05)	933 (98.21)	0.042
Illicit drug use	Yes	1,829 (24.52)	255 (26.84)	
	No	5,631 (75.48)	695 (73.16)	0.119

Patients included in the operative validation dataset were more likely to be male (93.3 vs. 77.6%, p<0.001) and were younger than the patients in the derivation dataset. This was due to a significantly higher proportion of patients in both the <18 and 18-29 age category, as well as fewer patients in all age categories greater than 39 years of age. The operative cohort had a higher proportion of Black patients (56.2% vs. 45.7%, p<0.001) and a lower proportion of Indian (0.7% vs. 1.8%, p = 0.015) and White (0.8% vs. 6.7%, p<0.001) patients compared to the derivation dataset. The derivation dataset

had a higher proportion of diabetic patients (2.5% vs. 0.8%, p = 0.001) and non-smokers (97.1% vs. 98.2%, p = 0.042).

Table 9.9 Comparison of pre-hospital circumstances of the injury events between the Trauma Surgery derivation and Operative validation datasets

Prehospital circumstances	Category	Derivation N (%)	Operative N (%)	P-value
Complaint	Fall	970 (13.0)	4 (0.42)	<0.001
	Motor vehicle collision	1,033 (13.85)	55 (5.8)	<0.001
	Struck by or against	1,236 (16.57)	12 (1.26)	<0.001
	Bite	45 (0.6)	3 (0.32)	0.279
	Firearm	1,004 (13.46)	416 (43.84)	<0.001
	Stab	1,743 (23.36)	359 (37.83)	<0.001
	Other	1429 (19.16)	103 (10.85)	<0.001
Mechanism	Blunt	4,534 (60.78)	173 (18.21)	<0.001
	Penetrating	2,926 (39.22)	777 (81.79)	<0.001
Intentional	Yes	3,708 (49.71)	720 (80.45)	<0.001
	No	3,435 (46.05)	175 (19.55)	<0.001
Transport	Air	11 (0.15)	1 (0.12)	0.82
	Ambulance	4249 (56.96)	757 (92.88)	<0.001
	Police	80 (1.07)	0 (0)	0.001
	Vehicle	1622 (21.74)	49 (6.01)	<0.001
	Walked	519 (6.96)	4 (0.49)	<0.001
	Other	75 (1.01)	4 (0.49)	0.12

The operative validation dataset had a significantly higher overall proportion of penetrating injuries (81.7% vs. 39.22%, p<0.001) including a higher proportion of both firearm injuries (43.84% vs. 13.46%, p<0.001) and stab injuries (37.83% vs. 23.36%, p<0.001), but not bite injuries. All three categories of blunt complaints were proportionally higher in the derivation dataset (fall: 13.0% vs. 0.4%, p<0.001; motor vehicle collision: 13.8% vs. 5.8%, p<0.001; and struck by or against: 16.5% vs. 1.3%, p<0.001). Intentional injuries were significantly higher in the operative validation cohort (80.4% vs. 49.7%, p<0.001). Operatively managed patients were also more likely to be

transferred by ambulance (92.8% vs. 56.9%, p <0.001) whereas in the derivation dataset patients were more likely to arrive escorted by the police (1.1% vs. 0.0%, p = 0.01), by a private vehicle (21.7% vs. 6.1%, p<0.001) or by walking (6.9% vs. 0.5%, p<0.001).

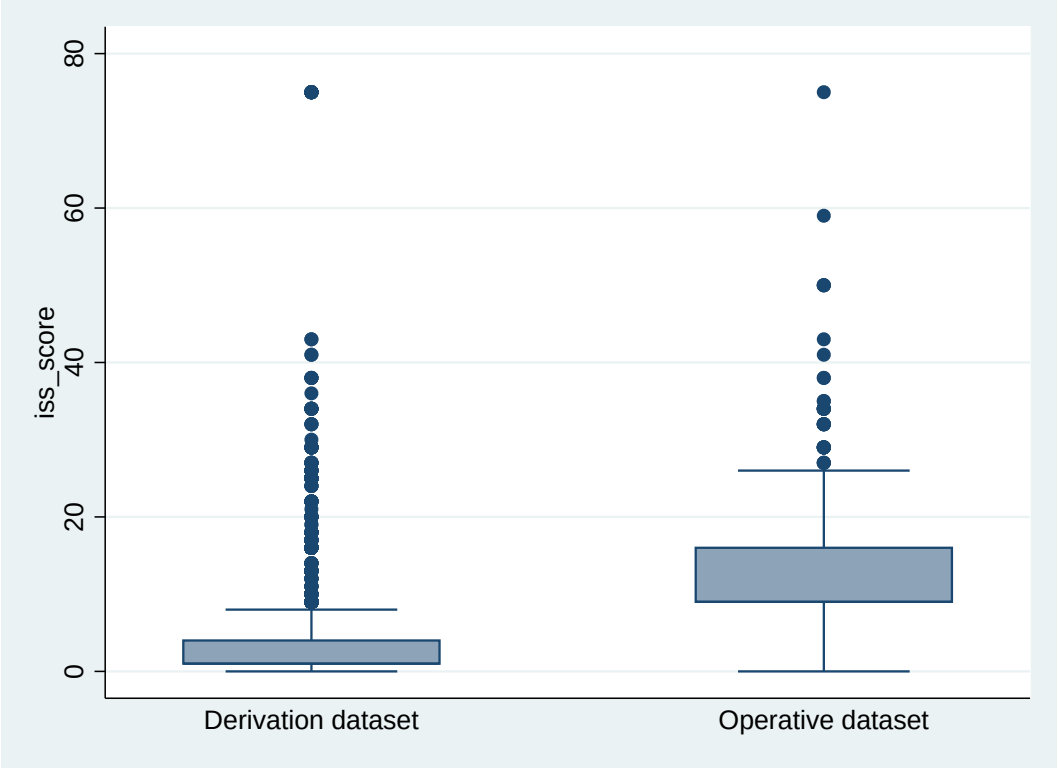
Table 9.10 A comparison of the patient characteristics describing the physiological injury sustained by the Trauma Surgery derivation and Operative validation datasets

Physiologic injury	Characteristics	Derivation N (%)	Operative N (%)	P-value
Hypothermic	Yes	1,189 (15.94)	244 (31.81)	
	No	4,801 (64.36)	523 (68.19)	<0.001
Pulse	Normal	4,442 (59.54)	395 (45.09)	<0.001
	Tachycardic	2,405 (32.24)	473 (54.0)	<0.001
	Bradycardic	133 (1.78)	8 (0.91)	0.05
Hypotensive	Yes	201 (2.69)	86 (9.05)	
	No	7,259 (97.31)	864 (90.95)	<0.001
Respiratory rate	<9	19 (0.25)	7 (0.74)	0.01
	9 - 30	6,343 (85.03)	817 (86.0)	0.429
	>30	639 (8.57)	126 (13.26)	<0.001
GCS	12-15	6,656 (89.22)	834 (87.79)	0.036
	9-11	125 (1.68)	19 (2.0)	0.473
	6-8	128 (1.72)	26 (2.74)	0.027
	<6	551 (7.39)	71 (7.47)	0.929
Airway	Yes	389 (5.21)	53 (5.58)	
	No	7,071 (94.79)	897 (94.42)	0.63

The operatively managed patients presented with a higher degree of physiological injury in that they were more likely to be hypothermic (31.8% vs. 15.9%, p<0.001), tachycardic (54.0% vs. 32.2%, p<0.001), hypotensive (9.1% vs. 2.7%, p<0.001) and tachypnoeic (13.2% vs. 8.5%, p<0.001) on arrival. The derivation data set had a higher proportion of patients in the highest GCS category (GCS 12-15: 89.2% vs. 87.9%, p = 0.036). The rates of ventilatory support on arrival were similar between the two datasets (5.2% vs. 5.5%, p = 0.63%).

In the derivation dataset, 6,501 (87.1%) patients had ISS coding completed on discharge. The median ISS was 5 (IQR 1-9). In the operative validation dataset, 834 (87.7%) patients had ISS coding completed on discharge. The median ISS was 9 (range 9 - 17). Figure 9.1 below compares the ISS of the two datasets.

Figure 9.1 Box plot comparing the Injury Severity Scores in the Trauma Surgery derivation and Operative validation datasets.



The Kruskal-Wallis rank sum test revealed a significant difference in the ranks of the ISS of the operatively and non-operatively managed patients in the derivation dataset ($p<0.001$). A comparison of the rates of endpoints in the derivation dataset, as well as the operatively and non-operatively managed patients is presented in table 9.11.

Table 9.11 Rates of endpoints in the Trauma Surgery derivation dataset, Operative validation dataset and non-operatively managed patients

End point measure	Derivation Rate (%)	95% CI	Operative Rate (%)	95% CI	Non-Operative Rate (%)	95% CI	P-value*
In-hospital mortality	2.52	2.18 - 2.90	2.53	1.68 - 3.77	2.52	2.17 - 2.93	0.99
In-hospital adverse event	5.2	4.72 - 5.73	19.08	16.66 - 21.76	3.27	2.86 - 3.73	<0.001
ICU admission	3.59	3.11 - 4.16	15.93	13.29 - 18.98	1.72	1.36 - 2.16	<0.001
LOS>30	7.94	7.34 - 8.57	18.31	15.93 - 20.96	6.49	5.92 - 7.11	<0.001

P-value derived after proportion test comparing operative and non-operative endpoint rates*

The operatively managed patients within the derivation dataset had significantly higher rates of in-hospital adverse events (19.1% vs. 3.3%, p<0.001), ICU admission (15.9% vs. 1.7%, p<0.001) and LOS of greater than 30 days (18.3% vs. 6.5%, p<0.001) compared to the non-operatively managed patients in the derivation dataset. There was no difference in rates of the primary endpoint of in-hospital mortality between operatively and non-operatively managed patients (p = 0.99).

b. Identification of predictors

The univariate association between the categorical variables collected on admission in the derivation dataset and the endpoints for prediction are presented in Table 9.12. Significance was set at p<0.1 and significant associations are highlighted in red.

Table 9.12 The univariate association between categorical predictors on admission and the outcomes for prediction in the derivation dataset

	In-hospital mortality	In-hospital adverse events	ICU admission	LOS>30
Patient reserve characteristics				
Gender	<0.0001	<0.0001	<0.0001	0.25
Age	0.05	0.344	0.352	0.22
Race	0.08	0.007	0.251	0.32
Charlson's comorbidity index	0.612	0.596	0.468	0.21
Diabetic status	0.291	0.227	0.164	0.14
HIV status	0.45	0.097	0.553	0.43
Smoking status	0.15	0.09	0.199	0.58
Illicit drug use	0.03	0.038	0.011	0.14
Prehospital circumstance characteristics				
Complaint	0.0001	<0.0001	<0.0001	0.007
Mechanism	0.114	<0.0001	0.001	0.034
Reason	0.005	0.11	0.143	0.028
Intention	0.007	0.054	0.032	0.43
Referral	0.003	<0.0001	0.04	0.27
Transport	<0.0001	<0.0001	<0.0001	<0.0001
Physiological injury characteristics			<0.0001	
Triage	<0.0001	<0.0001	<0.0001	<0.0001
Hypothermia	<0.0001	<0.0001	<0.0001	<0.0001
Pulse rate	<0.0001	<0.0001	<0.0001	<0.0001
Hypotension	<0.0001	<0.0001	<0.0001	<0.0001
Respiratory rate	<0.0001	<0.0001	0.06	0.004
GCS	<0.0001	<0.0001	<0.0001	<0.0001
Level of consciousness	<0.0001	<0.0001	<0.0001	<0.0001
Acidotic	<0.0001	<0.0001	0.002	<0.0001
Anaemia	0.003	<0.0001	0.022	0.003
Hyperlactataemia	<0.0001	<0.0001	0.008	<0.0001
Airway	<0.0001	<0.0001	<0.0001	<0.0001
Anatomical injury scoring				<0.0001
ISS	<0.0001	<0.0001	<0.0001	<0.0001
Number of serious injuries	<0.0001	<0.0001	<0.0001	<0.0001

Twenty two of the twenty seven collected categorical predictors (81.48%) had a significant univariate association to the primary outcome measure of in-hospital death at the 0.1 significance level. There was a degree of agreement between an association with the primary outcome measure and the other endpoint measures. Some exceptions were

that age had a univariate association with in-hospital death ($p = 0.05$) but not the other outcome measures. Race was associated with in-hospital mortality ($p = 0.08$) and in-hospital adverse event ($p = 0.007$) but was not associated with ICU admission ($p = 0.25$) or LOS>30 ($p = 0.32$). HIV and smoking status were only associated with in-hospital adverse event but no other endpoint measure. Finally mechanism of injury (blunt/ penetrating) was not associated with an in-hospital death ($p = 0.114$) but was associated with an in-hospital adverse event ($P<0.0001$), ICU admission ($p = 0.001$) and LOS>30 days ($p = 0.034$). Overall, there were stronger associations with the primary outcome and both physiological injury characteristics and anatomical injury scoring, than with patient reserve or pre-hospital circumstance characteristics.

Categorical predictors which were greater than 80% complete, and had a univariate association with the primary outcome for prediction at the $p<0.1$ significance level, were then considered for inclusion in the multivariate models. These were the predictors considered: gender, age, race, illicit drug use, complaint, reason, intention, referral, transport, triage, hypothermia, pulse rate, hypotension, respiratory rate, GCS, level of consciousness and airway. Additionally, both the derived ISS and number of serious injuries were considered for the injury severity prediction rule.

9.3 A triage prediction rule for GSH Trauma Surgery

The seventeen variables were considered individually for inclusion in the triage prediction rule. The final multivariate model with the best performance in predicting an in-hospital death, using the Trauma Surgery derivation dataset, which had been collected on admission, is shown in Table 9.13.

Table 9.13 Multivariate model predicting in-hospital death in the Trauma Surgery derivation dataset using data collected on admission

Predictor (reference level)	Odds ratio	95% CI	Co-efficient	95% CI	P-value
Airway	2.79	1.54 – 2.76	1.02	0.51 - 1.53	<0.001
Hypotension	3.73	2.76 – 4.21	1.31	0.76 - 1.86	<0.001
Triage red	22.07	12.09 – 45.6	3.09	2.21 - 3.97	<0.001
GCS motor (GCS 6)	0.47	0.32 – 0.56	-0.76	-0.88 - -0.63	<0.001
Age >54	2.94	2.12 – 3.76	1.07	0.29 - 1.86	0.007
Constant			-2.52	-3.62 - -1.42	<0.0001

The final five variable model included, four binary independent predictors of in-hospital death and one ordinal independent predictor (GCS motor). These co-efficients were the log odds for an in-hospital death, and therefore, by exponentiating the co-efficients we were able to derive the odds ratios. A patient requiring intubation on or prior to admission had a 2.79 fold increased odds of an in-hospital death ($p < 0.001$). The presence of hypotension on admission (Systolic BP<90mmHg) was associated with a 3.73 fold increased odds of in-hospital death. A red triage colour (defined by a life threatening injury according to the admitting physician) was associated with a 22.07 fold increased odds of in-hospital death ($p<0.001$). Each increase in the GCS motor score was associated with a 0.47 fold decreased odds of in-hospital death ($p<0.001$). Finally, age on admission of greater than 54 years was associated with a 2.94 fold increased odds of in-hospital death ($p = 0.007$). The final model had a ROC of 0.9668 (95% CI 0.954 – 0.979) and a GOF statistic of 1.96 ($p = 0.375$). Notably, the final model only included patient reserve and physiological injury characteristics. The only pre-hospital circumstance characteristic, which was an independent predictor in the multivariate model was referral status. Patients transferred directly from the scene had a 1.53 fold

increased risk of in-hospital death compared to those patients referred from a referral hospital or from home (p = 0.044). However, addition of the referral status did not add any value to the overall discrimination and calibration of the model and was therefore not included. The GSH Trauma Surgery triage prediction rule was developed for application on admission to predict an in-hospital death. This was based on the following formula:

log(P/(1-P)) = -2.528225 + 1.025246*(Airway status) +1.316493*(Presence of hypotension) + 3.094101*(Assessed as life threatening injury by admitting physician) -0.7618551*(GCS motor score) + 1.079287*(Age>54 years).

9.4 GSH Trauma Score (GSTS)

The beta-co-efficients of the GSH Trauma Surgery triage prediction rule were then divided and rounded to the nearest integer to create a scoring system ranging from 1 to 12, called the GSH Trauma Score (GSTS). The scoring rubric for GSTS is shown in Table 9.14.

Table 9.14 Components of the Groote Schuur Trauma Score

Component	Score
Age	
<55	1
≥55	0
Trauma triage colour	
Green/Yellow	3
Red	0
Intubation status	
No	1
Yes	0
Systolic blood pressure (mmHg)	
≥90	1
<90	0
GCS motor	
None	1
Extension to pain	2
Flexion to pain	3
Withdraws from pain	4
Localises pain	5
Obeys commands	6
Total score	1-12

For ease of application, the GCS motor score was included in the score in its original ordinal scale. This had no impact on the performance of the score in the derivation dataset compared to when using a co-efficient-derived integer for each category of the GCS motor score. For this reason, the GSTS was developed so that an increasing score indicated a less severe injury. The GSTS had an ROC to predict in-hospital death of 0.9668 (95% CI 0.95419 - 0.97950), which was not different to that of the original GSH triage prediction rule from which it was derived (p = 0.2241). The GOF of the GSTS was 5.63 (p = 0.0599). Table 9.15 presents the number of deaths and mortality rate per GSTS in the derivation dataset.

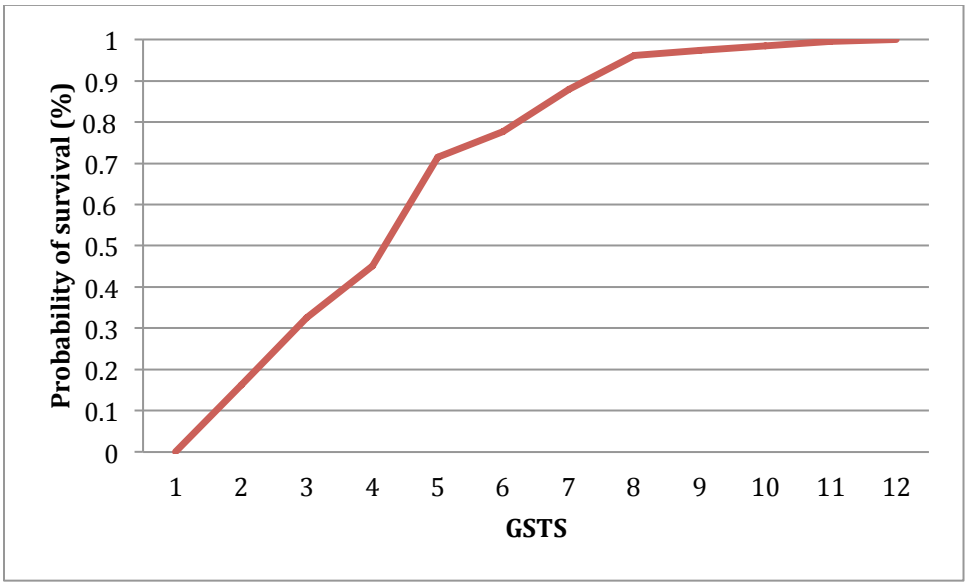
Table 9.15 The Mortality rate by Groote Schuur Trauma Score in the Trauma Surgery derivation dataset

GSTS	Frequency (%)	Deaths	Mortality rate (%)
1	2	2	100
2	43	36	83.72
3	80	54	67.51
4	31	17	54.83
5	35	10	28.57
6	67	15	22.388
7	91	11	12.08
8	205	8	3.90
9	39	1	2.56
10	1,052	16	1.52
11	731	3	0.41
12	4,559	2	0.043

In the derivation dataset the median GSTS was 12 (IQR 11-12) and mean was 10.94 (range 1-12). A GSTS of less than 8 had specificity of 98.34% and sensitivity of

72.87% to predict an in-hospital death. An increase in GSTS was associated with a stepwise increase in survival following trauma at GSH as shown in Figure 9.2.

Figure 9.2 Probability of survival following major trauma as predicted by the Groote Schuur Trauma Score (GSTS)



9.5 An Injury severity prediction rule for GSH Trauma Surgery

The GSH Trauma Surgery triage prediction rule and its subsequent scoring system, took into account patient reserve and physiological injury characteristics only, and no anatomical description of the injury sustained. The addition of the number of serious injuries did not add to the triage prediction rule, and the number of serious injuries was not found to be an independent predictor of in-hospital death. The addition of anatomical injury descriptions using the derived ISS scores, however, resulted in the best overall performance to predict an in-hospital death. The final multivariate model, which delivered the best overall results in predicting in-hospital death in the Trauma

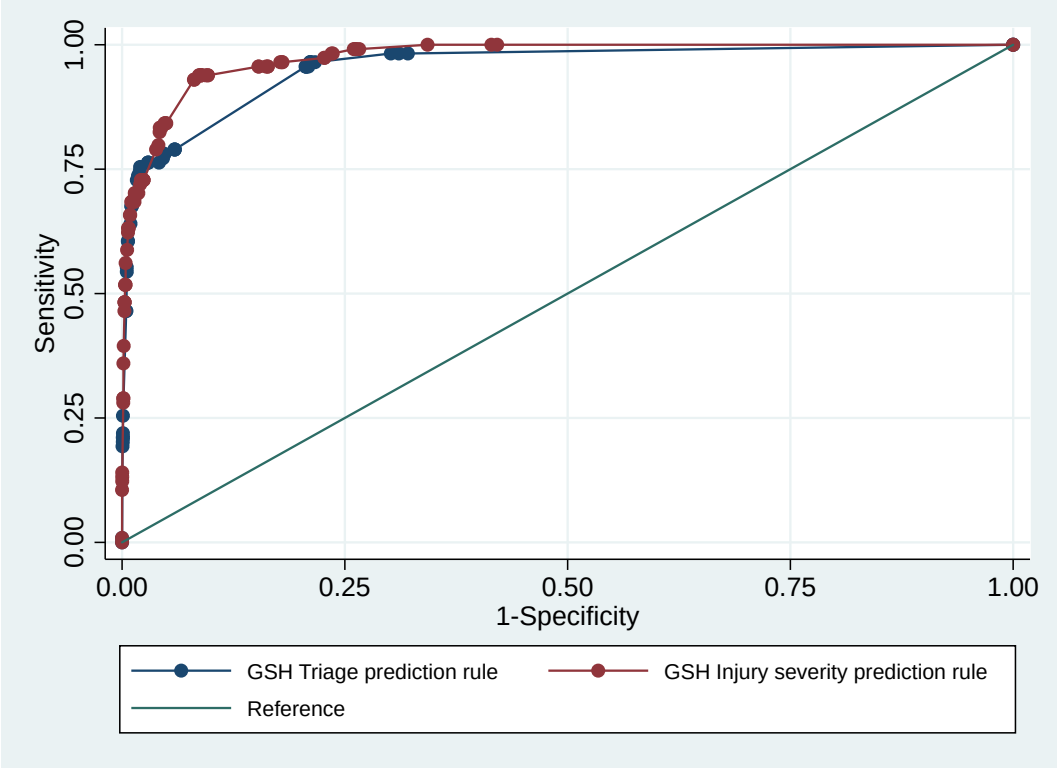
Surgery derivation dataset (using all the available data collected) is presented in Table 9.16.

Table 9.16 Multivariate model predicting in-hospital death in the Trauma Surgery derivation dataset including detailed anatomical description data

In-hospital death Model	Co-efficient	P-value	95% CI
Hypotension	1.23	<0.0001	0.61- 1.85
Triage red	2.64	<0.0001	1.66 -3.62
GCS motor score	-0.81	<0.0001	-0.93 - -0.69
Age > 54	1.08	0.022	0.15 - 2.01
ISS score	0.05	<0.0001	0.03 -0.07
Constant	-2.56	<0.0001	-3.72 - -1.41

The airway status was no longer a significant predictor of in-hospital death with the addition of ISS and was therefore excluded. Every one point increase in ISS was associated with a 1.06 fold increased odds of in-hospital death ($p<0.0001$). The addition of an anatomical description did improve overall performance of the model, compared to the triage prediction rule presented in figure 9.3.

Figure 9.3 Comparison of discriminatory performance by GSH Triage and Injury Severity Prediction Rules



p=0.0033

In the cohort with complete ISS scoring, the ROC for Injury Severity Prediction Rule was 0.9711 (95% CI 0.96017–0.981696) compared to the ROC for the Triage Prediction Rule of 0.9552 (95% CI 0.9364–0.97392). This difference was statistically significant (p = 0.0033). The calibration of the Injury Severity model was also better demonstrating a higher GOF statistic of 4.46 (p = 0.6150).

Therefore the GSH Trauma Surgery Injury Severity prediction rule to predict an in-hospital death was based on the following formula:

log(P/(1-P)) = -2.568611+1.233606*(Presence of hypotension) + 2.644827*(Assessed as life threatening injury by admitting physician) -0.8163333*(GCS motor score) + 1.084836*(Age>54 years)+0.0573972*(ISS).

a. Validation of the GSH Trauma Prediction rule

The detailed results of the validation of the GSH Trauma Prediction rule can be found in Appendix 4. Below are the summary findings of the head-to-head analyses with the GSH derived prediction rules and the following validated trauma-scoring systems: Revised Trauma Score (RTS), Kampala Trauma Score (KTS) and the Trauma and Injury Severity Score (TRISS).

Validation against the Revised Trauma Score

This comparative validation step showed that the data burden for the two scores were not different but the performance of the GSTS was superior with regards to both discrimination and calibration in all three datasets.

Validation against the Kampala Trauma Score

This comparative validation step showed that the data burden for the GSTS was significantly less ($p<0.001$) than the KTS. The GSTS discriminatory ability was superior in all datasets apart from the operative validation dataset where the ROC's were no different from each other ($p = 0.1566$). The GSTS ability to calibrate was superior in the operative validation and the derivation datasets, and both scores had adequate calibration in the severely injured cohort.

Validation against the Trauma Injury Severity Score

This comparative validation step showed that the data burden for the GSH injury-severity prediction was significantly less than for the TRISS prediction, in all three datasets. The discriminatory ability for both prediction rules were no different in all datasets, and the inclusion of an anatomical description in the rule improved the ROC's

compared to the previous scores analysed. The ability of the prediction rules to calibrate was comparable, and only performed adequately in the severely injured cohort.

In each validation step, the a priori degree of performance stated in the hypothesis was achieved. Clinician-entered data using eTHR can be used to predict in-hospital mortality following major trauma with adequate precision (c statistic >0.7) and calibration (p>0.05).

b. Risk-adjusted benchmarking analysis

Finally, using the co-efficient based scores associated with RTS and TRISS, a risk-adjusted comparison can be made benchmarking the outcomes of patients in the MTOS and those managed at GSH. The results of this analysis are shown in Table 9.17.

Table 9.17 Using co-efficient based scores to compare outcomes following trauma managed at GSH to the MTOS

	Observed deaths with complete data for the score	Expected deaths based on the score	O/E Ratio (95% CI)	P-value
Derivation cohort				
RTS	183	333	0.55 (0.47 - 0.64)	<0.001
TRISS	115	114	1.01 (0.83 - 1.21)	NS
Severely injured cohort				
RTS	86	65	1.31 (1.05 - 1.62)	<0.05
TRISS	82	54	1.52 (1.21 - 1.88)	<0.001
Calibration in operatively managed				
RTS	24	42	0.57 (0.36 - 0.85)	<0.001
TRISS	20	23	0.85 (0.52 - 1.32)	NS

There was no difference in outcomes for the whole cohort and operatively managed patients between GSH and what would be expected from the MTOS, according to TRISS. According to both the TRISS and RTS, patients who were severely-injured

were more likely to die if managed at GSH than as a participant of the MTOS. However, according to the RTS, patients who required an operation were almost twice as likely to survive being managed at GSH compared to being a participant of the MTOS.

Chapter 10

Validating the ISS coding at Groote Schuur Hospital:

Man versus m-Health technology

Using eTHR and clinician entered data to generate ISS was a novel application and therefore had to be validated.

10.1 Materials and Methods

a. AIS coding at Groote Schuur Hospital (ISS eTHR)

A random sample of 60 patients who underwent an operation was selected from the trauma derivation dataset. Only patients who underwent an operation were chosen as a starting point because a detailed anatomical description of injuries in operatively managed patients was thought to be more valuable. The admission records, operative records, radiology reports and discharge summaries of these randomly selected 60 patients were then retrospectively extracted, and uploaded into a database generated for the purpose of this study using REDCap (Research Electronic Data Capture).

b. AIS coding at The Vancouver General Hospital (ISS VGH)

The Vancouver General Hospital (VGH), in Canada, has an accredited adult Level 1 Trauma Center and has a working collaboration with GSH. Three data analysts from VGH, each with trauma coding experience of over 5 years and all certified trained in coding AIS 2005 and Update 2008, were given access to our REDCap database. They were asked to code the ISS for each patient according to the Association for the Advancement of Automotive Medicine (AAAM) guidelines of coding conservatively, and not coding queried or unconfirmed diagnoses. Each injury was scored and the ISS

was then manually calculated and entered into REDCap. Coders were blinded to the ISS eTHR scores. After all patients were coded, the ISS VGH scores were extracted and the REDCap database was closed. The inter-rater agreement of ISS eTHR and ISS VGH scores were then compared.

c. Analysis

A descriptive analysis of the patients who were included in the ISS validation step was performed. The resultant ISS data for both ISS GSH and ISS VGH were reasonably normal and therefore both parametric and non-parametric tests were described to comprehensively illustrate the difference between the two sets of ISS scores.

Measure of agreement

Inter-rater agreement was assessed using the Bland-Altman limits of agreement (LoA) method (183). This method compared the estimated variation in the data to a clinical evaluation of what was an acceptable variation for measurements to be considered 'not different.' The inter-rater agreement between VGH and GSH was further assessed by calculating the kappa statistic of the ISS grouped into validated ordinal categories (170). The ISS categories were 0-8, 9-15, 16-24 and 25-75. A kappa statistic was interpreted according to the following: 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement and 0.81-0.99 almost perfect agreement (184). The purpose of the ISS is to rank order severity of injuries for prognostication. For this reason, the Kruskal-Wallis Rank test was also employed to compare the ranks of the ISS coded by GSH and VGH.

Measures of reliability

Reliability was defined as "the ratio of the variation between measurements to the total variation of all the measurements it is intended to measure". Reliability was estimated by intraclass correlation co-efficient (ICC) statistics and corresponding 95% CI, using a two-way mixed model with absolute agreement index (185). This was repeated for the two categorical ISS scores. ICC statistics give a number on a scale from 0 to 1, where 0 indicates agreement no better than chance, and 1 indicates perfect agreement.

ISS as a predictor

In order to compare the performance of both scores, the univariate association between ISS and the presence or absence of an adverse in-hospital event was tested, using logistic regression for each set of ISS. The two resulting estimates and their confidence intervals were calculated. The calibration and discrimination of these two regression models were then compared.

10.2 Results

a. Descriptive statistics

The ISS of 57 patients were used in this validation analysis; 3 patients were excluded due to inadequate data. The mean age of the cohort was 27.2 years (range 14-62) and 96.5% were male. The mechanism of injury was penetrating in 93.4% of the patients, of which 52.8% were GSWs. The operative cohort included 33 exploratory laparotomies, 15 cardiothoracic procedures, six vascular procedures, two videoscopy procedures, and one neck dissection. An in-hospital complication occurred in 25.2% (CI 17.76 – 33.98), including a mortality rate of 4.5% (CI 2.71 – 7.57). The summary

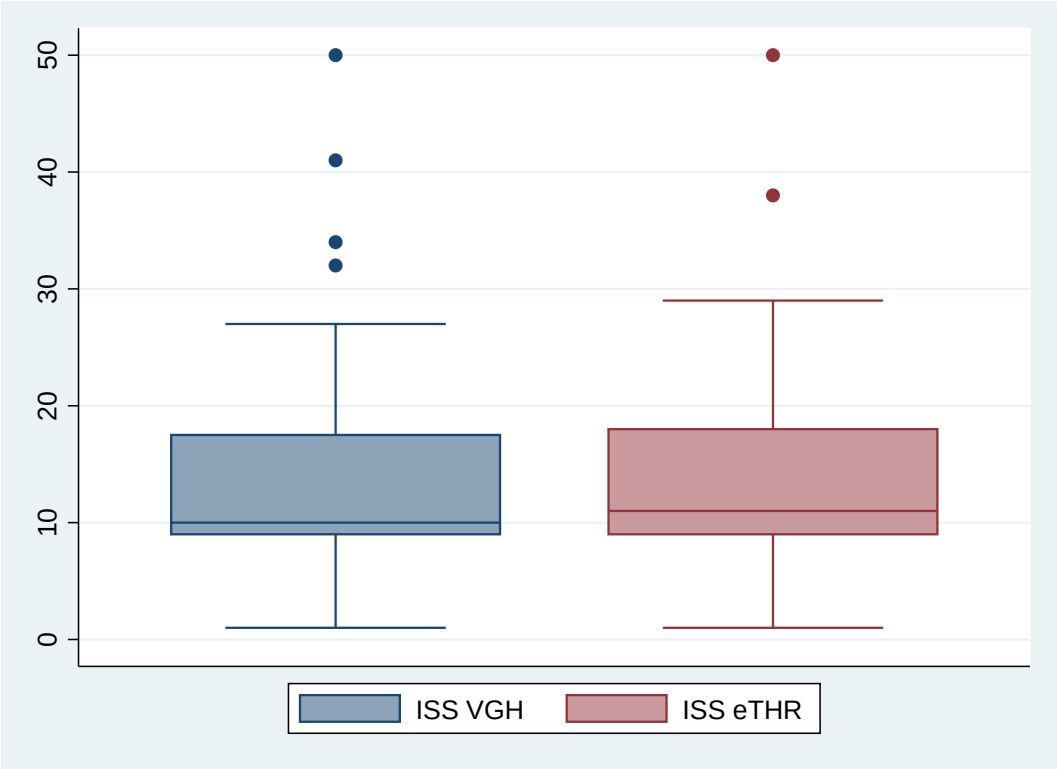
descriptive statistics including that of the ISS are presented in Table 10.1.

Table 10.1 Summary statistics of the 57 patients used for Injury Severity Scoring validation

Characteristic	Measurement	Statistic
Gender	Female N (%)	2 (3.5)
	Male N (%)	55 (96.5)
Age	Mean (Range)	27.24 (14-62)
Mechanism	Blunt N (%)	4 (7.0)
	Penetrating N (%)	53 (93.0)
Penetrating type	GSW N (%)	28 (52.8)
	Knife N (%)	25 (47.2)
	Other N (%)	0 (0)
Operation	Exploratory laparotomies N (%)	33 (57.9)
	Cardiothoracic procedures N (%)	15 (26.3)
	Vascular dissections N (%)	6 (10.5)
	Videoscopy procedures N (%)	2(3.5)
	Neck dissections N (%)	1 (1.8)
ISS VGH	Median (IQR)	10 (9 - 17)
	Mean (95% CI)	14.92 (12.38 - 17.48)
	Variance (Sd)	90.87 (9.53)
ISS eTHR	Median (IQR)	11 (9 - 18)
	Mean (95% CI)	14.51 (12.13 - 16.89)
	Variance (Sd)	78.98 (8.89)

The median ISS VGH was 10 (IQR 9-17.5) and the median ISS eTHR was 11 (IQR 9 – 18). The mean ISS VGH was 14.9 (95% CI 12.38 – 17.48) compared to the ISS eTHR mean of 14.5 (95% CI 12.14 – 16. 89). The variance of the ISS VGH was 90.8 compared to a variance of 78.9 from the ISS eTHR. Figure 10.1 represents the box plots of both ISS VGH and ISS eTHR.

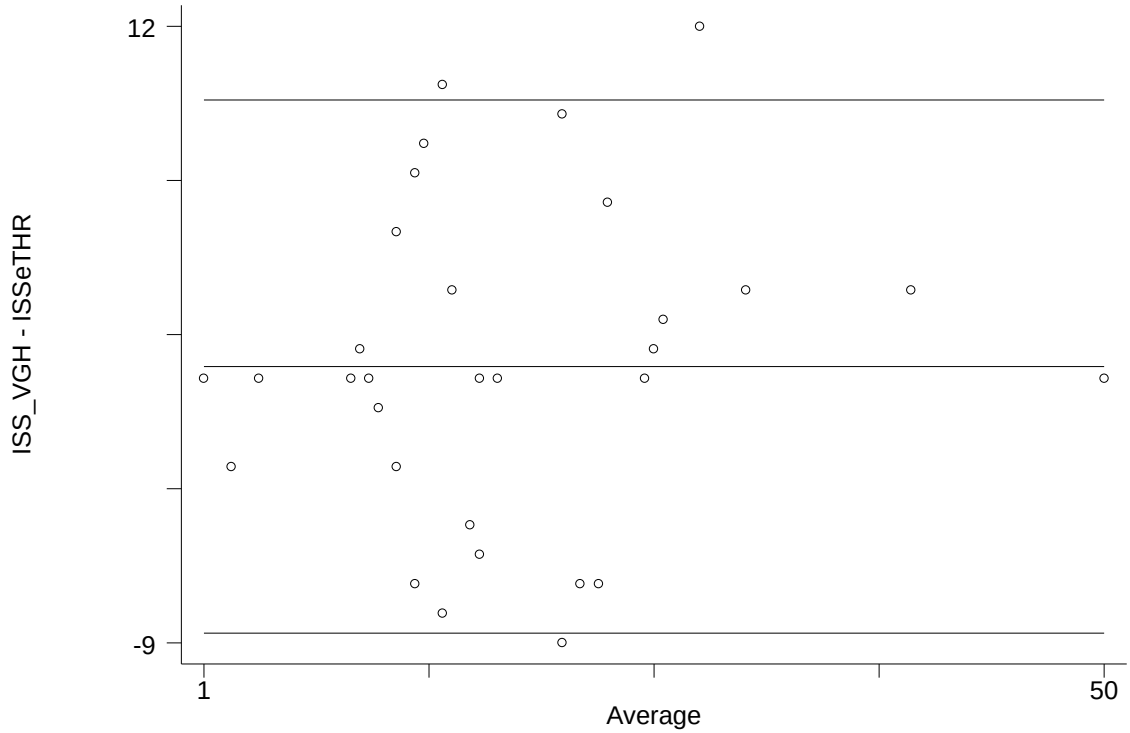
Figure 10.1 Box plots comparing ISS VGH and ISS eTHR



b. Agreement

The Bland-Altman comparison of ISS eTHR and VGH showed that the limits of agreement were within -8.669 to 9.490. The mean ISS difference was 0.411, which was not statistically significant (95% CI -0.805 – 1.626). The Pitman's Test for the difference in variance was r=0.147 (p = 0.281). Figure 10.2 represents the resultant Bland-Altman plot. The kappa statistic was 0.53 (95% CI 0.49 – 0.61; p<0.0001), which implies a moderate agreement.

Figure 10.2. Bland-Altman plot comparing rater agreements between coders and eTHR



The Kruskal-Wallis rank test resulted in a rank sum of 3,270.50 in the ISS VGH compared to a rank sum of 3,057.50 in the ISS eTHR. The difference in ranks was not statistically significant in the two groups (p = 0.54).

c. Reliability

The ICC of the individual ISS was 0.8797 (95% CI 0.8034 – 0.9276), which equates to excellent reliability between ISS VGH and eTHR. The ICC of the categorical ISS was 0.8016 (95%CI 0.6833 – 0.8788).

d. ISS as a predictor

The risk estimates when using both individual and categorical ISS for VGH and eTHR are shown in Table 10.2. The odds ratios estimated using individual ISS were very

comparable. The OR using ISS VGH was 1.24 (95% CI 1.10 – 1.39) compared to an OR of 1.19 (95% CI 1.07 – 1.34) when using ISS eTHR in the logistic regression models predicting an in-hospital complication.

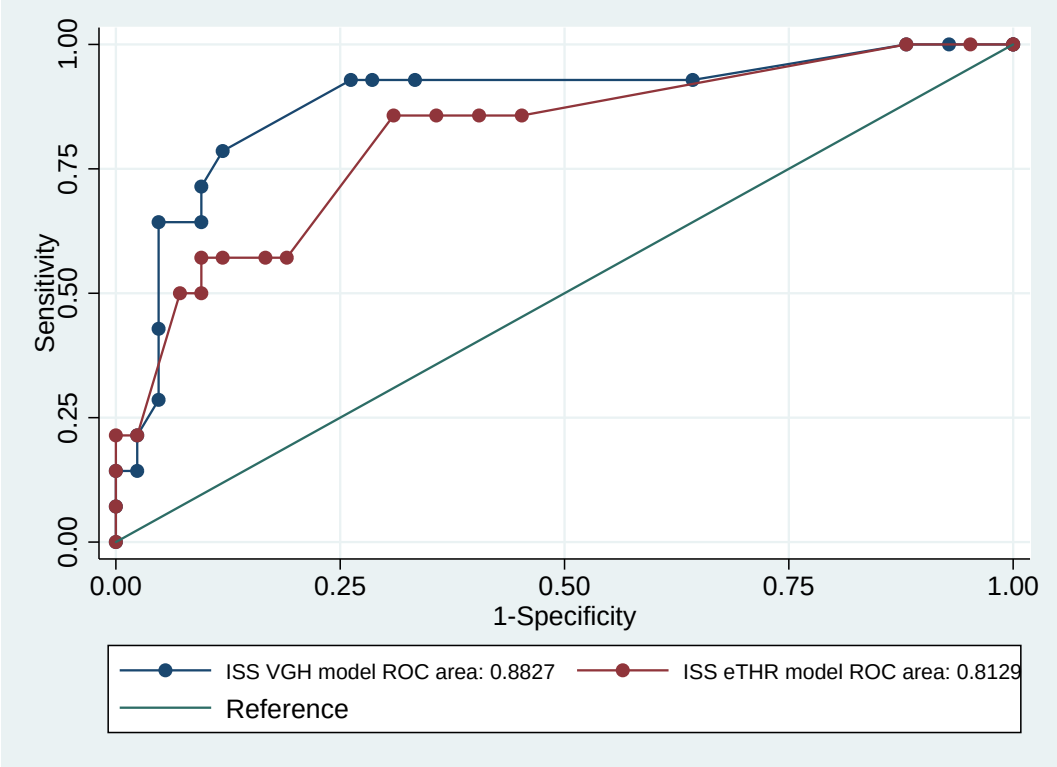
Table 10.2 Comparison of the univariate models predicting an adverse outcome using ISS eTHR and ISS VGH as predictors

Method of coding	Odds ratio	95% CI	ROC*	95% CI	GOF P-value^
Individual ISS VGH	1.24	1.10 - 1.39	0.8827	0.7677 - 0.9976	0.21
Individual ISS eTHR	1.19	1.07 - 1.34	0.8129	0.6802 - 0.9456	0.67
Categorical ISS VGH	5.29	2.01 - 13.98	0.8172	0.6947 - 0.9395	0.91
Categorical ISS eTHR	10.85	3.13 - 37.69	0.8912	0.7967 - 0.9855	0.84

ROC Area under Receiver Operating Curve. GOF P-Value^ Hosmer Lemeshow Goodness-of-Fit test*

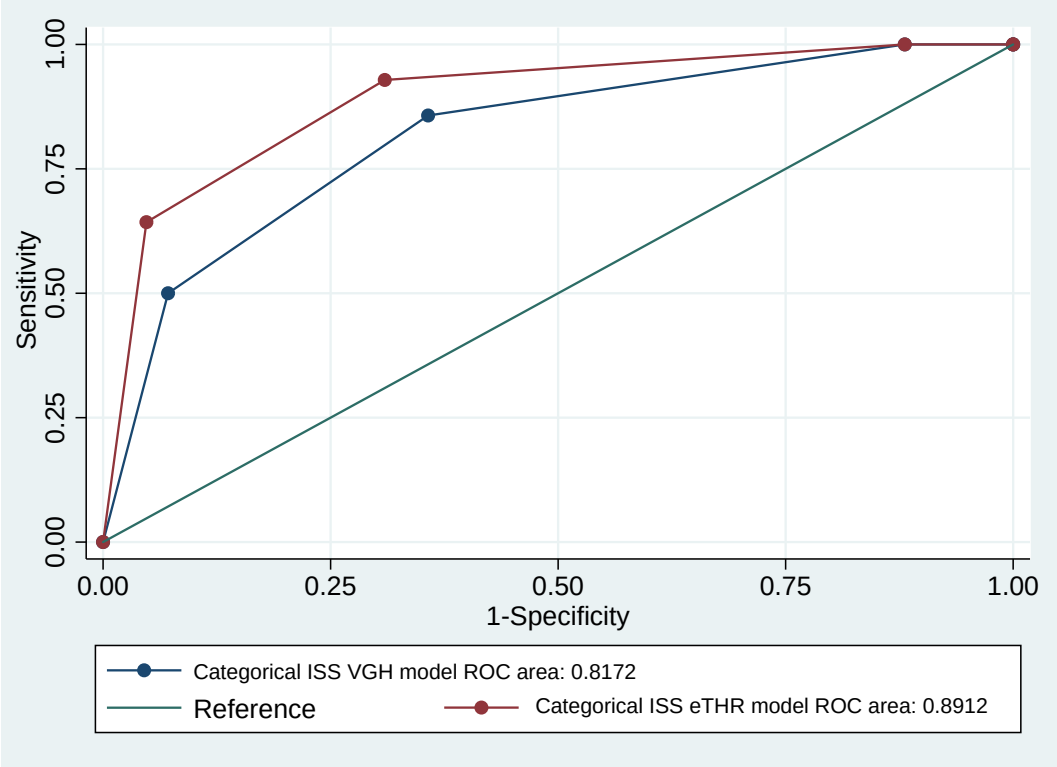
Both of these models had outstanding predictability with an ROC using ISS VGH of 0.8827 (95% CI 0.7677 – 0.9976) compared to an ROC of 0.8129 (95% CI 0.6802 – 0.9456), which was not significantly different (p = 0.33). These curves are plotted against each other in Figure 10.3. Both models had acceptable ability to calibrate in-hospital adverse events; GOF statistics were non-significant for both models (p = 0.21 and p = 0.67, respectively).

Figure 10.3 Comparison of coder and eTHR continuous models to discriminate adverse events



The models using categorical ISS had much larger standard errors and these estimates were less precise as shown in Table 10.2. Using categories for ISS VGH in the logistic regression resulted in an OR of 5.29 (95% CI 2.01 – 13.98) compared to when using categories of ISS eTHR resulted in an OR of 10.85 (95% CI 3.13 – 37.69). However, both of these models maintained excellent ability to predict with an ROC using categorical ISS VGH of 0.8172 (95% CI 0.6947 – 0.9395) compared to an ROC of 0.8912 (95% CI 0.7967 – 0.9855), which was not significantly different (p = 0.23). These curves are plotted against each other in Figure 10.4. Both models had acceptable ability to discriminate in-hospital adverse events; GOF statistics were non-significant for both models (p = 0.91 and p = 0.84, respectively).

Figure 10.4 Comparison of coder and eTHR categorical models to discriminate adverse events

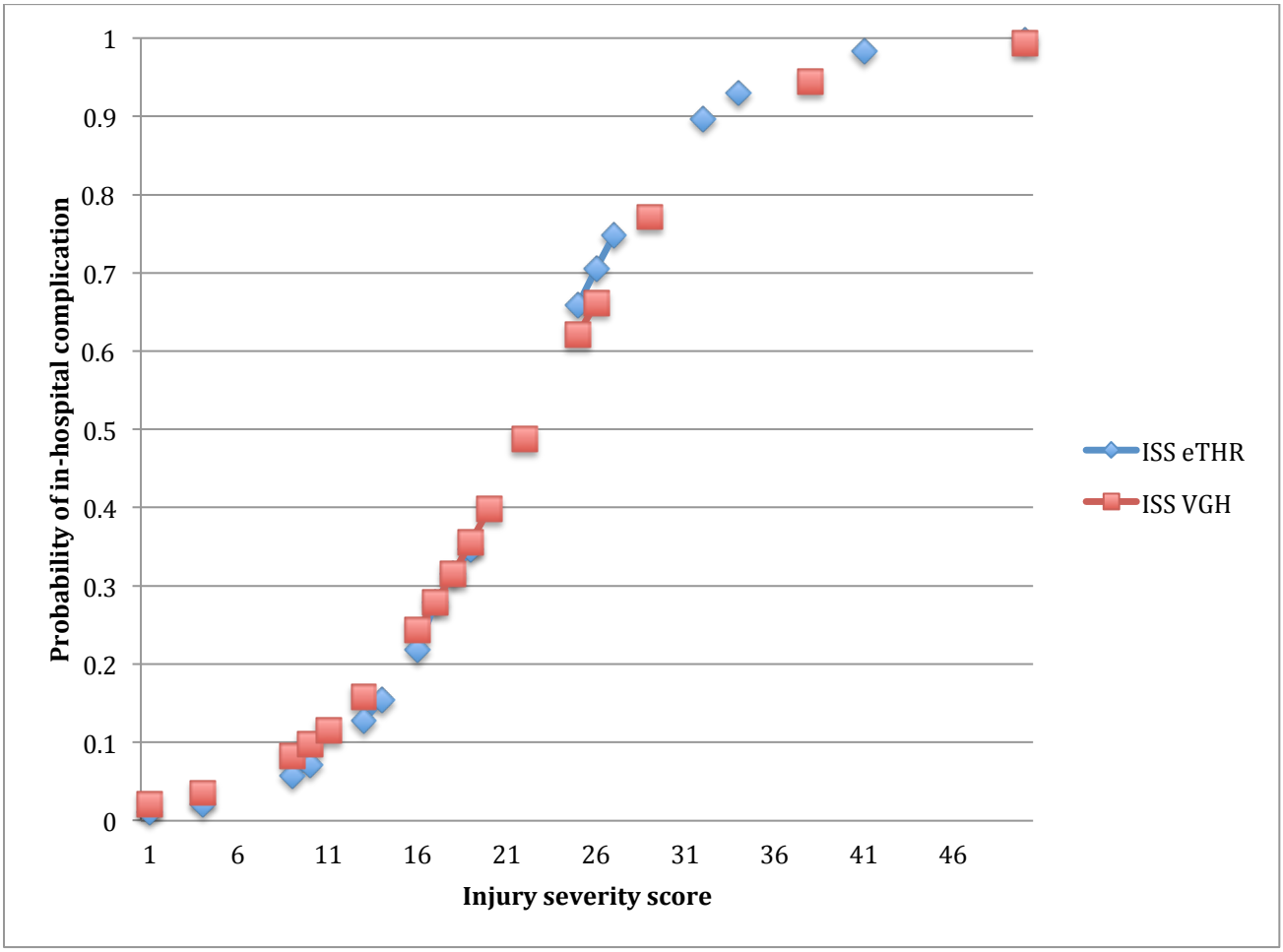


The result of the Hosmer-Lemeshow goodness-of-fit using quintiles of 10 for grouping the data of both sets of individual ISS is presented in table 10.3 and graphically in Figure 10.5. The probability of adverse events are comparable at similar VGH and eTHR ISS.

Table 10.3 Estimated probabilities of in-hospital adverse events at increasing ISS VGH and ISS eTHR scores

ISS eTHR	Probability of in-hospital complication	ISS VGH	Probability of in-hospital complication
1	0.011	1	0.021
4	0.0208	4	0.0355
9	0.0584	9	0.0834
10	0.0714	10	0.0982
13	0.1278	11	0.1154
14	0.1537	13	0.1578
16	0.2182	16	0.2436
17	0.2784	17	0.2784
18	0.3161	18	0.3161
19	0.3472	19	0.3563
25	0.6587	20	0.3988
26	0.7052	22	0.4877
27	0.7478	25	0.6207
32	0.8967	26	0.6622
34	0.9303	29	0.7712
41	0.9836	38	0.9448
50	0.9976	50	0.9934

Figure 10.5 Estimated probabilities of in-hospital adverse events at increasing ISS VGH and ISS eTHR scores



This study showed that ISS calculated by eTHR and gold standard coding were comparable and further supports the hypothesis that emerging m-Health technology provides a solution to mitigate the lack of reliable surgical outcomes research and surgical quality improvement programs in LMICs.

Chapter 11

Summary findings

11.1 General surgery

Two general surgery databases were developed de novo during the study period. An Essentials database modeled around the ACS-NSQIP was developed for general and vascular surgery patients at GSH. This included 130 variables for each of the 373 patients in the cohort. A second multi-centre Procedure-targeted database, describing the outcomes of 320 emergency exploratory laparotomies performed in the Metro West, was prospectively developed. Clinician-entered data and m-health technology were used. In the Essentials derivation dataset, a post-operative major complication occurred at a rate of 20.9%, in-hospital mortality at a rate of 7.7% and LOS>14 days at a rate of 22.7%. In the Procedure-targeted dataset, a post-operative major complication occurred at a rate of 14.1%, in-hospital mortality at a rate of 6.5% and LOS>14 days at a rate of 8.8%. Using these newly acquired databases, Essentials and Procedure-targeted prediction rules for major in-hospital complications and LOS greater than 14 days were derived. In order to demonstrate their geographical and methodological transportability, the derived prediction rules were applied to two separate databases from the ACS-NSQIP consortium including 320,816 and 41,633 patients, respectively. All four of the derived rules demonstrated excellent discriminatory ability, superior to the a priori level of 0.7 stated in the hypothesis. However, the ability to calibrate observed versus expected outcomes according to GOF techniques in the validation datasets ($p<0.05$) was poor. This disconnect in calibration was further explored by generating risk-adjusted O/E ratios benchmarking the local hospitals against those in the ACS-NSQIP consortium. These

findings suggested that a patient undergoing a general surgery or vascular operation at GSH was twice as likely to experience a major complication or almost ten times as likely to spend longer than two weeks in hospital, if operated on at GSH compared to an average performing hospital in the ACS-NSQIP consortium, despite controlling for a number of confounders. A patient undergoing an emergency exploratory laparotomy at a hospital in the Cape Metro West district of South Africa had a 10% increased risk of experiencing a major complication or 20% increased risk of spending longer than two weeks in hospital, compared to if the operation was performed at an average performing hospital in the ACS-NSQIP consortium.

As a quality improvement initiative, the prediction rules for a major in-hospital complication developed for both the Essentials and Procedure-targeted programmes were used to derive simple scoring algorithms which could be used pre-operatively to identify which patients had a high probability of a major complication following surgery. In general and vascular surgery, the 7-point Groote Schuur Surgical Risk score (GSSRS) was established and included five binary variables with minimal associated data burden: emergency status (2), pre-operative packed red blood cell transfusion of greater than four units (2), wound class status contaminated or dirty (1), ASA category greater than two (1), and age older than 60 years (1). The predictive performance of the GSSRS was similar to that of the original prediction rule from which it was derived. A cut-off threshold GSSRS of greater than three had a sensitivity of 76.8% and specificity of 75.0% to detect a major complication in the Essentials derivation dataset. An increase in the GSSRS was associated with a stepwise increase in the probability of a major complication following general or vascular surgery at GSH. In the Procedure-targeted

programme, the 7-point University of Cape Town Surgical Risk Score (UCTSRS) was established and included the following four binary variables with minimal data burden: sepsis or septic shock (3), ASA category greater than two (2), pre-operative urea of greater than 7,1 mmol/l (1), and age older than 60 years (1). The predictive performance of the UCTSRS was similar to the original prediction rule from which it was derived. A UCTSRS Score of greater than 2 had a sensitivity of 82.77% and specificity of 65.91 % to detect a major complication in the Procedure-targeted derivation dataset. An increase in UCTSRS was associated with a stepwise increase in a major complication following an emergency exploratory laparotomy in the Metro West health district.

Finally, in order to illustrate the notion that emerging m-Health technology provides a solution to mitigate the lack of reliable surgical outcomes research and surgical quality improvement programmes in LMICs, the use of the ACS Universal risk calculator as a tool for generating global benchmarking was piloted. In this study, it was demonstrated that the availability of these risk calculators could provide the 'E', which is necessary for the O/E ratio to perform risk-adjusted benchmarking of surgical outcomes at GSH against the ACS-NSQIP consortium. Comparing observed versus expected outcomes, ratios were greater than one for all adverse events audited at GSH except for venous thromboembolism. These ratios reached statistical significance for cardiac complications, surgical site infections, renal failure and death following general or vascular surgery at GSH. In the validation step of this study, the performance of the models predicting 30-day mortality, 30-day morbidity and LOS>30 days included in the risk calculator were excellent and surpassed the a priori degree of performance defined in the thesis hypothesis.

11.2 Trauma surgery

During the study period, 15,861 patient encounters were logged into a newly implemented eTHR. A prospectively developed clinician-entered derivation outcomes database of 7,460 patients, including 1,770 severely injured patients and 950 operatively managed patients was developed. A striking 49.7% of all injuries were intentionally inflicted. There were 188 in-hospital deaths in the derivation dataset, resulting in an in-hospital mortality rate of 2.5%. Forty two (22.3%) of these deaths were considered to be preventable. The derivation and validation datasets were analysed according to Osler's algorithm, which stated that the outcome following trauma is dependent on patient reserve characteristics, and the anatomical and physiological injury sustained. Both severely injured and operatively managed patients were more likely to be male, have evidence of substance abuse, have an intentionally-inflicted penetrating injury, have higher degrees of physiological injury on admission, require ICU, and experience an in-hospital adverse event. A total of 6,501 patients had injury severity scores calculated by a prospectively generated AIS score. Using only data, which was greater than 80% complete, two prediction rules were derived:

1 A triage prediction rule, which was useful for application using data available to a clinician on admission and,

2 An injury severity prediction rule, which was derived using all three components of Osler's algorithm.

These prediction rules were validated for use in severely injured and operatively managed validation cohorts, and their performances were compared to well established trauma scores including prospectively generated RTS, KTS and TRISS scores. This was

done by performing head-to-head analyses comparing ability of the scores to discriminate and calibrate survivors and non-survivors, as well as comparing their associated data burden. The GSH prediction rules discriminatory ability was superior to the RTS in both validation datasets, superior to KTS in the seriously injured cohort and was no different to TRISS in both validation datasets. The GSH prediction rules ability to calibrate observed and expected in-hospital deaths was superior to RTS in both validation datasets, superior to KTS in the operatively managed dataset, and comparable to TRISS in both validation datasets. The data burden associated with the GSH prediction rules was less than both the KTS and TRISS and comparable to the RTS. During the analysis it became apparent that patient reserve characteristics were the least important component of an injury severity score. Detailed descriptions of anatomical injury resulted in better performance in an operatively managed cohort, and detailed descriptions of physiological injury resulted in better performance in a severely injured cohort. In every validation step, the a priori degree of performance stated in the thesis hypothesis was achieved.

As a risk-adjusted benchmarking step, co-efficient based injury severity scores were used to compare the outcomes of all admissions, severely injured patients, and operatively managed patients managed at GSH against the outcomes expected from the MTOS. There was no difference in outcomes for the whole cohort and operatively managed patients between GSH and what would be expected from the MTOS, according to TRISS. According to both the TRISS and RTS, patients who were severely injured were more likely to die if managed at GSH than as a participant of the MTOS. According to the RTS, however, patients who required an operation were 50% less likely to die being managed at GSH compared to being a participant of the MTOS.

As a quality improvement initiative, the triage prediction rule was used to derive a simple scoring algorithm, which could be applied to a trauma victim to generate a probability of survival on admission to GSH. The 12-point Groote Schuur Trauma Score (GSTS) was developed with a higher score indicating a higher survival probability. The GSTS included five binary variables and one categorical variable with minimal associated data burden: age less than 55 years (1), triage colour of green or yellow (3), maintaining own airway on arrival (1), normotensive on arrival (SBP≥90), and GCS motor score on arrival (1-6). The predictive performance of the GSTS was similar to that of the original prediction rule from which it was derived. A GSTS of less than 8 had specificity of 98.34% and sensitivity of 72.87% to predict an in-hospital death. An increase in GSTS was associated with a stepwise increase in survival following a traumatic injury managed at GSH.

Finally, in order to illustrate the notion that emerging m-Health technology provides a solution to mitigate the lack of reliable surgical outcomes research and surgical quality improvement programmes in LMIC's, ISS coded by trained coders at VGH were compared against those coded by clinicians on the eTHR application. In this study, multiple methods were used to illustrate that the resultant ISS for traumatic injuries coded by the m-Health application were comparable to those generated by the current gold standard AIS coding. Furthermore, the estimated probabilities of in-hospital adverse events using AIS coded by data analysts or clinician-entered scoring were virtually identical.

A flow diagram that summarizes the derivation of risk prediction models and their associated scoring systems for all three programmes can be found in appendix 5.

The derivation and validation of these structured quality improvement initiatives for both general and trauma surgery raised many discussion points, such as regarding which type of QI programme should be implemented, which specific variables should be collected, how best to collect them, how to analyse these data, and the ideal quality metric to be reported in our setting. These are discussed in the following chapter with an emphasis on measuring the quality of surgical care in a limited resource setting.

Chapter 12

Discussion

12.1 Speciality-specific, procedure-specific or generic quality improvement programmes

Surgical quality improvement programmes have been shown to improve outcomes, and it would be important to develop such initiatives where they are not currently established, such as in the Cape Metro West health district. However, there is very little guidance as to how to do this in a resource-limited setting. In the United States of America, the ACS-NSQIP remains the most robust risk-adjusted and reliable tool available, and most importantly the only tool, which is readily accepted by most surgeons (186). The ACS-NSQIP has been shown to reduce both morbidity and mortality in enrolled hospitals, with the initially worse performing hospitals having the greater likelihood of improvement (150,187). However, a significant problem with the ACS-NSQIP is that it is expensive, and therefore, limits the number of hospitals participating, and excludes many smaller and rural hospitals, even in the United States of America. These are the hospitals about which one might legitimately wish to ask certain quality and safety questions (186,188). In addition, the programme requires the retrospective collection of over 130 variables per audited patient, further limiting the generalisability of the ACS-NSQIP.

Although one of the initial objectives of this thesis was to implement the ACS-NSQIP in its current format at GSH, such a programme has limited applicability in the public sector in South Africa, where arguably there was a high degree of variation in

surgical outcomes and therefore great potential for quality improvement. Despite employing a full-time clinical auditor at GSH for data collection, this process was not sustainable, and only lasted four months. Key to its failure was the general poor documentation by nursing staff and, particularly, doctors. The lack of standardisation of entries and included data points for admission, operative and discharge records were problematic. For example, both ASA and BMI are very well validated independent predictors for death following surgery and are included in most risk-adjusted models. Relying on documentation alone, these variables were missing in 18.8% and 67.3% of patient records, respectively. The only solution was for the clinical auditor to see the patients personally on the ward to get the required detail. This situation was not sustainable because of the current haphazard booking system and delayed record keeping of elective surgery performed at GSH. The operating theatre register at GSH was updated by volunteer staff who only complete the records a number of days after the operation. In many instances, the clinical auditor was only alerted of eligible cases after patients were discharged from the hospital, and as a result, the incomplete patient records, laboratory results and telephonic follow-up became the principal methods of retrospective data collection. In hospitals enrolled in the United States of America, the hospital adopts the QI programme and all staff are involved. The clerical staff based in theatre are aware of which cases meet the inclusion criteria, and the clinical reviewers are alerted electronically. Clinical reviewers have remote access to electronic patient records and the records are 98.8% complete for all 130 variables included in the datasheet. Obviously, this was far from being the reality at GSH. However, there were a number of crucial learning points taken from this pilot of the ACS-NSQIP, which enabled the

development of a Procedure-targeted QI programme relying on clinician-entered data by clinicians primarily responsible for the care of the patients.

The ACS-NSQIP Essentials programme focused on patients undergoing general and vascular surgical procedures. The same data was collected on every patient, regardless of procedure or specialty. This method of reporting has been criticised for not being sufficiently granular for targeting specific procedures or subspecialty areas for improvement. This approach may even be falsely reassuring and result in missed opportunities for improvement (158). For example, poor performance in colorectal surgery in a hospital may be masked by better than average outcomes in vascular surgery, or vice versa. Procedure-specific performance measures have been proposed to alleviate such problems, and has particular appeal in a resource limited setting (158). Effort needs to be directed on where there is the greatest return on investment of time and resources. High-risk or high-volume procedures must be prioritised for audit. The exploratory laparotomy has been proposed by the Lancet Commission of Global Surgery as one of three Bellwether procedures, which should be provided safely by a district hospital with a surgical service (189). The investigators of the SASOS study concluded that most surgical patients in South Africa underwent urgent and emergent surgery, which was strongly associated with an increased risk of mortality and unplanned critical care admissions (15). For these reasons, the emergency abdominal laparotomy was chosen for the Procedure-targeted programme.

For the purpose of this thesis, the Procedure-targeted programme was designed to also be piloted for 3-months. There was departmental commitment and all four hospitals in the Cape Metro were willing to participate. The outcomes of 450 exploratory

laparotomies were described in detail and a retrospective review of local hospital theatre registries confirmed that no eligible patients were missed. This pilot was a success, but it was not the ideal solution. A number of discussion points were raised, including the number of variables/data points requested and the reliability of clinician-entered data, particularly for complications. These will be discussed further below. A major finding was that trauma and general surgery could not be combined into one abbreviated dataset. The average trauma patient undergoing an emergency laparotomy was a young male, who was previously well and was at risk for hypovolemic shock. In comparison, the average general surgery patient was more likely to be female, older, with associated co-morbidities and at risk for septic shock. There is extensive literature validating injury scoring systems but these are limited for emergency general surgery and are not transferrable. Due to the subjectivity of the score, something even as simple as ASA becomes very unreliable in trauma (194). For these reasons, the description and the analysis of the Procedure-targeted programme in this thesis only included general surgery patients, and the 130 trauma laparotomies were excluded. The experience of implementing these two programmes has led to the proposal of a speciality-specific, Procedure-specific programme.

A similar suggestion was made by Birkmeyer *et al*, who recommended that the new ACS-NSQIP collect data on all patients undergoing a specified set of procedures within each specialty (158). Because a primary interest of ACS-NSQIP was to reduce morbidity and mortality, procedures should be selected, in part, according to their contribution to the overall number of major adverse events within each specialty. This approach would embrace both the relative frequency and the risk associated with each

procedure. For some specialties, the large majority of adverse outcomes occur in a small set of procedures. For example, according to ACS-NSQIP data from 2005 to 2006, four procedures alone - carotid endarterectomy, abdominal aneurysm repair, lower extremity revascularisation, and leg amputation accounted for approximately 75% of all complications in vascular surgery (158). In the current study, this concept was adopted in the GSH trauma centre, and in consultation with the senior staff exploratory laparotomies for penetrating trauma were identified as the index procedure to audit. A database using REDCap was developed with procedure-specific variables, and data capture commenced prospectively from the 1st May 2015. The database was currently in its 6th month and detailed outcomes of 257 exploratory laparotomies were described at the time of writing. As the database had as many as 188 optional variables for every patient, the sustainability of this database remains questionable. Which variables to include is essential in the development of any surgical outcomes database as a QI initiative.

12.2 Variables for inclusion: The 5 P's

The Donnabedian paradigm, "Measuring the quality of surgical care: structure, process or outcome?", has been adopted to evaluate the quality of medical care, specifically in the context of surgery (23). Direct outcome measures have at least two advantages. Most consider patient outcomes the "bottom line" of surgical practice, and assessment of quality with direct outcomes measures has obvious face validity and is likely to get the greatest buy-in from surgeons. Secondly, measurement alone may improve outcomes - the so-called "Hawthorne effect" (23,158). However, process measures are directly actionable, and structural measures, like surgeon experience, are important to consider in a resource-limited environment where such variables may have

greater variability. In the Procedure-targeted programme, having a surgical or anaesthetic trainee as the most qualified technician rather than a specialist actually tended towards a decreased risk. Although this association did not hold in the multivariate analysis, this finding suggested that specialists were only called upon in the most high risk or technically challenging cases. A broader model for inclusion was needed at this exploratory stage of surgical outcomes research, pertaining to more resource-limited settings. However, this needed to be balanced against not having too many variables under consideration in order to create a sustainable model. During the design and analysis of these programmes, a useful framework was adopted where factors were separated into the domains of Patient, Presenting problem, Provider, Process and Potentially preventable occurrences. In this manner, the 5 'P's of surgical outcomes research has been proposed.

In trauma, Osler's algorithm was extremely useful, showing that outcomes following trauma were dependent on patient reserve characteristics, anatomical injury and physiological injury. Considering anatomical and physiological injuries as problem variables, these too can be incorporated in the 5 P framework. AH Haider *et al* studied the influence of the NTDB on trauma outcomes, and proposed a "bare minimum" set of co-variates, which could be used to perform a risk-adjusted analysis for mortality (190). The authors suggested that the following co-variates were known to impact survival after trauma: 1) patient age 2) patient sex 3) any type of anatomical severity, 4) any type of physiological severity, and 5) mechanism or type of injury.

However, there was limited consensus on which data to collect for both trauma and general surgery in a resource-limited setting (191). Moreover, when there was

consensus, but without prior experience, many of the data collection proposals end up "casting a wide net", creating a huge data burden and thus impeding real world progress. For this reason, selected comments on data variables collected for both trauma and general surgery in this thesis will be discussed under the proposed 5P framework in order to make experience-based suggestions for a minimum dataset for a resource-limited setting.

1. Patient variables

Patient variables are a vital component of any surgical outcomes database. The following four patient variables are important:

(a) Age

The importance of including age warrants little debate. Although in the current study, the general surgery cohorts were most likely underpowered to demonstrate an association between age and a major complication, it was included for face validity. Using a binary cut-off for age does raise concern in the context of international benchmarking. TRISS for example included a binary cut-off for age (older than 55 years). Since TRISS was the gold-standard score, this age has been adopted by the KTS as well as the GSTS. However, in a subgroup analysis of penetrating injuries, this binary cut-off was not significant, since the 99th percentile for age in the penetrating injury cohort was 60 years, and only 1.1% of patients with penetrating injuries were older than 55 years. This suggested that a lower age cut-off was needed for trauma in LMICs.

(b) ASA

The American Society of Anesthesiologists' (ASA) classification of physical status was introduced in 1941 by Skalad in an attempt to provide a basis for comparison of data in anaesthesia (192). The classification was revised in 1963 with the number of classes being reduced from seven to five (193). For the last five decades it has remained a 5 class score, with or without emergent status, and it has become established as one of the single most predictive scores in outcomes research.

A number of comments need to be made regarding ASA score in the current study. In general surgery, it was one of the most predictive variables in both the validation and derivation databases, and ASA score alone had an ROC of 0.745. In the Essentials programme, each increase in ASA class was associated with a 1.86 fold increased odds of a major complication, and 1.67 fold increased odds of a LOS greater than 14 days. However, ASA was not documented in 19.64% of the anaesthetic records reviewed during the study period. In contrast, the ACS-NSQIP validation database only 0.35% of the ASA scores were missing. Despite being a simple score, ASA is still prone to interrater variability, and this was important to consider when conducting international benchmarking analyses (193,194). Comparing the derivation and validation Essential databases, the ASA scores were higher in the ACS-NSQIP sample. However, the GSH sample had greater proportions of systemic sepsis, emergency cases, disseminated cancer, diabetes, previous cardiac events, congestive heart failure, severe COPD, worse effort tolerance and renal failure. These findings question the reliability of ASA rating in such diverse settings. The reporting of ASA is not standardised in trauma, with some anaesthetists reporting premorbid ASA, and others reporting an ASA based on the

clinical findings at the time of the induction of the anaesthesia. Finally, ASA score is a good example of the need to simplify scoring systems and compromise on data granularity in limited resource settings in order to increase their utility. A binary ASA score with a cut-off of three was as predictive as the full five category score in our experience.

(c) Body Mass Index

Obesity (BMI ≥30) has been well described as an independent risk factor following general surgery, and has since been included in most parsimonious models predicting surgical outcomes (67,153,195,196). In the United States, a patient with a BMI >35 is classified as an ASA 3. The association between a low BMI and adverse outcomes after surgery has been poorly documented. However, a low albumin is a well-known predictor of adverse outcomes following surgery, although albumin is not measured routinely in many settings. In our experience, height and weight variables for calculating BMI were missing in over 65% of in-hospital records, thereby excluding BMI as a variable for consideration in our risk-adjustment models. In contrast, BMI was reported in every entry in the 320,830 ACS-NSQIP validation dataset. The importance of documenting weight and height variables would need to be brought to the attention of the nursing staff admitting patients.

(d) HIV

South Africa has one of the highest indices of human immunodeficiency virus (HIV) infection in Africa. However, the impact of HIV infection at various levels of immune suppression on surgical outcomes remains essentially unknown (197). Based at GSH, Muller *et al* challenged the prior contraindication of using donors with HIV-

infection in transplantation and has since described similar outcomes in HIV negative and HIV positive donor kidney transplant recipients (197-201). There is an unparalleled opportunity to contribute to this field of HIV and outcomes following surgery but looking at the situation currently, this will be a challenge. The HIV status was unknown for over 70% of all participating patients. In the general surgery cohorts, the highest confirmed HIV prevalence was at New Somerset Hospital (12.8%). In the trauma derivation database, the reported HIV positive status was 2.9%, but the remaining patients were recorded to not know their status (97.1%). This precluded any meaningful interrogation of the association between HIV and surgical outcomes. In light of a very successful Anti-retroviral (ARV) rollout program in South Africa, the practice of routine HIV testing prior to major surgery needs to be further explored.

2. Presenting problem variables

In the current trauma cohort, 60.7% of the injuries were classified as blunt and 39.2% were penetrating. Of all injuries presenting to the GSH trauma centre, 1,006 (13%) were secondary to firearms injuries. Penetrating injuries were more likely to be severe and require an operation compared to blunt injuries. In this analysis, the mechanism of injury (blunt/ penetrating) was not associated with an in-hospital death. In contrast, the TRISS identified a penetrating mechanism of injury as an independent risk factor for death (162). Further analysis of the comparative outcome of blunt versus penetrating injuries needs to be conducted in areas with a high incidence of penetrating injury. These findings may support the frequently cited association of surgical outcomes and provider volume (202). Injury severity scores, which focused on anatomical injury performed better in an operatively managed cohort, whereas those, which focused on

physiological injury, performed better in a cohort of severely injured patients. Predictive performance was not as high for any score in the head-to-head analysis in the severely injured cohort. This suggests the need to review scores such as the APACHE-II in severely injured patients in our setting (203).

The general surgery cohorts were underpowered to assess the association between adverse outcomes and individual diagnoses. A proxy for diagnosis was the surgical specialty involved, which strictly speaking was a provider variable. However, in both the Essentials and Procedure-targeted programmes surgical specialty had a univariate association with all outcome measures, but only a diagnosis of vascular disease was associated with a prolonged length of stay in the multivariate analysis. Similar to how AIS coding was done prospectively by clinicians in trauma, an efficient method of coding for International Statistical Classification of Diseases (ICD) needs to be developed for GSH. In HICs, procedure codes are used over diagnosis codes in generic risk-adjustment models and the proposal of procedure and specialty-specific programmes would eliminate the need for such granularity of data collection including individual diagnosis codes.

3. Provider variables

There is increasing evidence that surgical outcomes vary according to provider status (14,168,189,204-206). In fact, the most frequently cited association of surgical outcomes and provider volume has recently been put into policy in the US, and three major health systems in the US, Dartmouth-Hitchcock Medical Center, Johns Hopkins Medicine and the University of Michigan, have proposed a system-wide minimum-volume standard for 10 procedures (202, 207-209). In current surgical practice, patients and their families are turning to the Internet and other resources to make better informed

decisions about where and by whom to undergo surgery. Surgeons and hospitals are increasingly being asked to provide evidence of the quality of care, which they deliver (23). The GlobalSurg collaborative is a network of over 3000 clinicians across 67 countries. The GlobalSurg-1 project, which was launched in 2014 was an international cohort study of over 10,000 patients undergoing emergency abdominal surgery. This study showed that mortality after emergency abdominal surgery was two to three times higher in LMIC's compared to HIC's. This difference was not attributable to patient baseline characteristics alone (210).

The 'resident effect' on perioperative outcomes has been a popular theme to research (211-213). According to the literature, the resident effect implies that residents operate as assistants, and resident autonomy has been difficult to achieve in many training institutions (214-216). In this study, surgical trainees were the most senior operating surgeon in 46% of the patients in the Essentials programme, 67% in the Procedure-targeted programme, and 68% in the Trauma programme. In the Procedure-targeted programme, a trainee surgeon operating independently was associated with a decreased risk of adverse events. This may have been related to the fact that consultant surgeons are only present in more complicated cases. However, this association needs to be explored further and simple risk-adjusted models can be developed to benchmark outcomes of surgical trainees as an objective assessment of their progress. An individual surgeon is ultimately the product of the system in which he or she trained, and surgical trainees are part of a bigger system, which needs to take responsibility for adequately training those individuals. However, objective assessments in surgery are difficult to achieve. On one hand, in a resource-limited context surgical trainees are allowed to

perform a greater amount of time operating independently. On the other hand, trainees may be disadvantaged as they learn bad habits, which are harder to correct later, or may find themselves out of their depth without the necessary support, which could ultimately lead to worse patient outcomes. As a result of the previous lack of reliable data acquisition, outcomes of trainee-led operations have never been adequately audited. This must be highlighted as an area for further study.

4. Process variables

A common yet erroneous perception is that surgery involves a surgeon and an anaesthetist in a sterile environment (189). However, a more accurate assessment would indicate that surgery involves an interdependent network of individuals, processes and institutions, all of which are essential for the delivery of safe, timely and affordable surgical care and anaesthesia (6). An important advantage of measuring process variables is that they can be actionable from a provider perspective e.g. perioperative beta blockers in high-risk surgical patients (23). However measuring process variables can be challenging because of the multi-faceted nature of surgical systems and processes. An example of this, is measuring the implementation of the WHO Surgical Checklist, which has been validated for use in very diverse settings to improve surgical outcomes (217). In the current study, a binary variable was included in all three datasets to ascertain whether the WHO checklist was used. In trauma, it was included in the electronic operative form and over 99.8% of the reviewed operative records were compliant. However, this is not a reliable measure of whether the 19-item checklist was actually adhered to. Furthermore, without knowing baseline rates of adverse outcomes the impact of implementing any process cannot be assessed. The following process variables may

influence surgical outcomes:

(a) Operative procedure

Just like sample size limitations affected outcome comparisons by diagnosis, the same was true for generating risk estimates for each procedure performed in the study period. The risk quantification index developed by Dalton JE *et al* using the ACS-NSQIP database identified 2,555 CPT codes present in the derivation data set, and of these, 1,721 codes were represented by 30% or fewer patients (218). The investigators developed a risk index based on these CPT codes, however, these estimates could not be validated for use in our setting. In addition, not all coded procedures are the same e.g. an exploratory laparotomy may represent an open appendectomy or a total colectomy with clearly very different risk profiles, and teasing this out requires very granular data. During the initiation of the ACS-NSQIP, operative complexity was graded by local expert consensus (155,156). After review of this classification system, a similar local expert consensus grading system describing operating complexity in the Essentials dataset was developed. The classification system (grades 1-5) only had a univariate association with only LOS>14 days and ICU admission, but did not hold significance in the multivariate analyses. In the Procedure-targeted approach, which only included exploratory laparotomies, a risk classification based on anatomical structures involved at the time of surgery was assessed. This also failed to be meaningfully predictive. Nonetheless, these variables are important to collect and do stimulate further research including, for example, a review of local utilisation of open versus laparoscopic approaches. Collection of both procedure codes and procedure-specific variables would be strongly encouraged.

(b) Perioperative blood transfusion

Many aspects of a safe surgical network form part of a shared delivery infrastructure, which is the basis of a functional health system (189). The goals of achieving a functional health system and surgical system are not separate, and the timely availability of blood products has been used as a proxy for measuring the maturity of a healthcare system (6). A pre-operative blood transfusion of greater than or equal to four packed red cells was associated with a three-fold increased risk of a major complication in the Essentials programme. Furthermore, the ACS-NSQIP included post-operative blood transfusion within 72-hours as a complication. An ongoing audit of blood product practices in any hospital and a review of how this affects patient outcomes should be in place.

(c) Referral criteria

The analysis of the patient characteristics in the Procedure-targeted programme showed that GSH had the highest proportion of patients in the following categories, age greater than 60 years, diabetic, dependent or partially dependent functional status scores, septic or septic shock, higher mean Charlson's comorbidity indices, and patients with an ASA greater than or equal to three. These findings suggested, to some degree, that the highest risk patients were appropriately referred and in part explained the higher rates of adverse events at GSH compared to other hospitals in the Metro. There is a need for the development of more objective referral criteria based on pre-operative risk assessments according to the application of the scores developed in this thesis, and these variables need to be continuously collected to review referral criteria in the Cape Metro.

(d) ICU admission

Post-operative ICU admission is not a burdensome variable to track. Arguably, the poor performance of the scores developed in this thesis to predict post-operative ICU admission suggested an underutilisation, incorrect utilisation or a shortage of ICU beds. The current GSH ICU admission criteria are based on anecdotal evidence, subjective opinion by the physicians on call, and bed availability. A more critical analysis of our ICU admission practices is needed and these need to be audited as part of a greater QI initiative.

(e) The three delays framework

According to the Lancet Commission, there are potentially three delays, which affect timely surgical care (189). The First Delay, the delay in seeking care occurs when patients wait to seek health care because of financial and geographic restrictions, cultural beliefs, poor education, a history of being disconnected from formal health systems, a lack of awareness of available services or lack of confidence in those services. Patients often turn to informal providers (traditional healers) because they are accessible, trusted, and inexpensive. WHO reports that up to 80% of the population in low-resource settings rely on informal providers who are often poorly connected to the broader health system. This option can lead to a further delay in surgical referral. The Second Delay, the delay in reaching care occurs when hospitals with surgical capacity are scarce and the nearest facility can be hours to days away depending on the mode of transportation. The Third Delay, the delay in receiving care occurs when attendance at a hospital does not guarantee treatment since few level one hospitals can provide comprehensive emergent operative care (189).

In this thesis, data describing the third delay was reviewed. In the Procedure-targeted programme time variables describing pre-operative delays were missing in 70.9% of patients. In contrast, in the trauma dataset, these variables were complete for 95.6% of the operatively managed patients. Although, the operatively managed trauma patients were only used as a validation dataset in this thesis, in a separate analysis, a delay to theatre of greater than 6 hours for patients triaged 'red' was significantly associated with an in-hospital death (unpublished data). This analysis is still ongoing, as the pre-hospital delay also needs to be considered and addressed. At GSH, a review of the association of the third delay and adverse outcomes is very feasible with the electronic theatre booking system in place for emergency procedures, and this must be highlighted for further investigation.

5. Potentially-preventable occurrences

Patient outcomes are considered by most to be the "bottom line" of surgical practice. An estimated 10,000 operations are performed annually at GSH, and yet the risk-adjusted outcomes are not formally audited or known. In the Essentials derivation dataset, a major post-operative complication occurred at the rate of 20.97% (95% CI 16.7 – 25.78), a death at the rate of 7.67 (95% CI 4.97 – 11.2), ICU admission at the rate of 14.21% (95% CI 110.83 – 18.17), and LOS> 14 days at the rate of 22.73% (95% CI18.32 – 27.63). In the Procedure-targeted dataset, a major post-operative complication occurred at the rate of 14.15% (95% CI 10.47 – 18.52), a death at the rate of 6.56% (95% CI 4.11 – 9.86), ICU admission at the rate of 12.69% (95% CI 9.23 – 16.89), and LOS> 14 days at the rate of 8.83 % (95% CI 5.95 – 12.52). This is in contrast to the SASOS study, which collected data from hospitals over a 2-week period in South Africa, and reported

unadjusted in-hospital mortality rates of 3.1%, (95% CI 2.6 to 3.7) and post-operative admission to ICU of 6.5% (95% CI 5.7 to 7.3).

In the Trauma programme, there were 188 in-hospital deaths during the study period, which equated to an in-hospital mortality rate of 2.52% (95% CI 2.1 – 2.9). In addition, there were 205 reported complications, equivalent to an overall in-hospital adverse event rate of 5.23% (95% CI 4.7 – 5.7). During the 15-month study period, 174 patients were admitted to an ICU, giving an ICU admission rate of 3.59% (95% CI 3.1 – 4.1). There were 592 patients whose hospital stay was greater than 30 days, giving a LOS greater than 30 days rate of 7.9% (95% 7.34 – 8.57). Finally, all in-hospital deaths were discussed at the monthly in-house M&M meeting and after a multi-disciplinary discussion, 42 of the 188 deaths (22.34%) were classified as preventable deaths, equivalent to a preventable death rate for the entire cohort of 0.5% (95% CI 0.4 – 0.76).

The commonest adverse event in all programmes was a post-operative SSI. As many as one in every five patients in the Essentials programme developed a SSI. This is not an unexpected finding. SSI is the commonest complication after surgery affecting up to 25% of patients after midline laparotomy in HIC settings (210). The GlobalSurg-1 study found that the incidence of SSI for patients undergoing intraperitoneal surgery more than doubled from high (7.4%), to middle (14.4%) to low (20.0%) income countries (210). For this reason, the GlobalSurg-2 study has been designed with the primary aim to determine worldwide SSI rates following gastrointestinal surgery. All hospitals in the Cape Metro will contribute to this international evaluation including a planned re-audit following an on-line evidence based SSI education package as a QI initiative.

The ACS-NSQIP definitions of complications should serve as the backbone of post-operative outcome assessment. Standardised definitions empower researchers and clinicians alike by clarifying whether a particular event meets the criteria for inclusion as a complication, and provide a much needed degree of transparency. The ACS-NSQIP defined events into the following categories: 1) Wound, 2) Respiratory, 3) Urinary tract, 4) CNS, 5) Cardiac and 6) Other, including bleeding requiring transfusion, deep vein thrombosis requiring therapy, and sepsis and septic shock (219). These have also been adopted by the ACS-TQIP, and as we have documented, are applicable to both general surgery and trauma. With the shift towards more sub-specialisation with procedure-specific programmes, a list of standardised specific outcome variables, such as, a pancreatic fistula occurrence post Whipple's procedure will need to be defined. This has already been done by the ACS consortium for certain vascular, surgical oncology, colorectal, hepatobiliary and breast procedures (67,195,220).

There is little consensus on the period of follow-up and the frequently adopted 30-day follow-up is arbitrary. In the current study, in the Procedure-targeted approach the agreement between in-hospital and 30-day follow-up adverse events in a randomly selected cohort of 90 patients was very high, which suggested that this extra burden may not be necessary. This would need further exploration, guided by the clinical condition being audited and the available resources.

12.3 Methods of data capture

a. Relying on the traditional M&M conference

Since the introduction of the morbidity and mortality (M&M) conferences by Dr. Ernest A. Codman in the early 20th century, they have become standard practice in

modern medicine (221). The surgical morbidity and mortality (M&M) conference has been the most critical aspect of quality assurance and education in surgery departments since the turn of the last century. This method of reporting surgical complications is practised in units throughout the world and has mostly remained unchallenged and unchanged. However, there is little description or research about how best to conduct these conferences. Rules, conduct and definitions have been verbally passed down from consultants to trainees to students over the years (222). A study by Hutter M *et al* compared the data reported in a traditional M&M conference at Massachusetts General Hospital in Boston, to the data collected using the ACS-NSQIP (222). Mortality rates calculated by the traditional M&M conference (53 deaths in 5,905 patients), compared with the ACS-NSQIP nurse reviewer (28 deaths in a 24% sample of 1,439 patients), were 0.9% versus 1.9%, respectively ($p = 0.001$). Complication rates reported in M&M were 6.4% versus 28.9% ACS-NSQIP ($p<0.001$). In summary, the study suggested that one of two deaths and three of four complications were not presented in the M&M. Other than a study of 311 patients, which found similar results, literature on the accuracy of M&M conferences remains limited (222).

Our pilot study of the ACS-NSQIP did result in certainly increased mortality and morbidity reporting at GSH. However, due to the reasons discussed above, the clinical auditor missed certain complications, which were reported at the M&M conference. In order to increase morbidity and mortality reporting at the Massachusetts General Hospital, the investigators developed a web-based reporting system, which restructured the M&M conference based on the ACS-NSQIP. The ACS-NSQIP definitions of complications served as the backbone of the new system. The system still functions

today, and requires clinicians to report the complications using this web-based system, which is available as an application on a portable device (222).

b. Emerging m-Health technology

Exploiting the internet-capabilities of smart phones and mobile devices, near real-time transfer of data collected using electronic forms on mobile applications can be achieved. Emerging m-Health technology may be the solution to the lack of reliable surgical outcomes research in LMICs.

At the initiation of this study, GSH did not have a structured surgical outcomes database for research or clinical application. The development of these programmes required de novo establishment of these databases. There were varying levels of dependency on m-Health technology, as well as, clinician involvement with each database. For the Essentials programme, an adapted version of the ACS-NSQIP was made on-line using the web-based application REDCap for the use of investigators who were not directly involved in patient care. Entry occurred at the bedside where data was extracted from patient records or from independent consultation by clinicians. In the Procedure-targeted program, the database was designed using REDCap, and the data fields were populated by clinicians who had volunteered to take part in the study and were directly involved in the care of patients. The workflow in these participating hospitals did not change, and the programme was therefore an additional burden to clinicians. In the trauma programme, clinicians entered data into eTHR, which simultaneously generated patient records, and contributed to the development of a prospective database. It was possible in all three programmes, to generate validated prediction rules for benchmarking of risk-adjusted outcomes, which met the a priori

levels of performance stated in the hypothesis. However, only the Trauma programme is still ongoing after accrual of over 15,000 entries since its implementation in January 2014, and is the only programme, which utilises electronic health records. As stated previously, one of the main reasons for the unsustainability of an adapted ACS-NSQIP Essentials programme was the poor documentation by nursing staff and, in particular, doctors. The lack of standardisation of both entries and included data points for admission, operative and discharge records were identified as further limitations. Relying on the retrospective follow-up design of the ACS-NSQIP Essentials programme resulted in missing primary outcome data of a major complication in 11.8% and death in 16.09%. In contrast, the missing data in the Procedure-targeted programme for the primary outcome was 2.81% and 0% for death. In the Trauma Programme, there were no missing outcomes data for in-hospital death or any complication. Following our experience with the implementation and use of eTHR in a high-volume trauma centre, the following table summarises some advantages and disadvantages of electronic health records to measure the quality of surgical care.

Table 12.1 Pros and cons of using electronic health records to measure the quality of surgical care provided in a limited resource setting

Pros
Standardises data collection
Prospective data capture
Important data fields can be set as mandatory to ensure 100% capture
Data dictionaries and documents can be exported for communication and research purposes
Remote access from multiple devices is feasible
Surgical providers, including trainees, can track their case volumes

Records are not lost
Coding can be done prospectively including injury severity, diagnosis, billing and procedure coding
Once developed, the application can be implemented elsewhere, further standardising data collection and promoting collaboration
Prevents duplication of entries with the use of unique patient identifiers and ensures recognition of unexpected readmissions
Cost-effective in the long term
Less variability in data entry translates to less data cleaning required at the analysis stage
Data can be exported directly to common software programmes including excel, SAS or STATA for analysis
Cons
Significant development costs
Mobile devices are prone to theft
Requires significant motivation from clinical providers
Great variability exists in the applications available
Relies on a reliable mechanism of printing
A delay in clinical records may occur

An alternative approach is to rely on administrative data, which is collected routinely. Unfortunately, this occurs rarely to any meaningful degree in most LMICs. Furthermore, in HICs there is significant criticism of the use of administrative databases for assessing the quality of care, as highlighted, by an article by Ioannidis JPA, 'Are mortality differences detected by administrative data reliable and actionable?' (223). It is apparent, that computer technology is far superior to paper technology as a means for storing clinical data. By embracing computer technologies and being creative in a resource-limited setting, it could be possible to implement automated quality assessment,

which is more accurate and less costly than manual chart review. The use of risk calculators is a good example to demonstrate this point. In this thesis, the use of freely available on-line risk calculators, to better quantify the considerable global variability, which exist in surgical outcomes, was piloted. The generation of O/E ratios for targeted complications enabled a more focused analysis of such variability, and identified specific opportunities for quality improvement.

Another example, of being creative with the use of technology to improve data capture and the care we provide our patients could be the use of a Customer Relations Manager (CRM) for patient follow-up (224). A limitation of the trauma and targeted programmes was the poor follow-up of patients after discharge. CRM is an approach to managing a company's interaction with current and future customers. The CRM approach tries to analyse data about the customer's experience with a company in order to improve relationships with customers, and specifically focuses on retaining customers in order to drive sales growth. In surgery, one could envisage a similar system where a patient is entered into a CRM on the day of surgery, and a clinical reviewer or clinician is alerted the day that the patient is due for telephonic, laboratory or clinical follow-up. Additionally, an automated patient survey could be sent out, at this time, to encourage patient-reported outcomes, including quality of life assessments. This could be done for a selected group of patients according to the resources available.

12.4 Methods of analysis

a. Efficient risk-adjustment models and scoring algorithms

David Chang and colleagues, created a model with the fewest number of variables necessary to perform adequate risk adjustment to predict any in-patient adverse event for

use in resource-limited settings (225). Using the ACS-NSQIP database, the single variable with the highest ROC was ASA classification (ROC = 0.7127). ROC values reached 0.7923 with the following five variables: ASA classification, wound classification, functional status prior to surgery, albumin and age. The ROC was 0.7945 with six variables. The sixth variable was one of the following: alkaline phosphatase, weight loss, principal anaesthesia technique, gender, or emergency status. Including all 66 pre-operative variables produced little additional gain (AUC = 0.8006). They concluded, that six variables were sufficient to develop a risk adjustment tool for in-patient surgical mortality and morbidity (225).

In this thesis, we validated their findings in a resource-limited setting, but noted that the performance of the models developed in this study were superior. In both general surgery programmes, ASA was the single variable with the highest ROC to predict an in-hospital death viz. 0.8643 in the Procedure-targeted programme and 0.7851 in the Essentials programme. In the Essentials programme, adding emergency status, wound classification, pre-operative blood transfusion and age, resulted in a ROC of 0.8622. In other words, the outcome of 86.2% of randomly selected patients was correctly predicted by the model (225). In the Procedure-targeted programme, adding age, pre-operative sepsis status and a urea value, resulted in a ROC of 0.8869 to predict an in-hospital death.

The performance of these models was validated on the ACS-NSQIP database, and demonstrated similar discriminatory performance. To increase the utility of these validated prediction rules, scoring algorithms were then developed by assigning weights (points) to each predictor in the models. The weights were proportional in size to the regression co-efficients in the model, shrunken to zero. In both programmes, these scores

performed similarly compared to the original models from which they had been derived. These scores can now define individual pre-operative probabilities of an adverse outcome, as each value in the scoring system had an associated predictive probability of both morbidity and mortality derived from the ACS-NSQIP, within which they had been validated.

There are multiple validated risk-assessment and quality benchmarking tools available in surgery, and therefore deriving a score to predict adverse events following surgery was not novel. However, prior to the development of these scores, we lacked a clinically meaningful object metric, which could be applied pre-operatively to a general surgery cohort, with minimal data burden in a resource-limited setting. The surgical risk score, POSSUM, Portsmouth predictor equation, Emergency Surgery acuity Score (ESAS), and even the abbreviated ESAS are all too complex and data burdensome for routine use in our setting (226). The Surgical Apgar Score has been well validated globally, but is based on data collected intra-operatively at 5-minute intervals, which limits its application (227).

Similarly in trauma, most injury scoring systems are resource intensive, and are difficult to implement and maintain in LMICs (170). As a result, investigators in Uganda developed a simple, easy-to-calculate tool, which would be more appropriate in such settings (181). The Kampala Trauma Score (KTS), which relied on the number of serious injuries, age, systolic blood pressure, respiratory rate and neurologic status was shown to be a robust predictor of mortality. It compared favourably against the Revised Trauma Score, Injury Severity Score and Trauma Injury Severity Score in both HIC and LMICs (228-230). However, it was suggested that this score could be modified to

improve data burden and predictive performance, and therefore utility. After review of the collected predictors, which were more than 80% complete and had a univariate association with an in-hospital death, components of the KTS, which were targeted for modification, included a respiratory rate component, the neurological assessment and the number of severe injuries. Respiratory rate is data burdensome in a trauma setting, and has been criticised as being an unreliable predictor of an adverse outcome (231,232). The AVPU neurological system is not routinely collected globally, and there is evidence that the motor component is as effective, as all three components of the GCS (180,233). In a centre with a high volume of penetrating injuries, a single variable describing the number of severe injuries could lead to frequent misclassification, as one bullet wound does not necessarily translate to one severe injury sustained. This resulted in a very simple score suitable for a trauma triage, even in a pre-operative setting, as well as a validated prediction rule, which takes in all three components of Osler's algorithm.

The final five variable trauma triage model included airway, hypotension, triage colour 'red', the motor component of GCS and age greater than 54. This model was used to derive a scoring algorithm, which outperformed the KTS in a pair-wise comparison of missing data, discrimination and calibration. The data burden for the GSTS was significantly less, and the discriminatory ability superior in all datasets, apart from the operative validation dataset where the ROC's were similar. The GSTS's ability to calibrate was superior in the operative validation and the derivation datasets, and both scores had adequate calibration in the severely injured cohort. Notably the GSTS only included patient reserve and physiological injury characteristics. The addition of an anatomical description, did improve overall performance of the model compared to the

triage prediction rule. Therefore, the final injury severity prediction rule included hypotension, triage 'red', GCS motor, age greater than 54 and ISS. ISS has limited application in a simple scoring algorithm, and this prediction rule is rather designed for retrospective benchmarking of operatively managed patients.

b. External benchmarking

External benchmarking, which allows direct inter-hospital performance comparisons, has been the cornerstone of quality improvement programmes such as the ACS-NSQIP and TQIP. These comparisons exploit inter-hospital variation in risk-adjusted outcome estimates to identify centres performing significantly better or worse than their peers (190). The established method to rank trauma centres on mortality, compares the observed or actual mortality of the centre with its "expected" mortality, given its patient case-mix. The "expected" mortality is computed by risk adjusting for patient demographic and illness severity characteristics, and is then used to compute an observed-to-expected (O/E) mortality ratio (with 95% confidence interval; CI). The O/E ratio for each facility is then plotted on a graph for comparison with other centres. This information can be used to provide feedback to the individual centres and to assist them with developing targeted and informed quality improvement initiatives (157,158).

In the area of healthcare, the World Health Organisation first applied international benchmarking during the 1980s at the policy and system levels. In recent studies, international benchmarking has been extended to also cover trauma centres. Gabbe *et al* compared the outcomes following major trauma in an inclusive trauma system (Victoria, Australia) against a setting where rationalisation of trauma services was absent (England

and Wales) (59). *Schuetz et al* benchmarked trauma care performance in a tertiary hospital in Queensland and in European trauma centres (234). In general surgery, the European Surgical Outcomes Study (EuSOS) was conducted in 2011 with the primary objective to describe mortality rates and patterns of critical care resource utilisation for patients undergoing non-cardiac surgery across 28 European countries. The results showed mortality rates varied twenty-fold between countries (from 1·2% for Iceland to 21·5% for Latvia) despite adjustment for confounding variables (14,206). This meant that after adjustment for variations in perioperative factors, a patient was up to seven times more likely to die post-operatively, simply because of the hospital or country location of the surgery.

Extension of these international comparisons to LMICs has been limited. The post-operative mortality rate (POMR) following noncardiac surgery was compared among 1,514,242 patients from surgical units in New Zealand, Australia, Papua New Guinea and South Africa (191). The investigators found the risk-adjusted POMR for the unit in South Africa was 7.1-fold higher (95% CI 6.04 – 8.33) than the unit in New Zealand (191). Haider AH *et al* conducted a proof of concept study for benchmarking of trauma care worldwide, comparing risk-adjusted in-hospital mortality for two trauma centres in HIC's and one in a LMIC against the National Trauma Databank of North America (170). The LMIC centre showed significantly worse survival (O/E = 1.52, 95% CI 1.23-1.88) and the authors concluded that using only a few key co-variates, aggregated global trauma data can be used to conduct international trauma centre benchmarking adequately. There is major variability in surgical outcomes globally. However, methods as to how to improve quality reliably and sustainably are limited.

In this thesis, direct risk-adjusted comparisons of outcomes between GSH and a sample from ACS-NSQIP were made. Furthermore, novel methods of risk-adjusted comparisons were explored, including the use of the Universal calculator in general surgery and using co-efficient-based injury-severity scores in trauma. In light of the preliminary and exploratory nature of this work, emphasis should not be placed on the exact risk estimates, but rather that the findings are significantly different from the null hypothesis. A pillar for benchmarking needs to be set before the pillar can be moved in one direction or the other. Part of the original incentive for pursuing this study, was to provide solutions that contribute towards addressing a growing need for hospitals, in more resource-limited environments, to set their own benchmark for surgical quality improvement.

As a concept in general, global benchmarking of outcomes may be inappropriate in certain contexts e.g. comparing outcomes from sub-Saharan Africa with level 1-trauma centres in the United States. It would also be unlikely, that HIC trauma units would accept simpler injury-severity scoring for international comparison. Rather than pursuing comparisons, which are geographically global, a more reasonable approach has been proposed to compare similarly resourced centres, because not all HIC hospitals are abundantly resourced and not all LMIC hospitals are ill-resourced (190). As described by Adil Haider, 'One can imagine a tiered-system, where the original and lighter versions can co-exist. Using this system, low-resourced centres could use simpler injury-scoring systems, while higher-resourced centres could use more elaborate systems. This would create resource-based global benchmarking tiers, perhaps similar to how trauma centres in the United States are designated, and hospitals could choose, based on their resources,

the level of trauma and injury-severity scoring systems, in which they want to participate.' A similar system could be applied to general surgery, and in this manner there is every reason to support the development of surgical collaboratives across the Cape Metro, South Africa, and further afield with the primary objective of improving the surgical care we provide to patients. In order to do so, the following recommendations should be considered.

Chapter 13

Recommendations

1. Align surgical quality initiatives with the development of research collaboratives

As the old adage says, 'You can't manage what you can't measure,' and without hospital-level data quality improvement initiatives cannot be implemented, targeted or tracked. Whether the incentive is primarily for surgical research or for quality improvement, in this era of emerging m-Health technology, research collaboratives can provide the necessary platform to generate the data required to stimulate both research and external risk-adjusted benchmarking of surgical outcomes. The founders of the ACS-NSQIP and TQIP programmes always believed that the programmes have dual responsibilities to both surgical service management and surgical research.

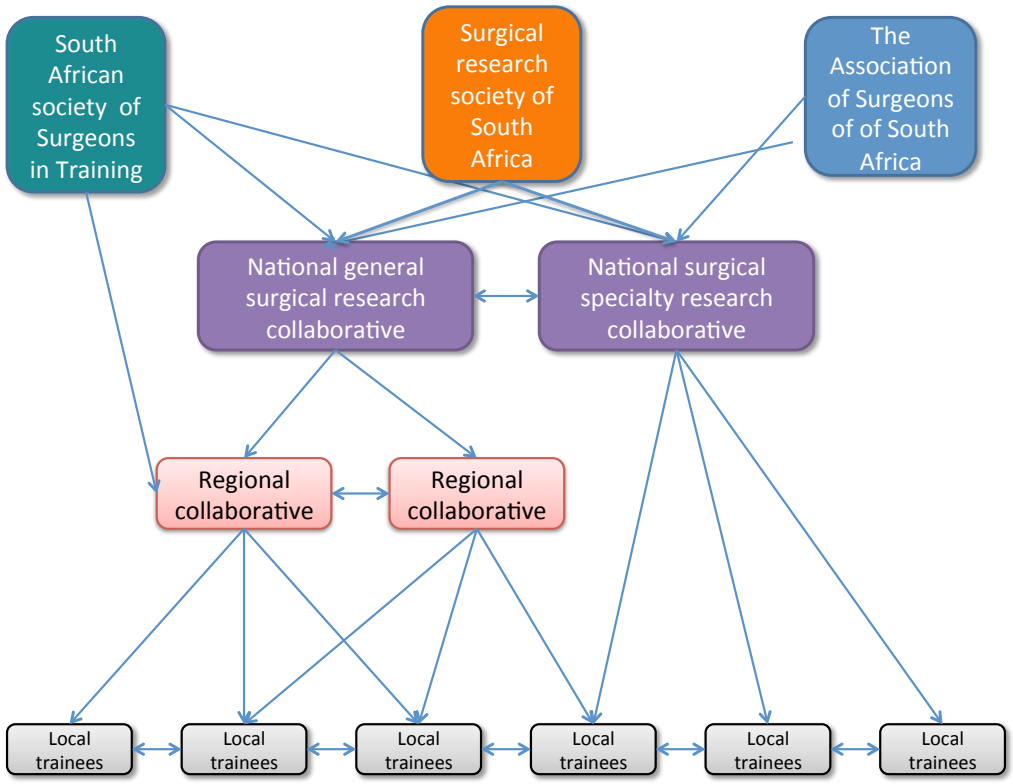
a. Surgical research collaboratives provide a solution

In the United Kingdom, over the past seven years trainee-led regional networks in general surgery have been developed to adopt a novel collaborative approach to research. Collaboration between trainees in several hospitals, allows for a larger number of patients to be included in studies over a shorter time, prevents repetition, and makes the results more applicable than those arising from single-centre studies. Trainees are ideally placed to deliver this model: they follow a rotational pattern through several hospitals, they are in regular contact with each other, they are motivated and they are expected to produce evidence of research and audit (236).

All collaboratives from the United Kingdom, have found alignment with existing organised structures to be beneficial in providing advice and support including

professional specialty associations (236). In South Africa, the Association of Surgeons of South Africa (ASSA), Surgical Research Society of Southern Africa (SRS) and the South African Society of Surgeons in training (SASSiT), could provide the academic, structural and logistical support to adopt a similar model. This would encourage surgical trainee collaborative development in South Africa to strengthen surgical outcomes research and drive quality improvement initiatives (Figure 13.1).

Figure 13.1 A proposed model for organisation and communication of national research collaboratives in South Africa



2. A minimum dataset for general and vascular surgery

Standardising data collection is a key first step to global benchmarking. However, there is little consensus on what data to collect (191). Moreover, even in cases with consensus but without prior experience, many of the data collection proposals end up "casting a wide net", creating a huge data burden and thus impeding real world progress.

Following the acquisition of very large surgical outcomes databases, such as the ACS-NSQIP, multiple logistic regression has been used to identify independent predictors, as well as to calculate individualised probabilities of an adverse event following major surgery. And with them, risk calculators based on these models have been created (4). These calculators are predominantly utilised today as a decision aide, and an informed consent tool within the context of "informed-informed consent" (220).

However, the data required as input by the ACS Universal calculator developed by Bilimoria *et al* are more appropriate candidates for a global surgical minimal dataset. Developed from the ACS-NSQIP, these data variables have very clear consensus definitions. Additionally, because these calculators were intended to be useful at the bedside or in consultation, they were developed to be very succinct, thereby greatly reducing data burden. Finally, these data elements were selected based on rigorous regression modeling on large volumes of high-quality clinical data. The data presented in table 13.1 is based on the data required by the ACS Universal calculator and the variables are presented in the 5P framework. This makes the tool more applicable to measuring quality in a limited-resource setting, rather than for 'informed- informed consent.' Those marked with an asterisk are locally relevant additions not required by the calculator.

Additions have been made following the findings of this thesis. As a patient variable, HIV status (with CD4 count, when appropriate) should be added. A specific diagnosis (ICD coded if possible) should be collected as a problem variable. In terms of the provider, a named hospital, surgical subspecialty and the qualification of the most senior operative surgeon and anaesthetist should be added. Additional recommended process variables include post-operative ICU admission (No/Planned/Unplanned) as well as, the third delay (time from admission to knife-to-skin for emergency cases).

Table 13.1 A comprehensive generic dataset for measuring the quality of general surgery care in a limited-resource setting

Patient
Age
Sex
Height
Weight
Steroid use for chronic condition (Yes/ No)
Ascites within 30 days prior to surgery (Yes/ No)
Systemic sepsis within 48 hours prior to surgery (None/ SIRS/ Sepsis/ Septic shock)
Functional status (Independent/ Partially dependent/ Independent)
Ventilator dependent (Yes/ No)
Disseminated cancer (Yes/ No)
Diabetes (None/ Oral medication/ Insulin medication)
Hypertension requiring medication (Yes/ No)
Previous cardiac event (Yes/ No)
Congestive heart failure in 30 days prior to surgery (Yes/ No)
Dyspnea (None/ With moderate exertion/ At rest)
Current smoker within 1 year (Yes/ No)
History of severe COPD (Yes/ No)
Dialysis (Yes/ No)
Acute renal failure (Yes/ No)
HIV status (HIV positive confirmed/ HIV negative confirmed/ HIV status unknown)*
ASA class (1-5)
Problem
Diagnosis (ICD-9 coded preferably)*

Emergency case (yes/no)
Provider
Hospital (Named)*
Subspecialty (Colorectal/ Hepatobiliary/ Surgical oncology/ Acute care/ Vascular)*
Most senior surgeon qualification (Trainee/ Consultant/ Sub-specialist)*
Most senior anaesthetic qualification (Trainee/ Consultant/ Sub-specialist)*
Process
Name of procedure (Converted to CPT by the risk calculator)
Wound class (Clean / Clean-contaminated/ Contaminated/ Dirty)
Third delay (Time from admission to knife-to-skin for emergency cases)*
ICU admission (No/ Planned/ Unplanned)*
Pre-operative blood transfusion ≥4 packed red blood cells (Yes/ No)*
Potentially preventable occurrence
Wound occurrence (Yes/ No)
Superficial incisional SSI
Deep incisional SSI
Organ/ space SSI
Wound disruption
Respiratory occurrence (Yes/ No)
Pneumonia
Unplanned ventilation
Pulmonary embolus
On ventilator >48hrs
Cardiac occurrence (Yes/ No)
Cardiac arrest requiring CPR
Myocardial infarction
Urinary tract occurrence (Yes/ No)
Progressive renal insufficiency
Acute renal failure
Urinary tract infection
CNS occurrence (Yes/ No)
Stroke/ CVA
Other occurrence (Yes/ No)
Bleeding requiring transfusion
Deep vein thrombosis requiring therapy
Sepsis
Septic shock

Length of stay greater than 14 days
In-hospital death

Adapted from ACS universal calculator available at www.riskcalculator.facs
* *Denotes modifications*

3. A minimum dataset for trauma

Osler's algorithm must be considered when deciding on a minimum dataset for trauma, which states:

Outcome = Patient reserve characteristics + Anatomical injury + Physiological injury (179)

In keeping with this algorithm, Haider *et al* proposed the following "bare minimum" set of co-variates which should be included to perform a risk-adjusted analysis for mortality: 1) Patient Age 2) Patient Sex 3) Any type of anatomic severity, 4) Any type of physiological severity and 5) Mechanism or type of injury (190). Considering Osler's algorithm, the "bare minimum" proposal and the findings in this thesis, a minimum dataset for trauma patients is suggested in table 13.3 under the 5P framework.

Table 13.3 A minimum dataset for patients managed operatively or non-operatively following trauma

Patient
Age greater than 54 (Yes/ No)
Sex (M/F)
Problem
Anatomical severity
Number of severe injuries (None/ One/ Two or more)
OR
Number of systems with Abbreviated Injury Score greater than three (None/ One/ Two or more)
OR
Triage colour (Green/ Yellow/ Red)
Physiological severity
Systolic blood pressure on arrival greater than 89mmHg (Yes/ No)
AND
Airway support required during transfer or on arrival (Yes/ No)
Mechanism of injury

Blunt/ Penetrating
Neurological status
Alert/ Responds to verbal/ Responds to pain/ Unresponsive
OR
Glascow coma score
Provider
Hospital (Named)
Most senior surgeon qualification (Trainee/ Consultant/ Sub-specialist)*
Most senior anaesthetist qualification (Trainee/ Consultant/ Sub-specialist)*
Process
ICU admission (No/ Planned/ Unplanned)
Massive transfusion protocol implemented (Yes/ No)
Third delay (Time from admission to knife-to-skin for emergency cases)*
Operative categories (Cardiac/ Extremity amputation/ Neck/ Thoracic/ Gastrointestinal/ Genitourinary/ Vascular)
Potentially preventable occurrence
SSI (Yes/ No)*
Length of stay greater than 14 days (Yes/ No)
Death (Yes/ No)

Adopted from the article by Haider AH et al. Influence of the National Trauma Data Bank on the study of trauma outcomes
** For operatively managed patients*

The ACS-NSQIP potentially preventable occurrences were also suitable for both operatively and non-operatively managed patients following trauma further promoting data standardisation. These could include the full classification as presented in Table 13.1 or the abbreviated classification as suggested in Tables 13.2 and 13.3.

4. Simple scoring algorithms for risk-adjusted outcome predictions

As a quality improvement initiative, the scoring rubrics presented in table 13.4 were derived and validated in this thesis.

Table 13.4 Simple scoring systems to predict outcomes following noncardiac surgery in the Cape Metro West District

Groote Schuur Trauma Score	
Component variable	***Score***
Age	
<55	1
≥55	0
Trauma triage colour	
Green/Yellow	3
Red	0
Intubation status	
No	1
Yes	0
Systolic blood pressure (mmHg)	
≥90	1
<90	0
GCS motor	
None	1
Extension to pain	2
Flexion to pain	3
Withdraws from pain	4
Localises pain	5
Obeys commands	6
Total score	***1-12***

Groote Schuur Surgery Risk Score for General and Vascular surgery	
Component variable	
Age	
13-60 years	0
Greater than 60 years	1
ASA class	
ASA class 1-2	0
ASA class 3-5	1
Wound class	
Clean or clean-contaminated	0
Contaminated or dirty	1
Emergency status	

Elective	0
Emergency	2
Pre-operative *blood transfusion within 72 hours*	
0-3 Packed red blood cells	0
Greater than 3 packed red blood cells	2
Total score	**0-7**
University of Cape Town Risk Score for emergency exploratory laparotomies	
Component variable	
Age	
13-60 years	0
Greater than 60 years	1
Acute renal failure	
Normal pre-operative urea (<7.2mmol/l)	0
Abnormal pre-operative urea (≥7.2 mmol/l)	1
ASA class	
ASA class 1-2	0
ASA class 3-5	2
Pre-operative sepsis status	
None/ SIRS	0
Sepsis/ Septic shock	3
Total score	**0-7**

Each scoring system has a predicted probability of an adverse event associated for each resultant value of the score, which can be calculated from the figures 13.2-13.4.

Figure 13.2 Probability of survival following major trauma as predicted by the Groote Schuur Trauma Score (GSTS)

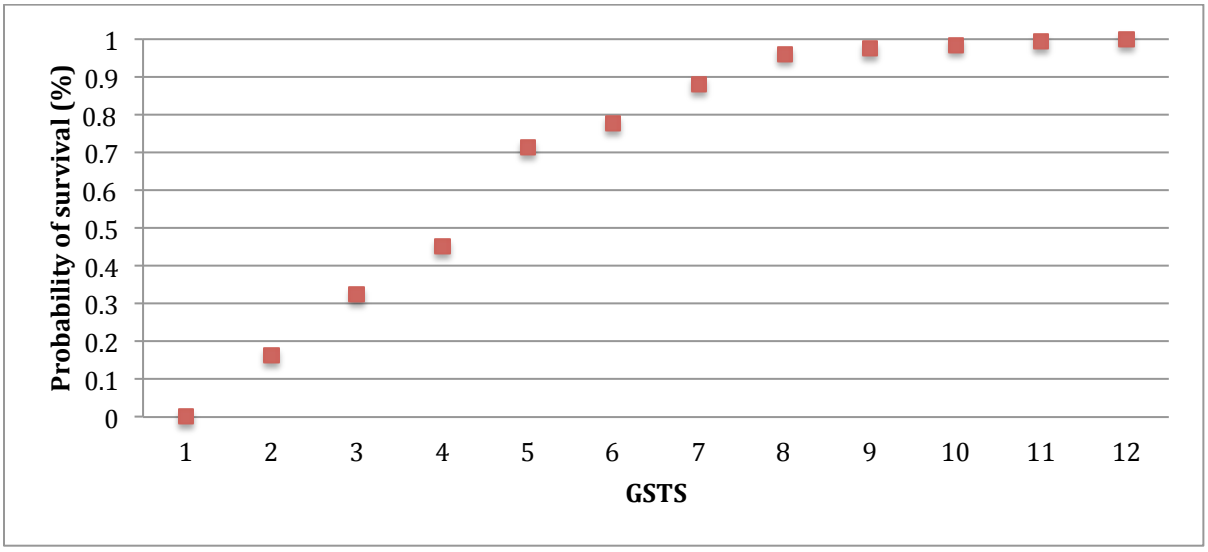


Figure 13.3 Probability of a major complication by Groote Schuur Surgery Risk Score (GSSRS)

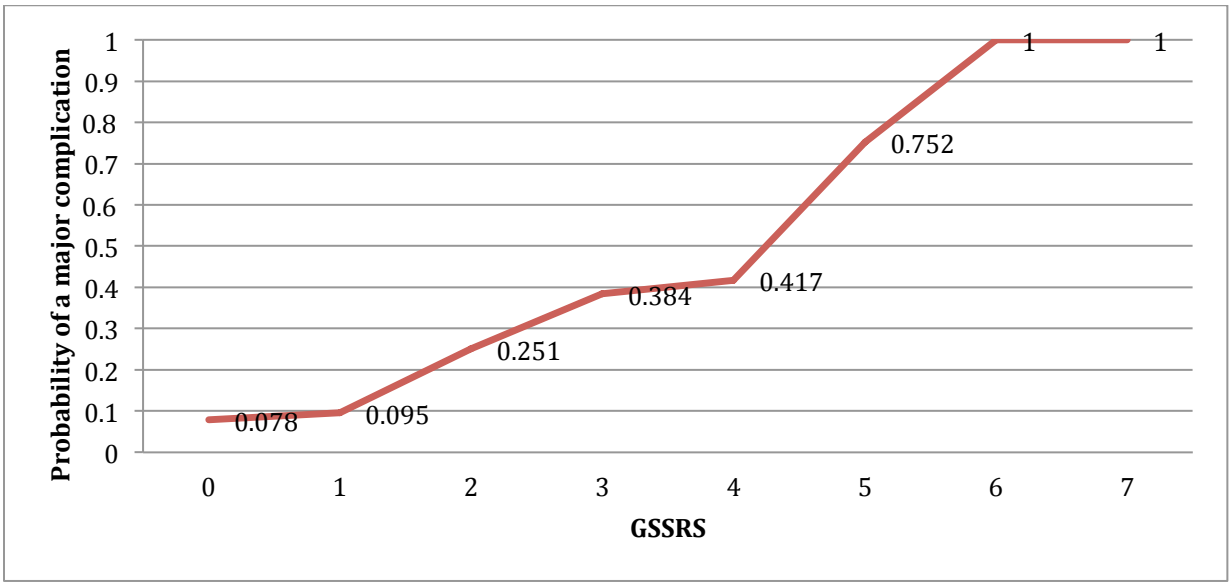
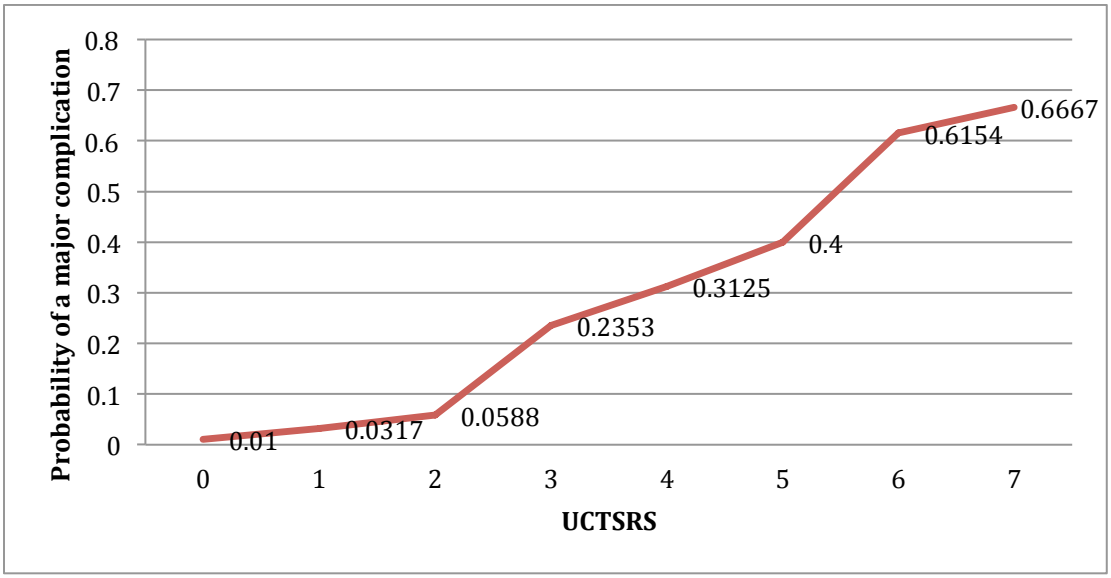


Figure 13.4 Probability of a major complication as predicted by University of Cape Town Surgical Risk Score (UCTSRS)



Designed to ensure low data burden, these scoring systems were developed for general, vascular and trauma surgery patients, managed in the Metro. These scoring systems could be applied pre-operatively to operatively managed surgery patients, or on admission for trauma patients, managed operatively or non-operatively, to achieve the following objectives where appropriate;

1. Improve communication with patients including pre-operative informed consent taking. Each value within a scoring system had a predicted probability of an in-hospital major complication, which could be communicated to the patient and family members.
2. Improve resource allocation. Cut-off threshold scores have been developed to objectively guide resource allocation like ICU admission, transfer to the tertiary level care and the need for a surgeon and anaesthetist to assist the trainee.

3. Improve both internal and external benchmarking. The scores can be used as prospectively collected quality improvement metrics to develop risk-adjusted O/E ratios (further explained below).

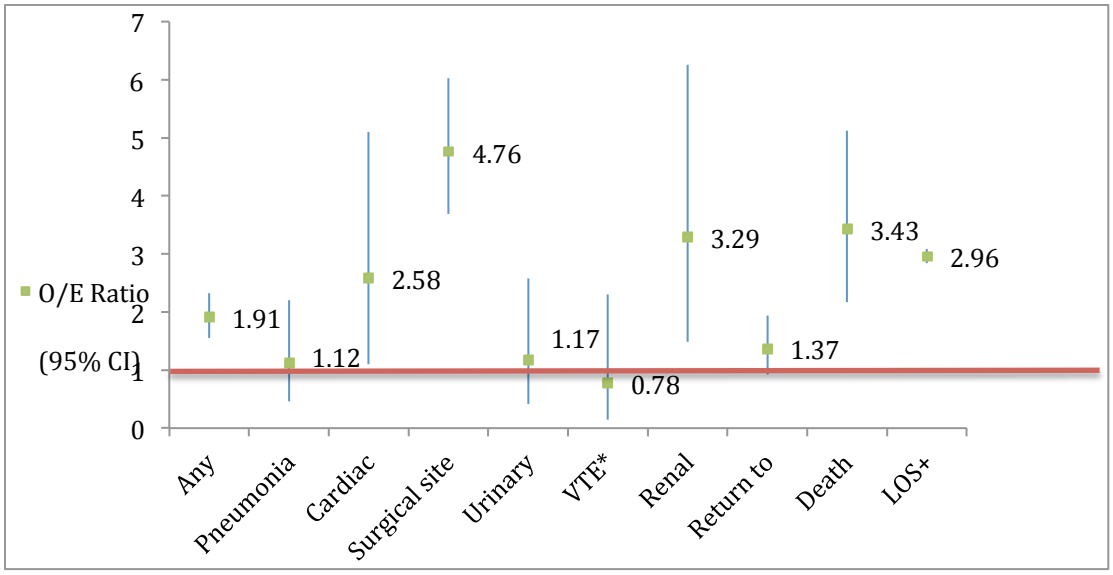
5. A meaningful quality metric: Beyond Post Operative Mortality Rate (POMR)

The Report on the Lancet Commission on Global Surgery recently highlighted six core measurable indicators, which were essential to achieving the goal of universal access to safe and affordable surgery and anaesthesia (189). One indicator, perioperative mortality (POMR) was identified as crucial for monitoring progress towards this goal. However, adverse events including post-operative mortality are an inevitable consequence of major surgery, and some adverse events may be expected or even acceptable. Without case mix adjustment and taking the heterogeneity of surgical patients and procedures into account, hospitals, which manage sicker patients, would appear to have worse outcomes. Reporting POMR alone, offers little in the way of meaningful comparisons or identifying opportunities for quality improvement.

A more meaningful surgical outcomes metric for measuring quality of care is the observed versus expected (O/E) ratio reported with 95% confidence intervals. The O/E ratio is a measure of the degree of agreement between the predicted outcome (E) and the actual outcome (O) (153). If the O/E ratio is above 1 and the lower bound of the confidence interval is less than 1, then the surgical unit has experienced a statistically significant larger number of adverse events than would have been expected on the basis of its patient characteristics. With these data, meaningful comparisons of risk-adjusted outcomes between surgical units can be achieved.

Nevertheless, the adoption of the O/E ratio as a metric for surgical quality has been limited, mostly because of the difficulty in calculating the 'E' in the O/E ratio. With the scoring systems proposed, however, the individual patient probabilities calculated can be summed across any surgical collaborative to obtain the expected number of adverse events (E) based upon the patient characteristics of the service. Having obtained the 'E', calculating the numerator of an O/E ratio is simple, and entails recording actual observed outcomes for the same predicted adverse events. Confidence intervals (CI) can then be calculated using a Poisson CI calculator on any standard statistical package to identify statistically significant low or high O/E ratios. The result of these analyses can be communicated back to surgical units within a collaborative in the form of interpretable caterpillar plots such as presented. The figure 13.5 shows a caterpillar plot benchmarking GSH against the ACS-NSQIP consortium. Using the O/E ratio, gives surgical units a benchmark to target corrective action and improvements. This process is far superior to reporting a POMR alone, which offers little opportunity for identifying opportunities for quality improvement.

Figure 13.5 Caterpillar plot benchmarking Groote Schuur against the ACS-NSQIP consortium



6. A Surgical Quality Improvement Programme for Groote Schuur Hospital: A specialty-specific, procedure-specific programme

In the current sampling scheme of the ACS-NSQIP, hospitals collected data on a sample of all patients undergoing any general or vascular surgical procedure under general or regional anaesthesia. Following the experience in this thesis, our findings support the recommendations made by Birkmeyer *et al* (158) and recommend that a QI programme for GSH should include data collection on all patients undergoing a specific procedure within a particular specialty. Because a primary interest of this programme was to reduce morbidity and mortality, procedures were selected, in part, according to their contribution to the overall number of major adverse events within each specialty. This approach would account for both the relative frequency and risk associated with each procedure. For some specialties, the large majority of adverse outcomes occurred in

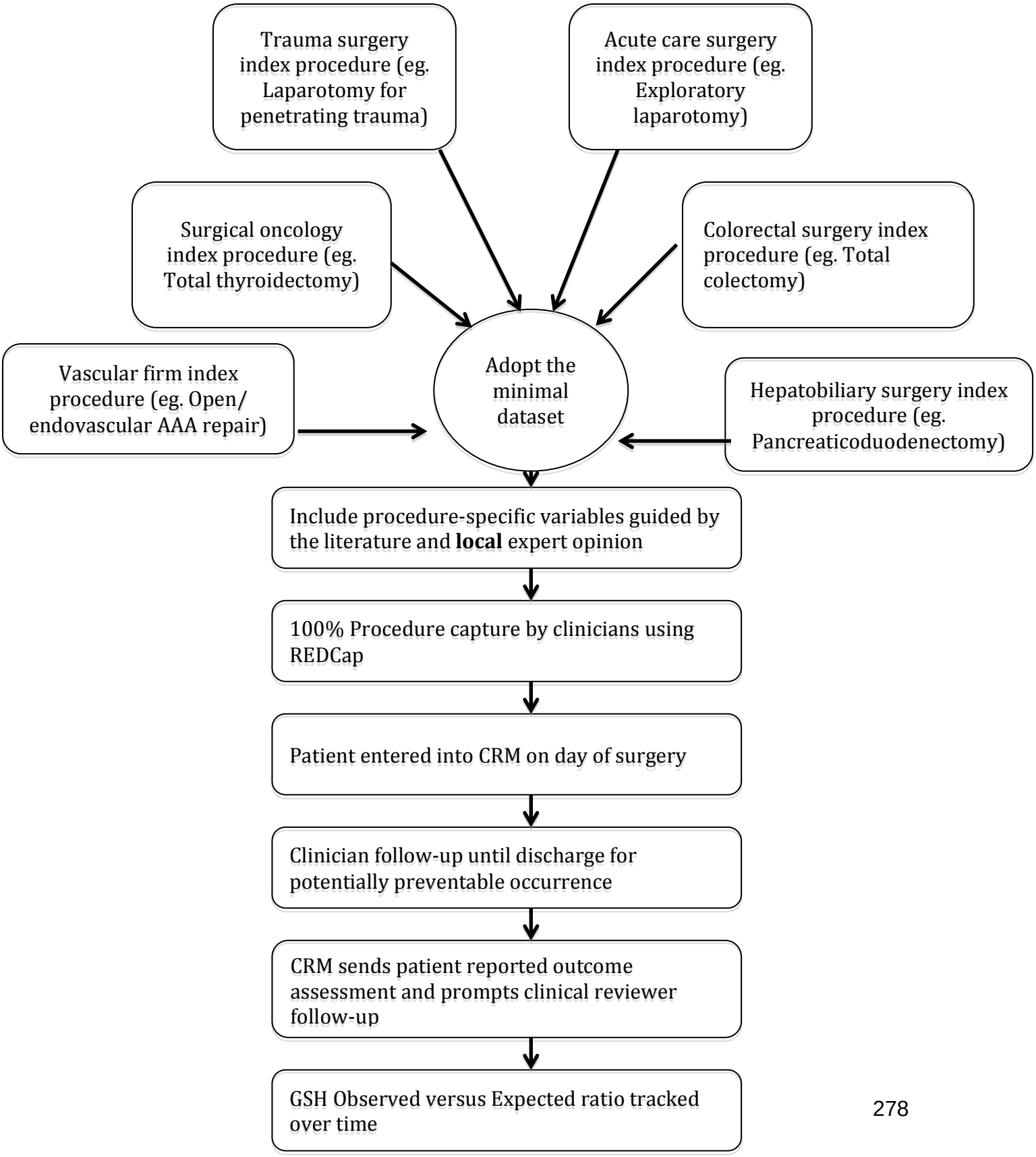
a small set of procedures. Procedure-specific outcomes assessment has obvious advantages for quality improvement. As currently provided by ACS-NSQIP, specialty-level outcome measures provide surgical leaders with a "bottom line" of their overall performance. They are also useful for monitoring specific outcomes relevant to almost any procedure (eg surgical site infection). However, specialty-level outcome measures are not sufficiently granular for targeting specific procedures or subspecialty areas for improvement. Importantly, focusing on procedure-specific outcome assessment would also reduce the amount of information needed for risk adjustment (158).

Adopting this procedure-specific, speciality specific concept, figure 13.6 describes a proposal for a GSH Surgical QI program. The leadership of each surgical firm should identify a procedure of interest to the firm, which, ideally, contributes significantly to the overall resource utilisation, morbidity and mortality within the subspecialty. Each surgical firm should adopt a minimal dataset by reviewing the more comprehensive generic dataset and the current literature. Based on local expertise, relevant procedure-specific variables (including outcome variables) should be identified for inclusion. A database can be developed based on these variables in REDCap. Data collection should be done prospectively by clinicians within each firm using a m-Health application. This process should be supported by the leadership of the firm to ensure reliable and accurate capture, as well as stimulate consensus decisions regarding adverse events. Ideally, patients should be entered into a Customer Relations Management System (CRM) on the day of surgery designed specifically for this purpose at GSH. This system can be developed at low cost and should be designed to send a patient-reported assessment form via text /email at 30-days from the day of surgery. Furthermore, the

CRM system should prompt an employed clinical auditor at the GSH department of Quality Assurance with a patient's contact details in order to follow up on the patient after 30 days by folder, laboratory and telephonic review.

The data generated by this programme, could be used to calculate general surgery and procedure-specific O/E ratios for risk-adjusted outcome tracking of the quality of care provided to patients. This could be used for internal and external benchmarking over time. It would also add value to the M&M conference, identify opportunities for quality improvement, stimulate research, and measure the impact of any corrective action subsequently implemented.

Figure 13.6 A surgical quality improvement program for GSH: A speciality-specific, procedure-specific programme



7. A Surgical Quality Improvement Programme for the Cape Metro West Collaborative: The Codman Calculator.

Ernest Amory Codman, a surgeon at the Massachusetts General Hospital in the early 1900s, developed an "End Results" system in which a detailed patient history and outcomes were documented, adverse events were systematically reviewed and their causative errors were categorised. Codman was instrumental in the development of the Hospital Standardisation Programme, which in turn became the Joint Commission on Accreditation of Health Care Organizations. His pioneering work in the early 1900's was far beyond his time and many of his recommendations are applicable to our setting today e.g. 'Comparisons are odious, but comparison is necessary in science.'

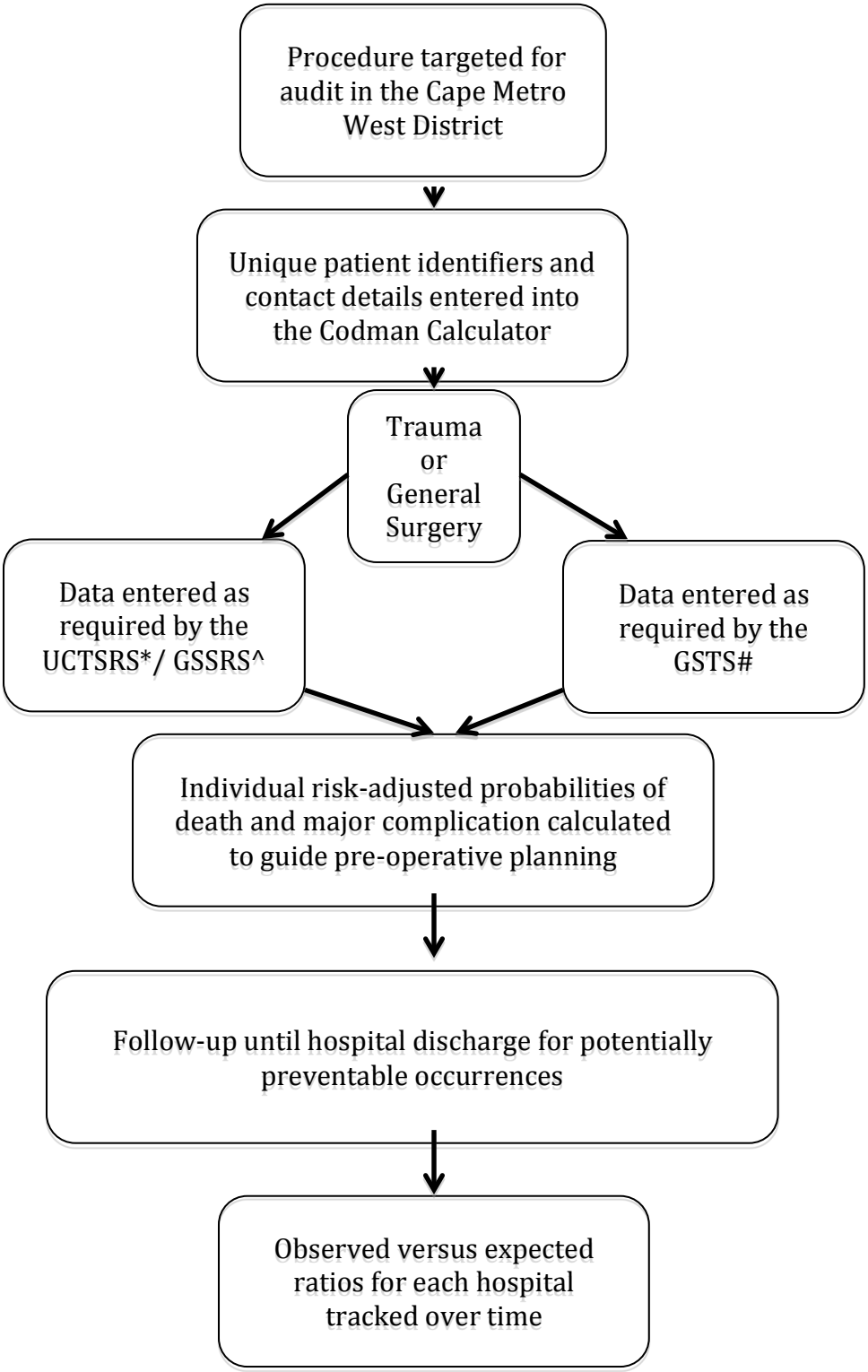
As demonstrated in the trauma department, by exploiting the internet-capabilities of smart phones and mobile devices near real-time transfer of data collected using electronic forms on mobile applications can be achieved. Following the acquisition of very large surgical outcomes databases by the ACS-NSQIP, risk calculators based on their regression models have been created (4), with relevance to general, colorectal, hepatobiliary and breast reconstruction surgery, as well as surgical oncology (195,240-242). These calculators are predominantly utilised today, as a decision aide and an informed consent tool within the context of "informed- informed consent" (220). However, in this thesis it has been demonstrated that beyond patient-level decision-making, the availability of these risk calculators makes risk-adjusted benchmarking of surgical outcomes feasible.

Therefore, the minimal datasets proposed for general and trauma surgery should be built into a m-Health application that I would propose should be called the, 'Codman

Calculator.' In comparison to the development and implementation of eTHR at GSH, this application would be very simple to develop.

Figure 13.7 demonstrates how such an application can be used in practice. Ideally all major surgery should be tracked, but this process should be piloted with a specific procedure, which has been targeted for audit purposes. Unique patient identifiers can be generated and the risk scores completed on the application. This will generate real-time pre-operative probabilities for both in-hospital morbidity and mortality. This can be used for taking informed consent and guiding pre-operative planning, including referral to a higher level of care, ICU transfer and senior surgical and anaesthetic cover. Following surgery, patients can be followed up until discharge and potentially preventable occurrences noted, and entered into the calculator. Similar to the proposed GSH programme, the data generated by this programme could be used to calculate O/E ratios for risk-adjusted outcome tracking of the quality of care provided to patients. This can be used for internal and external benchmarking over time. It would also add value to the M&M conference, identify opportunities for quality improvement, stimulate research and measure the impact of any corrective action subsequently implemented.

Figure 13.7 Demonstrated utility of the proposed Codman Calculator



UCTSRS University of Cape Town Surgical Risk Score, GSSRS^ Groote Schuur Surgery Risk Score, GSTS# Groote Schuur Trauma Score.*

8. An unexpected outcomes approach to the Morbidity and Mortality (M&M) meeting

‘‘For the man who practises surgery, there are two kinds of mortality-chance and intentional. Chance mortality is the kind, which occurs unexpectedly. Intentional mortality is incurred by the chief surgeon when he attempts cases in which the condition is acknowledged to be grave’’ E. A. Codman (221).

The Minimum Standards, adopted in 1919, required that “the staff review and analyse at regular intervals their clinical experience in the various departments of the hospital.” (222). Such meetings have been taking place in surgical departments globally ever since. In 1983, the Accreditation Council of Graduate Medical Education in the United States of America, made it a requirement for departments with surgical training programmes to hold “a weekly review of all current complications and deaths, including radiologic and pathologic correlation of surgical specimens and autopsies” (222).

How best to conduct M&M conferences remains unresolved. Rules, conduct, and definitions have been verbally passed down from senior to trainee to student over the years. The Institute of Medicine’s 1999 report, “To Err Is Human”, heightened awareness with regard to errors/near misses, quality measurement, and reporting (222). This Institute of Medicine report, and the follow-up report, “Crossing the Quality Chasm”, have focused attention on “errors caused by faulty systems, processes, and conditions which lead people to make mistakes,” rather than on individual culpability (222). The monthly surgical M&M meeting at GSH is conducted in an open peer-reviewed manner and cases are categorised after critical review by general consensus.

This process could be strengthened by the unexpected outcomes approach proposed by Bohnen J *et al* (221). Clinical scenarios represent four archetypes of patient outcomes in healthcare: expected successes, unexpected failures, unexpected successes, and expected failures, respectively (Fig. 13.8). Many patient outcomes can be categorised into these groups using clinical intuition. Bohnen J *et al* suggested that there were many benefits in using this approach and included: (1) Informed decisions could be made about whom to present during M&M; (2) Unexpected successes (box 3) could be introduced into M&M conferences and studied more systematically; (3) enhanced atmosphere of M&M by celebrating unexpected successes; and (4) creation of an evidence base to counter potential financial and reputational consequences for expected failures (box 4).

Figure 13.8 Archetypes of patient outcomes

		Outcome	
		Success	Failure
Expected Result	Success	1. Expected Successes	2. Unexpected Failures
	Failure	3. Unexpected Successes	4. Expected Failures

Adopted from, Bohnen JD. Et al. 'Reconceiving the morbidity and mortality conference in an era of big data.'

However, two key tools were needed to facilitate this process: (1) accurate, patient-level risk-adjustment models (i.e. risk scores) and (2) an objective definition of the optimal cut-off point between expected and unexpected results. In this thesis, we have developed and validated three risk scores, which can be used to generate expected predicted risk. Furthermore, using the Youdin index, an optimal cut-off point for each score was defined. Using these scores and cut-off thresholds, the unexpected outcomes approach could be used in our M&M conference to objectively guide and target the discussion, as well as to improve the efficiency of what has been termed the 'golden hour' of surgical training.

Chapter 14

Limitations

There are several limitations to this work and the findings must be interpreted within the context of the study design.

14.1 Missing data

In all three programmes, missing data was a recurring problem. Although, certain variables like gender and operative procedures were 100% complete, others particularly laboratory values were greater than 85% incomplete. This was not surprising. However, of concern were the following missing variables in the Essentials programme: ASA class, height and weight, and, particularly, missing outcomes data. It was not possible to go back and retrospectively measure the weight and height of patients who had already been discharged prior to their inclusion. This was due to the fact that the records of the elective general surgery GSH cases included in the study were updated about one week after the surgery had taken place. Relying on retrospective data capturing by a single clinical auditor in the Essentials Programme resulted in missing primary outcomes data for 11.8% of the cohort. The prospective nature of the clinician-entered electronic datasets minimised the dependence on retrospective folder review. Furthermore, variables, which were deemed essential, were made mandatory entries in the electronic datasets. Rates of missing data were therefore decreased and missing primary outcomes data in the Procedure-targeted and trauma cohorts were 2.89% and 0%, respectively. However, due to the exploratory nature of this study, a larger number of variables than necessary were included in the databases, as the most significant predictors which were

most feasible to collect were previously unknown. Therefore, even the real-time electronic datasets used in the Procedure-targeted and Trauma programmes still had some high rates of missing data.

Throughout the analysis, methods of dealing with missing data were made clear and all missing dependent and independent variables included in the models were entered as missing. Only independent variables, which, were greater than 80% complete, were included, which could be a source of bias. Although this was not ideal, and it would have been more valid statistically to include every variable in each original dataset, we needed to develop models where collecting data was feasible in our setting. Despite the existence of many risk-adjustment tools in the literature, there was no objective measure, which was feasible to use locally. Relying on data, which was less burdensome to collect, resulted in a more applicable and parsimonious model. Furthermore, as logistic regression only includes subjects with complete data for all variables, insisting on the inclusion of all variables would have grossly limited our already small sample size for analysis in the general surgery programmes.

14.2 Sample size

GSH is one of the busiest trauma referral hospitals in the world, and therefore once a sustainable method of data capture was implemented, sample size for the trauma programme was never going to be a limitation. In general surgery, however, both datasets were developed de novo and the sample size was dependent on the sustainability of the method of data capture used. The Essentials programme was never intended to last only three months. However it became very apparent that insisting on a longer collection time would have threatened the accuracy of the data collected. The Procedure-targeted

programme was modeled on a three-month study period. This was mainly due to the fact that surgical trainees rotate at 3-month intervals and was therefore not going to be available after a 3-month pilot period. Both cohorts were approximately 90% powered to test a 1.5 measure of association for a single predictor in a model predicting an event, which occurred at an estimated rate of 15%. This seemed reasonable for an exploratory study of this nature. An outcome of in-hospital mortality was therefore too rare an occurrence, for the General Surgery programmes given the 10:1 rule of ten predictors to one event. This was circumvented by deriving a score to predict a major complication and reviewing its performance to estimate in-hospital mortality. Alternatively, the ACS-NSQIP validation dataset of 320, 816 patients could have been used to derive the scoring systems and then validated on our datasets. However, similar work had been done previously, and although these efficient risk-adjustment models had not been validated in a resource-limited setting, these models relied on laboratory values as well as BMI (225). However, we found that both of these variables were not routinely collected in our setting. Furthermore, the most predictive variables in our setting needed to be identified instead of simply relying on variables, which are most predictive elsewhere.

A major complication was defined by an occurrence classification developed in the ACS-NSQIP. This classification included both wound and cardiac occurrences, and therefore, puts equal weight on the occurrence of a deep wound infection and a myocardial infarction, which is a major limitation. An alternative, which was considered was to further classify all complications according to the Clavien Dindo grading classification (178). An example of a primary outcome measure could be a Clavien Dindo classification of three or greater including any life-threatening complication, any

complication resulting in death and any complication requiring surgical, endoscopic or radiological intervention. This was not feasible, as in both programmes the only validation dataset available was the ACS-NSQIP. Outcomes for prediction were therefore, limited to those outcomes included in the ACS-NSQIP programmes. Interestingly, perhaps not surprisingly, the models developed to predict a major complication performed best when used to predict in-hospital mortality. Following the findings of this thesis, a limited dataset has since been defined. If adopted by the hospitals in the Metro, in time, a larger sample size would enable development of individual models specifically for death, wound, cardiac, respiratory, CNS and renal occurrences, as is the case in the current ACS-NSQIP (243,244).

14.3 Clinician-entered data

A major limitation of the ACS-NSQIP and TQIP were that they require retrospective collection of over 100 variables per patient in both programmes, which would limit their generalisability to LMICs. The availability of external auditors/clinical reviewers is a privilege that most surgical units around the world do not have. We therefore hypothesised that using available technology, clinician-entered data may provide a solution. In reality, it is probably the only solution for data acquisition.

Two of the programmes were developed based on clinician-entered data. A major criticism of clinician-entered data is that it is prone to misclassification bias of outcomes. Several steps were taken to minimise this, including peer-review of the outcomes recorded during consultant led ward rounds in the Trauma programme and during the combined morbidity and mortality meetings in both programmes. The primary outcome in the Trauma programme was in-hospital mortality, and each death was verified against

the records of the state mortuaries. In the Procedure-targeted programme, the primary outcome was a major in-hospital complication. GSH contributed 109 patients to this cohort. Of these, 90 (82.6%) were followed up at 30-days and the clinician entered outcomes and 30-day outcomes were compared. The Kappa statistic to compare the inter-rater reliability between these two methods of follow-up for all outcome measures were all greater than 0.81, translating to almost perfect agreement. This suggested. in-hospital outcome measures were sufficient to begin with, and that the clinician entered outcome assessment was reliable. Again, it must be highlighted that outcome assessments using clinician-entered data had significantly less missing data than the retrospective entry by a clinical auditor.

For a busy clinician, there are concerns that the data entered may not be as accurate as traditionally collected data with external reviewers. In the Trauma programme in this study, we compared the ISS calculated by clinicians against that calculated by the gold-standard method of AIS coders. Despite being a complicated system, the findings showed that the two methods resulted in very similar scores. A study by Anderson JE *et al* compared data collected by electronic health records against data collected by traditional methods in the ACS-NSQIP (245). They concluded, that rigorous risk-adjusted surgical quality assessment could be performed solely with objective variables in electronic health records (245). Importantly, only objective data variables were used in the electronic registries. In the current study, we were concerned about using subjective variables for external and international benchmarking. When comparing the Essentials derivation and validation datasets, although the patients in the GSH validation dataset appeared to have more advanced disease, the ACS-NSQIP dataset

had higher ASA scores. This challenges the use of subjective scores in such diverse settings. The study by Cohen ME *et al* explored the effect of subjective pre-operative variables on risk-adjusted assessment of hospital morbidity and mortality. Using ASA and functional health status, the authors showed that these variables had unique contributions in risk-adjustment models, and gave little indication that they were subject to institutional bias. They concluded that it was appropriate to use these variables to assess risk-adjusted surgical quality (246).

14.4 Electronic health records

The exponential growth in computing power and increasing usability of computer interfaces has created unprecedented opportunities for data acquisition and research. However, the use of electronic health records does have limitations and concerns in our setting. The implementation of electronic records and adoption of m-Health applications as in the Trauma programme were chaotic at first. For example, iPads were stolen or broken by intoxicated trauma victims. The database required an uninterrupted Internet connection in the development stages, and this resulted in the trauma unit having to revert to paper records twice in the implementation phase. eTHR had the capacity to collect over 600 variables, which resulted in a significant amount of redundancy. In contrast, the minimal dataset developed for trauma consisted of only 17 variables. For these reasons, the data collected in the first three months was disregarded as pilot data, and only data from the 1st of April 2015 was included in the Trauma programme. Only a few fields were made mandatory for completion, and during most of the data acquisition the author was a full-time participant in the trauma unit to cover any issues. This was the first pilot of electronic health records in GSH, and not surprisingly there were unexpected hurdles.

The counterargument to all these highlighted issues is that 18 months after initial implementation, the eTHR is still being used and the database continues to grow.

14.5 Statistical validation of predictive models

It is widely accepted, that the performance of risk prediction models should be assessed by both their calibration and discrimination (247). In this thesis, the method chosen to assess discrimination was the area under the receiver operator curve (ROC). This was the most popular method used in the literature and also the most widely accepted (175). A criticism of the ROC method was the insensitivity in model comparison in which the baseline model had performed well (247). Due to the exploratory nature of this study and the primary objective being to identify only the baseline model, the ROC seemed appropriate and a ROC of greater than 0.7 was chosen in the hypothesis. For a measure of calibration, the Hosmer-Lemeshow statistic was chosen. This was also chosen based on the fact that it was a widely accepted method of calibration. However, such traditional metrics of calibration are chi-squared distributed, and therefore a limitation with these statistics was that small deviations from perfect calibration in larger sample sizes tend to be statistically significant (248). This might have explained why none of the models developed in the Essentials programme calibrated well in the large validation datasets as they were over powered. However, the subsequent direct observed to expected comparisons performed, confirmed that although the predictors chosen adequately discriminated both datasets for the primary outcome of interest, there was a direct mismatch between estimates of the models and actual outcomes. Therefore this was interpreted as potential to improve the quality of surgical care provided.

14.6 Data granularity

The development of binary models was to ensure the most parsimonious and user-friendly models. In contrast, the major focus around data collection efforts in the American collaboratives was the creation of more granular data. The assumption was that more complex data collection translated into the accrual of more accurate data and better outcome predictions. Many would argue that dichotomising variables is oversimplifying a problem at the expense of a loss of data. In this thesis, this notion was challenged. The models developed were compared using the original variable definitions and binary definitions, which had equivalent predictive performance in all the derived models. A possible criticism could be how the cut-off threshold definitions were chosen. A similar debate, currently ongoing in the United States of America involves defining cut-off thresholds for 'high volume' hospitals. Hospital volume has been shown repeatedly to be an independent predictor for surgical outcomes, but the threshold for defining a 'high volume' hospital seems arbitrary (202,207,249). Therefore, it is important to use both clinical acumen and a review of the literature to identify the most plausible cut-offs for dichotomising data. As long as similar cut-offs are used, external benchmarking based on binary models may be criticised for oversimplifying a complex problem, but it is at least offering a more feasible and sustainable solution to acquire risk-adjusted outcomes.

14.7 Reliance on standardised data

Due to the paucity of well-validated surgical outcomes programmes and databases in LMICs, this study was reliant on programmes and data from HICs to design, implement and validate the work, as well as to externally benchmark the surgical

outcomes. This was an unavoidable limitation, and to compare outcomes in such diverse settings may not be justified. Using local data was an option as remarkable work had been done by Clarke DL *et al* in setting up electronic data registries (250). However, these local databases had been designed specifically for data acquisition and not for quality improvement. Standardisation of data is at the cornerstone of quality outcomes research. In trauma, we learnt from the TQIP to design and implement eTHR, and data from this programme can now be benchmarked against The National Trauma Databank in the United States of America. The eTHR application is immediately scalable for other units in similar settings to adopt, resulting in exciting future avenues of work to enable benchmarking of equally resourced centers against each other. In General Surgery, the researcher was able to benchmark the outcomes against the ACS-NSQIP. Data standardisation is therefore, at the cornerstone of any external benchmarking exercise.

14.8 A culture of service delivery over research and audit

Hospitals in HICs are increasingly being asked to provide evidence of the quality of surgery they provide. It is well accepted that the need to compare surgical outcomes across hospitals is of paramount importance to patients, physicians and even funders. In current surgical practice, patients and their families are turning to the Internet and other resources to make better informed decisions about where and by whom to have surgery. Faced with these external pressures, surgical communities in HIC have responded with programmes like ACS-NSQIP. Such external pressures are less locally, and there are few legislative requirements to produce evidence of quality or audit in surgery. Nonetheless, budget restrictions are greater, and there is evidence to suggest greater variability of outcomes, which will have direct resource implications. The minimal datasets proposed

in this thesis, need to be further interrogated, prospectively collected, validated and improved upon for this reason. Local benchmarks for quality are needed. However, it remains questionable whether the adoption of such evidence-based suggestions is currently realistic or idealistic within GSH, the Western Cape Metro or even on a national level. The continued culture, which emphasises service delivery over research and audit, may be the single biggest limitation of this work.

Chapter 15

Conclusion

As surgery assumes a greater position in the global health agenda, the need to not only improve access to surgical care, but also improve the quality of surgical care, is paramount. There is increasing evidence of the significant global variation in surgical outcomes. A better understanding of the reasons for these variations must be interrogated to identify opportunities for improvement.

Emerging m-Health technology has the capacity to transform surgical outcomes research and quality improvement programmes in LMICs, such as South Africa. Various m-Health applications were used in this study to address the primary aim of designing and implementing structured surgical programmes within the Cape Metro West health district. These programmes reliably described risk-adjusted outcomes following major surgery for benchmarking and quality improvement initiatives. This involved wireless, real-time, collection of the proposed minimal datasets by clinicians using mobile devices. The data was used to populate efficient risk-adjustment models and scores derived in this thesis, which have demonstrated adequate discrimination and calibration to reliably predict individual probabilities of a major complication in a surgical patient, managed within our collaborative.

Such predictions, could be used to enhance informed consent, guide pre-operative resource allocation, and enable internal and external benchmarking of our surgical outcomes, using an objective quality metric i.e. the observed over expected (O/E) ratio. Internal benchmarking provides the opportunity to measure the effect of any corrective strategy implemented, and a strategy to target surgical site infections should be

prioritised. External benchmarking, which allows direct inter-hospital performance comparisons, has been the cornerstone of quality improvement programmes in High Income Countries. Objective definitions of the optimal cut-off point between expected and unexpected results have been proposed promoting the adoption of the ‘unexpected outcomes approach’ to our Morbidity and Mortality meetings. Such a process would strengthen our existing surgical collaborative and promote capacity for research. Implementation of these recommendations will close the gap between data capture, analysis and action.

Appendices

Appendix 1

AMERICAN COLLEGE OF SURGEONS
DRAFT THESIS.DOCX NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM

ESSENTIALS WORKSHEET

*IDN _____ Cycle Number _____
LMRN_____ Case Number _____

DEMOGRAPHICS

Last Name: _____First: _____MI: _____

Street Address: _____-
_____

City/Town: _____State/Province: _____ Zip: _____
Country: _____

Home Phone (_____) _____Work Phone (_____) _____Cell Phone(_____) _____

*DOB: _____/_____/_____ *(mm/dd/yyyy)* Gender: ☐*Male* ☐ *Female*

Race: ☐White ☐American Indian / Alaska Native ☐Asian
☐Black / African American ☐ Native Hawaiian / Other Pacific Islander
☐Unknown

Ethnicity: Hispanic- ☐*YES*☐*NO* Preferred Language: ☐ENGLISH
☐ SPANISH

SURGICAL PROFILE

Principal Procedure _____ CPT Code _____

Patient Status:☐*Inpatient* ☐*Outpatient* Elective Surgery: ☐*YES* ☐*NO*
☐ *Unknown*

Transfer / Origin Status:

☐Not transferred, admitted directly from home

☐Transfer from other (i.e. Spinal Cord Injury Unit or other facility not listed)

☐Acute Care Hospital (inpatient status only)

☐Transfer from outside Emergency Department

☐Nursing Home/Chronic Care Facility/Intermediate Care Unit

☐Unknown (if transferred from unknown location or Facility)

Hospital Admission Date: _____/_____/_____ *Operation Date: _____/_____/_____

Anesthesia Technique:

☐ General ☐Spinal ☐ Epidural ☐ Regional ☐ Local ☐ MAC ☐None ☐Other ☐Unknown

*Surgical Specialty: *(select one)*

1. General Surgery	3. Thoracic	5. Orthopedics	7. Urology	9. Plastics
2. Vascular	4. Cardiac	6. Neurosurgery	8. Otolaryngology (ENT)	10. Gynecology

Attending Surgeon's Name: _____ Attending Surgeon's IDN: _____

LCN: _____ Encounter Number: _____

PRE-OPERATIVE RISK ASSESSMENT

GENERAL				RENAL		
Height	_____	*Inches*	*CM*	Acute Renal Failure w/in 24 hrs	*YES*	*NO*
Weight	_____	*Pounds*	*KG*	Currently requiring or on Dialysis w/in 2 wks	*YES*	*NO*
Diabetes Mellitus	*Non-Insulin*	*Insulin*	*NO*	NUTRITIONAL/IMMUNE/OTHER		
Current Smoker w/in 1 year		*YES*	*NO*	Disseminated Cancer	*YES*	*NO*
Dyspnea	*Mod. Exertion*	*At Rest*	*NONE*	Open Wound (w/ or w/out infection)	*YES*	*NO*
Functional Health Status prior to surgery	*I___ PD ____*	*TD ___*	*Unk ___*	Steroid use for chronic condition	*YES*	*NO*
				>10% loss of body wt. last 6 months	*YES*	*NO*
PULMONARY						
		YES	*NO*	Bleeding disorders	*YES*	*NO*
Vent. Dependent w/in 48 hrs						
		YES	*NO*	Preop Transfusions (RBC units w/in 72 hrs)	*YES*	*NO*
COPD (severe)						
				Sepsis w/in 48 hours	*SIRS*	*NO*
HEPATOBILIARY						
		YES	*NO*		*Sepsis*	
Ascites w/in 30 days						
CARDIAC					*Sep Shock*	
CHF w/in 30 days		*YES*	*NO*			
Hypertension req. meds.		*YES*	*NO*			

LABORATORY DATA

LABORATORY DATA: (report preop lab values closest to the Procedure/Surgery start date & time)

Preop values should be within 90 days prior to surgery

PRE-OPERATIVE LAB DATA	Value90 days	unknown	Date
Serum Sodium (Na)		☐	____/____/____

Blood Urea Nitrogen (BUN)		☐	____/____/____
Creatinine (Cr)		☐	____/____/____
Albumin (ALB)		☐	____/____/____
Total Bilirubin (TB)		☐	____/____/____
Serum Glutamic-Oxaloacetic Transaminase (SGOT)/(AST)		☐	____/____/____
Alkaline Phosphatase (AlkPhos)		☐	____/____/____
White Blood Count (WBC)		☐	____/____/____
Hematocrit (Hct)		☐	____/____/____
Platelets (Plt)		☐	____/____/____
Prothrombin Time (PT)		☐	____/____/____
Internat'l Normalised Ratio (INR)		☐	____/____/____

OPERATIVE INFORMATION

Emergency Case: ☐ YES ☐ NO

Wound Classification: ☐ *Clean* ☐ *Clean/Contaminated* ☐ *Contaminated*
☐ *Dirty/Infected*

ASA Class *(circle one)*: *1 2 3 4 5 6 None Assigned(for local anes. only)*

OPERATIVE TIMES: Procedure / Surgery Start: _____:_____ Procedure/Surgery Finish: _____:_____

ADDITIONAL OPERATIVE PROCEDURES

Other Procedure	Concurrent Procedure		CPT
1.	1.		
2.	2.		
3.	3.		
4.	4.		
5.	5.		
6.	6.		
7.	7.		

OCCURRENCES

POST-OPERATIVE OCCURRENCES: ☐ *YES* ☐ *NO*
(Although not required for this program, you may wish to document 'treatment' and 'outcome to date' of the occurrence for internal quality monitoring)

Date *Treatments / Outcomes / Comments*

Wound Occurrences

Superficial Incisional SSI _____/_____/_____

_____

Present at Time of Surgery? ☐ *YES* ☐ *NO*

Deep Incisional SSI _____/_____/_____

_____

Present at Time of Surgery? ☐ *YES* ☐ *NO*

Organ/Space SSI _____/_____/_____

_____

Present at Time of Surgery? ☐*YES* ☐*NO*

Wound Disruption

_____/_____/_____

Respiratory Occurrences

Pneumonia (PNA) _____/_____/_____

_____

Present at Time of Surgery? ☐*YES* ☐*NO*

Unplanned Intubation _____/_____/_____

_____

Pulmonary Embolism _____/_____/_____

_____

On ventilator > 48 hours _____/_____/_____

_____

Present at Time of Surgery? ☐*YES* ☐*NO*

Urinary Tract Occurrences

Progressive Renal Insufficiency _____/_____/_____

_____

Acute Renal Failure _____/_____/_____

_____

Urinary Tract Infection (UTI) _____/_____/_____

_____

Present at Time of Surgery? ☐*YES* ☐*NO*

CNS Occurrences

Stroke/CVA _____/_____/_____

_____

Cardiac Occurrences

Cardiac Arrest req. CPR _____/_____/_____

_____

Myocardial Infarction _____/_____/_____

_____

Other Occurrences

Bleeding Requiring Transfusion (72h of surgery start time)

(transfusion of 1-200 units) _____/_____/_____ # of units transfused: _____

Deep Vein Thrombosis (DVT) req. Therapy _____/_____/_____

_____

Sepsis: Sepsis _____/_____/_____

_____

Septic Shock _____/_____/_____

_____

Other Post-operative Occurrences (ICD-9 code): _____/_____/_____ (ICD-9 code)

_____

HOSPITAL DISCHARGE INFORMATION / READMISSIONS / MORTALITY / REOPERATIONS

Discharge Destination:

☐*Chronic Care Facility, not Home*
☐*Unskilled Facility, not Home*
☐*Facility which was Home*
☐*Home*
☐*Separate Acute care (transferred to another acute care facility)*
☐*Rehab*
☐*Expired*
☐*Unknown*

Post-op ICD.9 Code _____ Diagnosis:

_____

Readmisson:

Readmission for any reason within 30 days of the principle procedure? ☐*YES* ☐*NO* If yes, date: _____/_____/_____

Information Source*(select one)* ☐ Medical Record ☐ Patient/Family Report ☐ Other

Was this readmission unplanned at the time of the principle procedure? ☐*YES* ☐*NO*

Select the primary reason for the unplanned readmission from the post-operative occurrences:

Superficial SSI	Pulmonary Embolism	Coma > 24 hours	Deep Vein Thrombosis (DVT)
Deep SSI	On ventilator > 48 hours	Peripheral Nerve Injury	Sepsis
Organ / Space SSI	Progressive Renal Insufficiency	Cardiac Arrest req CPR	
Wound Disruption	Acute Renal Failure	Myocardial Infarction	
Pneumonia	Urinary Tract Infection (UTI)	Bleeding Requiring Transfusion	Other: ICD-9 code_____
Unplanned Intubation	Stroke / CVA	Graft / Prosthesis / Flap Failure	

Notes ~ If ICD-9 is unknown, describe the reason.

Was this readmission for a post operative occurrence likely related to the principle surgical procedure?
☐*YES* ☐*NO*

Still in hospital > 30 days: ☐*YES* ☐*NO*

Hospital Discharge Date: _____/_____/_____ *(mm/dd/yyy)*

Post-operative Death:

Postop Death w/in 30 days:	☐*YES* ☐*NO*	Postop Death > 30 days: *(if remained in acute care)*	☐*YES* ☐*NO*
Date of death: _____/_____/_____	☐ Unknown	Date of death: _____/_____/_____	☐ Unknown

Unplanned Reoperation:
Unplanned return to the operating room for a surgical procedure w/in30 day post-operative period? ☐*YES* ☐*NO*

Was the return to the OR for a postop occurrence likely related to the principle procedure, or to any additional surgery performed under the same anesthetic as the principle procedure? ☐*YES* ☐*NO*

If yes, Surgery Date _____/_____/_____ CPT code _____ ICD9 code _____

Source *(select one)* ☐ Medical Record ☐ Patient/Family Report ☐ Other

Notes ~ If CPT code is not documented, describe the surgery.

Was there a SECOND unplanned reoperation within 30 days? ☐*YES* ☐*NO*

Was the second return to the OR for a post-operative occurrence that was likely related to the principle procedure, or to any additional surgery (i.e., 'other' or 'concurrent') performed under the same anesthetic as the principle procedure? ☐*YES* ☐*NO*

If yes: Surgery Date _____/_____/_____ CPT code _____ ICD9 code _____

Source *(select one)* ☐ Medical Record ☐ Patient/Family Report ☐ Other

Notes ~ If CPT code is not documented, describe the surgery.

Were there more than two unplanned reoperations for an adverse outcome related to the principal surgery within 30 days?
☐*YES* ☐*NO*

FOLLOW-UP

Follow Up Within 30 Days

Were you able to follow the case for the full 30 days? ☐*YES*☐*NO*
(NOTE: Answer yes for death w/in 30 days)

If you were unable to obtain the 30-day follow up information:

A) How many days (0-29) were you able to follow this case? _____

B) Which attempt methods were used for follow-up? Select all that apply.
(A minimum of three attempts should be made to contact the patient)

Method	# of attempts	Method
☐Phone	_____	☐Documentation
☐Letter	_____	☐Other

Patient Contact Management:

Contact date:_____/_____/_____ Contact Action: ☐Phone ☐Letter ☐Document ☐Fax ☐E-mail ☐Other

Contact Results:

☐No Answer	☐Letter Received	☐Incorrect Number
☐Left Message	☐Talked to Patient	☐Patient Refused
☐Letter Sent	☐Talk to Family	

Contact Notes:

Appendix 2

Data dictionary for the procedure-targeted program

Field Label	Data Type	Field Note
Study ID	Continuous	
Patients name	Free text	Surname, name
Hospital number	Continuous	Hospital prefix not needed
Hospital	Categorical: 1, Groote Schuur | 2, Mitchell's Plain | 3, New Somerset | 4, Victoria	
Date subject signed consent	Date	DD-MM-YYYY
Date of admission	Date	DD-MM-YYYY
Age (years)	Continuous	
Race	Categorical: 0, Black | 1, White | 2, Coloured | 3, Indian | 4, Other	
Gender	Binary: 0, Female | 1, Male	
Individual comorbidities	Categorical: 1, Heart disease | 2, Cancer | 3, Liver disease | 4, Stroke | 5, Hypertension | 6, Renal disease | 7, Impaired sensorium | 8, Respiratory disease | 9, Bleeding disorder | 10, Peripheral vascular disease | 11, Steroid use | 12, Nil known	Multiple choices permitted
Functional health status prior to surgery	Categorical: 1, Independent | 2, Partially dependent | 3, Totally dependent | 4, Unknown	
HIV status	Categorical: 1, HIV negative | 2, HIV positive | 3, AIDS | 4, HIV status unknown	
If HIV positive, CD4 count	Categorical: 1, >350 | 2, <350 | 3, unknown	
Diabetes mellitus	Categorical: 1, Non-insulin | 2, Insulin | 3, No	
Smoking status	Binary: Yes | No	
Emergency CT scan performed	Binary: Yes | No	
Height (cm)	Continuous	
Weight (kilograms)	Continuous	
BMI	round(([weight2]*10000)/(([height2])^(2)),1)	
Open wound (w/ or w/out infection)		
Serum Albumin	Continuous	Preop values should be within 90 days prior to surgery
Creatinine	Continuous	Preop values should be within 90

		days prior to surgery
Urea	Continuous	Preop values should be within 90 days prior to surgery
Haematocrit (HCT)	Continuous	Preop values should be within 90 days prior to surgery
International Normalised Ratio (INR)	Continuous	Preop values should be within 90 days prior to surgery
Initial pH	Continuous	From earliest ABG done on arrival
Initial Bicarb	Continuous	From earliest ABG done on arrival
initial base excess	Continuous	From earliest ABG done on arrival
Initial lactate	Continuous	From earliest ABG done on arrival
Date of operation	Date	
ASA	Categorical: 1, 1 | 2, 2 | 3, 3 | 4, 4 | 5, 5 | 7, not recorded	
Wound classification	Categorical: 1, Clean | 2, Clean/ Contaminated | 3, Contaminated | 4, Dirty/ Infected	
Time of start operation (knife-to-skin)	Time	
Time of completion operation	Time	last recorded time entry on anaes note
Surgical safety checklist used	Binary: Yes | No	eg. WHO Surgical Safety Checklist
Training of most senior surgeon	1, Sub-specialist | 2, Fellow | 3, Consultant | 4, Registrar/ Resident | 5, Medical Officer/ House Officer	

present		
Training of most senior anaesthetist present	1, Sub-specialist | 2, Fellow | 3, Consultant | 4, Registrar/ Resident | 5, Medical Officer/ House Officer	
Anaesthetic type	1, General Anaesthetic | 2, Regional Anaesthetic | 3, Local Anaesthetic +- sedation (eg. ketamine)	
Incision	1, Midline | 2, Transverse | 3, Rooftop | 4, Right subcostal (Kocher) | 5, Right iliac fossa (Gridiron/ Lanz) | 6, Paramedian | 7, Laparoscopic (+- open specimen extraction) | 8, Laparoscopic converted to open | 9, Other (please specify)	
Incision if not listed	Free text	Only use if reasonable option not available above
Primary operation performed	1, Abdomen: Laparotomy with no other procedure | 2, Abdomen: Diagnostic laparoscopy with no other procedure | 3, Abdomen: Division of adhesions of peritoneum | 4, Abdomen: Repair of anterior abdominal wall | 5, Abdomen: Closure of gastroschisis/ exomphalos | 6, Oesophagus: Excision of oesophagus | 7, Oesophagus: Repair of oesophagus | 8, Oesophagus: Other operations on oesophagus | 9, Stomach: Total excision of stomach | 10, Stomach: Partial excision of stomach | 11, Stomach: Connection of stomach to jejunum | 12, Stomach: Operations on ulcer of stomach | 13, Stomach: Other repair of stomach | 14, Stomach: Incision of pylorus | 15, Stomach: Other open operations on stomach | 16, Duodenum: Operations on ulcer of duodenum | 17, Duodenum: Correction of malrotation | 18, Duodenum: Other open operations on duodenum | 19, Small bowel: Excision of small bowel | 20, Small bowel: Bypass of small bowel | 21, Small bowel: Excision of Meckel's diverticulum | 22, Small bowel: Reduction of intussuception without excision | 23, Small bowel: Formation of ileostomy | 24, Small bowel: Closure of perforation | 25, Small bowel: Other open operations on small bowel | 26, Appendix:Emergency excision of appendix | 27, Colon: Total excision of colon and rectum | 28, Colon: Total excision of colon | 29, Colon: Extended excision of right hemicolon | 30, Colon: Excision of right semicolon | 31, Colon: Excision of transverse colon | 32, Colon: Excision of left semicolon | 33, Colon: Excision of sigmoid colon | 34, Colon: Other excision of colon | 35, Colon: Reduction of intussuception/volvulus without excision | 36, Colon: Formation of any colonic stoma | 37, Colon: Other open operations on colon | 38, Rectum: Excision of rectum | 39, Rectum: Fixation of rectum for prolapse | 40, Rectum: Other open operations on rectum | 41, Liver: Partial excision of liver | 42, Liver: Repair of liver, including liver packing | 43, Liver: Other open operations on liver | 44, Gallbladder: Excision of gall bladder | 45, Gallbladder: Other open operations on gall bladder | 46, Bile duct: Repair of bile duct | 47, Bile duct: Incision of bile duct | 48, Bile duct: Other open operations on bile duct | 49, Pancreas: Excision of head of pancreas | 50, Pancreas: Open drainage of lesion of pancreas | 51, Pancreas: Open drainage of lesion of pancreas | 52, Pancreas: Other open operations on pancreas | 53, Spleen: Total excision of spleen | 54, Spleen: Other open operations on spleen | 55, Aorta/vessels: Any primary abdominal vascular operation | 56, Kidney: Total excision of kidney | 57, Kidney: Partial excision of kidney | 58, Kidney: Open repair of kidney | 59, Kidney: Other operations on kidney | 60, Ureter:	Single most appropriate operation actually performed

	Repair of ureter | 61, Ureter: Other open operations on ureter | 62, Bladder: Repair of bladder | 63, Bladder: Other open operations on bladder | 64, Uterus: Abdominal excision of uterus | 65, Uterus: Other open operations on uterus | 66, Ovary: Bilateral excision of ovary/ tube | 67, Ovary: Unilateral excision of ovary/ tube | 68, Ovary: Other open operations on ovary/ tube | 69, Diaphragm: Repair of rupture of diaphragm | 70, Diaphragm: Other operations on diaphragm | 71, Other abdominal procedure (please specify)	
Primary operation performed if not listed	Free text	Only use if reasonable option not available above
Bowel resection performed	Categorical: 1, No | 2, Yes-handsewn anastomosis | 3, Yes-stapled anastomosis (+-hand sewn reinforcement) | 4, Yes-stoma without anastomosis	Only if complete transection through bowel lumen. If appendectomy, answer 'no'
Stoma formed?	Categorical: 1, No | 2, Loop ileostomy | 3, loop colostomy | 4, End ileostomy | 5, End colostomy | 6, Other	
Diagnosis	Categorical: 1, Neoplasm: any malignant (cancer) | 2, Neoplasm: any benign | 3, Infection: Typhoid/ paratyphoid | 4, Infection: Other infectious gastroenteritis/ colitis | 5, Infection: HIV disease | 6, Infection: Abdominal TB | 7, Infection: Other | 8, Perforation of oesophagus | 9, Peptic ulcer: bleeding | 10, Peptic ulcer: perforation | 11, Peptic ulcer: bleeding and perforation | 12, Peptic ulcer without bleeding or perforation | 13, Appendicitis | 14, Hernia: any abdominal hernia | 15, Colitis/ gastroenteritis: Chrons disease | 16, Colitis/ gastroenteritis: Ulcerative colitis | 17, Colitis/ gastroenteritis: Other noninfectiveicl. ischaemic bowel | 18, Bleeding: small bowel/ colon with no malignancy | 19, Diverticular disease | 20, Stercoral perforation of colon | 21, Adhesion: no bowel obstruction | 22, Intestinal obstruction: Adhesions | 23, Intestinal obstruction: Intussusception | 24, Intestinal obstruction: Volvulus | 25, Cholelithiasis/ cholecystitis (gallstones) | 26, Acute pancreatitis | 27, Female reproductive: salpingitis/ oophoritis | 28, Female reproductive: ectopic pregnancy | 29, Female reproductive: Abnormal uterine/ vaginal bleeding | 30, Female reproductive: Post-partum haemorrhage | 31, Congenital: Pyloric stenosis | 32, Congenital: Diaphragmatic hernia | 33, Congenital: Meckel diverticulum | 34, Congenital: Gastroschisis | 35, Congenital: Other | 36, Complication of previous surgical operation/ procedure | 37, No disease identified | 38, Other diagnosis (please specify; please try avoid using)	Most appropriate diagnosis following surgery, not before. Wait for pathology result if necessary.
Diagnosis if not listed	Free text	Only use if reasonable option not available above
Was perforation of abdominal organ found at operation?	Binary: Yes | No	Do not include perforations which are

		sustained during this operation
Prophylactic antibiotics given?	Binary: Yes | No	
Type of prophylaxis	Time	antibiotics name
Whole blood or blood product(s) used?	Categorical: 1, No | 2, Yes, whole blood | 3, Yes, blood products e.g. packed red cells, plasma, platelets	
Thromboprophylaxis used?	Categorical: 1, No | 2, Yes, drug only | 3, Yes, mechanical only | 4, Yes, drug and mechanical | 5, Yes, other	
Post operative occurrence	Binary: Yes | No	
Wound occurrences	Categorical: 1, Superficial Incisional SSI | 2, Deep incisional SSI | 3, Organ/ Space SSI | 4, Wound disruption	
Respiratory occurrences	Categorical: 1, Pneumonia | 2, Unplanned intubation | 3, Pulmonary embolus | 4, On ventilator>48hrs	More than one allowed
Urinary tract occurrences	Categorical: 1, Acute renal failure not requiring dialysis | 2, Acute renal failure requiring dialysis | 3, Progressive renal insufficiency | 4, Urinary tract infection	
CNS occurrences	Categorical: 1, Transient Ischaemic Attack | 2, Stroke/ CVA	
Cardiac occurrences	Categorical: 1, Cardiac arrest req. CPR | 2, Myocardial infarction	
Other occurrences	Categorical: 1, Bleeding req transfusion (72 hours of surgery start time) | 2, Deep Vein Thrombosis req. Therapy | 3, Sepsis | 4, Septic shock	
Other complications	Free text	
Post operative admission to ICU?	Binary: Yes | No	
ICU length of stay	Continuous	Days
Discharge destination	Categorical: 1, Chronic care facility | 2, Rehab facility | 3, Step-down acute care facility | 4, Referral/ academic acute care facility | 5, Home | 6, Expired	
Discharge date	Date	DD-MM-YYYY
Readmission for any reason within 30 days of principal procedure?	Binary: Yes | No	
Was this readmission unplanned at the time of the principle procedure?	Binary: Yes | No	
Select the primary reason for unplanned readmission from the post-operative	Categorical: 1, Superficial SSI | 2, Deep SSI | 3, Organ/ Space SSI | 4, Wound disruption | 5, Pneumonia | 6, Unplanned intubation | 7, Pulmonary embolus | 8, On ventilator>48 hrs | 9, Progressive renal insufficiency | 10, Acute renal failure | 11, Urinary tract infection (UTI) | 12, Stroke/ CVA | 13, Coma> 24 hours | 14, Peripheral nerve injury | 15, Cardiac arrest req. CPR | 16, Myocardial infarction | 17, Bleeding	Primary reason, only 1 option allowed

occurrences:	requiring transfusion | 18, Graft/ Prosthesis/ Flap failure | 19, Deep vein thrombosis | 20, Sepsis | 21, Other	
Post-operative death	Categorical: 1, Yes | 2, No | 3, Unknown	
Date of death	Date	DD-MM-YYYY
Still in hospital > 30 days?	Binary: Yes | No	
Unplanned return to the operating room for a surgical procedure w/in 30 day post-operative period?	Binary: Yes | No	
Type of unplanned reoperation	Free text	
Date of unplanned reoperation	Date	DD-MM-YYYY
Was there a second unplanned reoperation within 30 days?	Binary: Yes | No	
Were there more than two unplanned reoperations for an adverse outcome related to the principal surgery within 30 days?	Binary: Yes| No	

Appendix 3

Cape Metro West Surgical Quality Improvement Programme
Patient information and consent

To whom it may concern,

You are receiving this document as you have recently had surgery and we would like to ask whether you will consider taking part in a research study. The purpose of the research study is to measure the type of surgeries patients have within our health district, the care patients receive after surgery, if patients develop any problems after surgery, and whether the surgery and the care of the surgery are of good quality. This is part of a drive to make sure we are offering our patients the best care we can within the Cape Metro West health district.

If you agree to take part in the study, we will collect data from your hospital folder regarding your health and the operation you had. We will check your folder for information before and after the surgery, and during your follow-up appointments. We will also call you one month after the date of your surgery to discuss your health after the surgery. The only other request for taking part in this study is to check that we have your correct contact details before you leave hospital.

Your surgical care and follow-up appointments after your surgery will remain unchanged. You will not have to make any extra visits to hospital because of the research study. It is important to note that this study does not replace the normal care after your surgery, and you should still attend the follow-up appointments arranged by the doctor or nurse. There are no risks associated with this study. There are no direct benefits to you for taking part in this study. There is also no payment for taking part in this study. We hope that the findings of this research study may help to improve the service and care we offer to future patients by allowing us to measure our surgical outcomes more closely. The data we collect will remain confidential by using a code instead of your name for the information we collect. Your privacy will be maintained as we will not share your personal information with anybody. The findings of the study may be published in a scientific journal or discussed at meetings, but no individual participants will be identified.

If there are any further questions, please contact the Principal Investigator Dr Richard Spence on 0613001212 or spnric004@myuct.ac.za
If you have any questions or concerns about your rights or welfare as a research participant, please contact Professor Marc Blockman, Chairperson of the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee on 021 406 6626 or marc.blockman@uct.ac.za.

Consent statement

I have read the above information. I have had the chance to ask questions about it and any questions that I have asked have been answered to my satisfaction. I consent voluntarily to take part as a participant in this study and I understand that I can withdraw at any time from the study without this in any way affecting my current or future medical care.

Participant signature: _____
Date: _____
Investigator signature: _____

Appendix 4

Validation of the GSH Trauma Prediction rules

Validation against the Revised Trauma Score

The built-in algorithm in eTHR calculated a RTS score for the 6,934 (92.9%) patients in the derivation dataset on whom the necessary data was completed. Table 4.1a presents a comparison of the summary statistics for both RTS and GSTS in the derivation and validation datasets.

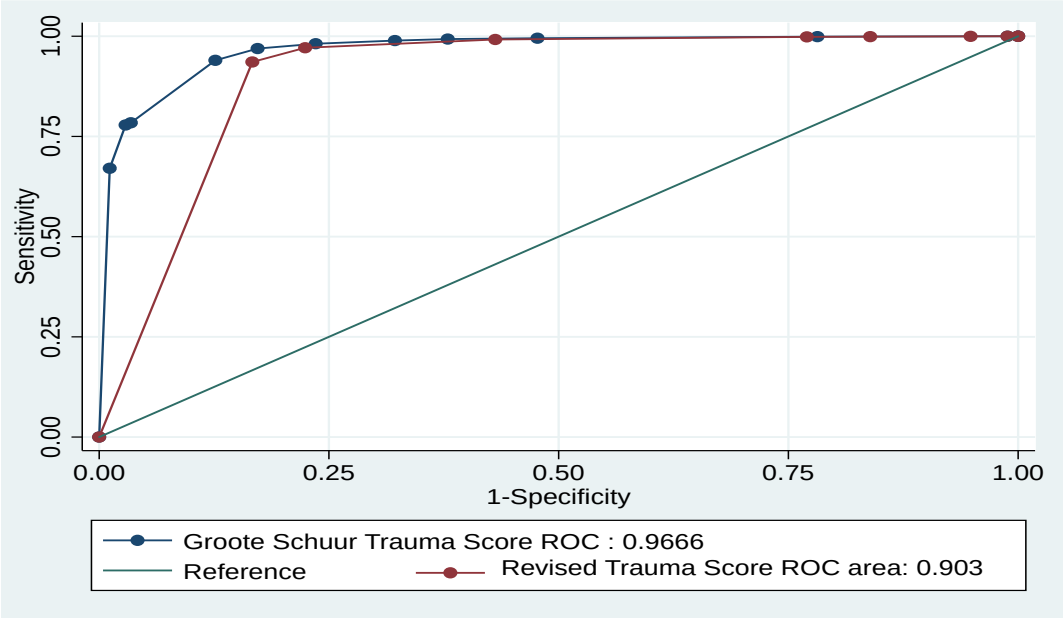
Table 4.1a Summary statistics for the Revised Trauma Score and the Groote Schuur Trauma Score in the Trauma Surgery derivation and validation datasets

Score	Frequency (%)	Median	IQR	Estimated probability of survival range	Missing (%)	P-Value*
RTS Derivation	6934 (92.95)	7	7 - 7.8408	0.027 - 0.988	526 (7.05%)	
GSTS Derivation	6935 (92.96)	11	11 - 12	0.042 - 0.999	525 (7.04%)	0.98
RTS Severely injured	1,687 (95.31)	7	7 - 7.8408	0.027 - 0.988	83 (4.69%)	
GSTS Severely injured	1,681 (94.97)	9	9 - 12	0.044 - 0.999	89 (5.03)	0.33
RTS Operative	877 (92.32)	7	7 - 7.8408	0.172 - 0.988	73 (7.68)	
GSTS Operative	879 (92.53)	9	9 - 12	0.127 - 0.999	71 (7.47)	0.62

P-value derived after probability proportion test comparing rates of missing data*

As an indicator of data burden, the proportion of missing data for each score was compared, but this was not significantly different in any of the datasets, indicating similar data burden was required for both scores. The ability of the two scores to discriminate between survivors and non-survivors in each dataset is compared in Figures 4.1a, 4.2a and 4.3a.

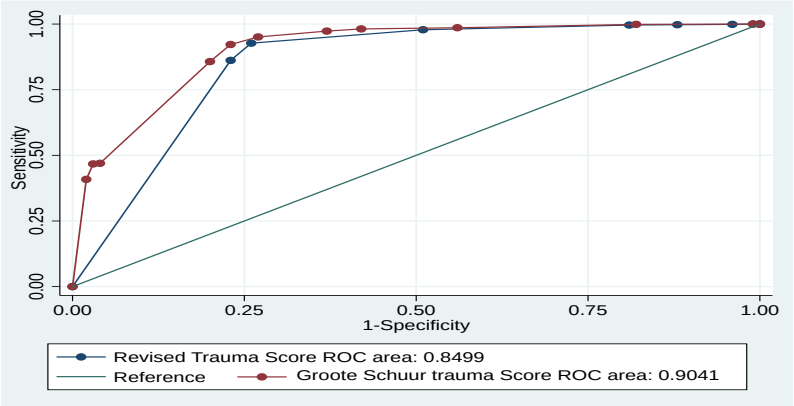
Figure 4.1a A Comparison of RTS and GSTS by ROC curves in the derivation dataset



P-Value <0.0001

The ROC for the GSTS was 0.966 (95% CI 0.95362 - 0.97958) compared to 0.9030 (95% CI 0.8739 – 0.9321) for the RTS. This difference was significant (p<0.0001). The GOF for the GSTS was 2.98 (p = 0.088), but there were insufficient degrees of freedom to generate a GOF for the RTS.

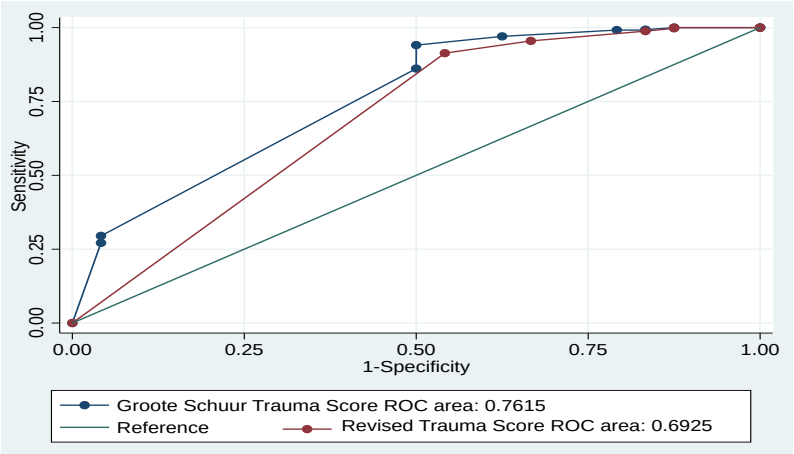
Figure 4.2a A Comparison of RTS and GSTS by ROC curves in the severely injured validation dataset



P-Value<0.0001

The ROC for the GSTS was 0.9041 (95% CI 0.87048 – 0.93777) compared to 0.8499 (95% CI 0.80425 – 0.89550) for the RTS and this difference was significant (p<0.0001). The GOF for the GSTS was 3.16 (p = 0.2057) in the severely injured validation. The GOF for the RTS could not be determined.

Figure 4.3a A Comparison of RTS and GSTS by ROC curves in the Operative validation dataset



P = 0.0150

The ROC for the GSTS was 0.7615 (95% CI 0.66395 – 0.85895) compared to 0.6925 (95% CI 0.58720 – 0.79778) for the RTS. This difference was statistically significant (p = 0.015). The GOF for the GSTS was 0.09 (p = 0.7661) in the Operative validation dataset. The GOF for the RTS could not be determined.

Validation against the Kampala Trauma Score

The built-in algorithm in eTHR calculated a KTS score for 6,485 (86.93%) patients in the derivation dataset on whom the necessary data was completed. Table 4.2a presents a comparison of the summary statistics for both KTS and GSTS in the derivation and validation datasets.

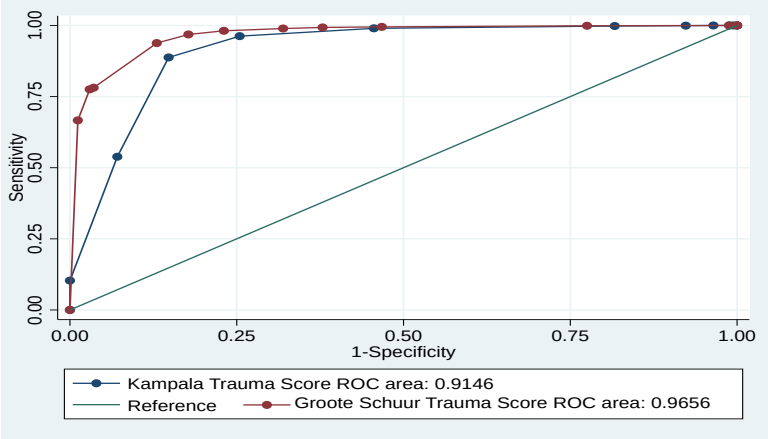
Table 4.2a Summary statistics for the Kampala Trauma Score and the Groote Schuur Trauma Score in the Trauma Surgery derivation and validation datasets

Score	Frequency (%)	Median	IQR	Predicted probability of survival range	Missing (%)	P-Value*
KTS Derivation	6,485 (86.93)	15	14 - 15	0.0007 - 0.999	975 (13.07)	
GSTS Derivation	6935 (92.96)	11	11 - 12	0.042 - 0.999	525 (7.04%)	<0.0001
KTS Severely injured	1,614 (91.19)	14	14 - 15	0.019 - 0.997	156 (8.81)	
GSTS Severely injured	1,681 (94.97)	9	9 - 12	0.044 - 0.999	89 (5.03)	<0.0001
KTS Operative	832 (87.58)	14	14 - 15	0.3685 - 0.9625	118 (12.42)	
GSTS Operative	879 (92.53)	9	9 -12	0.127 - 0.999	71 (7.47)	<0.0001

P-value derived after proportion test comparing rates of missing data*

As an indicator of data burden, the proportion of missing data for each score was compared. There was a significantly higher proportion of missing KTS data in all three datasets (p<0.0001). The ability of the two scores to discriminate between survivors and non-survivors in each dataset are compared in Figures 4.4a-4.6a.

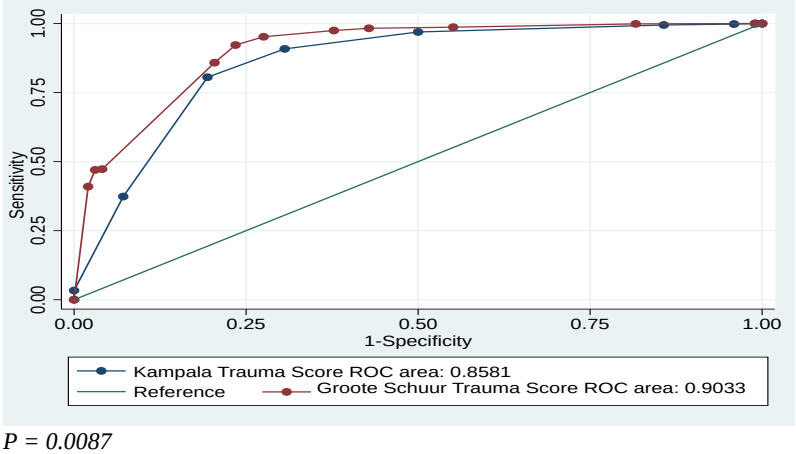
Figure 4.4a Comparison of KTS and GSTS by ROC curves in the derivation dataset



P-Value <0.0001

The ROC for the GSTS was 0.9656 (95% CI 0.9521 – 0.9790) compared to 0.9146 (95% CI 0.88724 – 0.94197) for the KTS. This difference was statistically significant (p<0.0001). The GOF for the GSTS was 2.98 (p = 0.088) and the GOF for the KTS was 5.49 (p = 0.019).

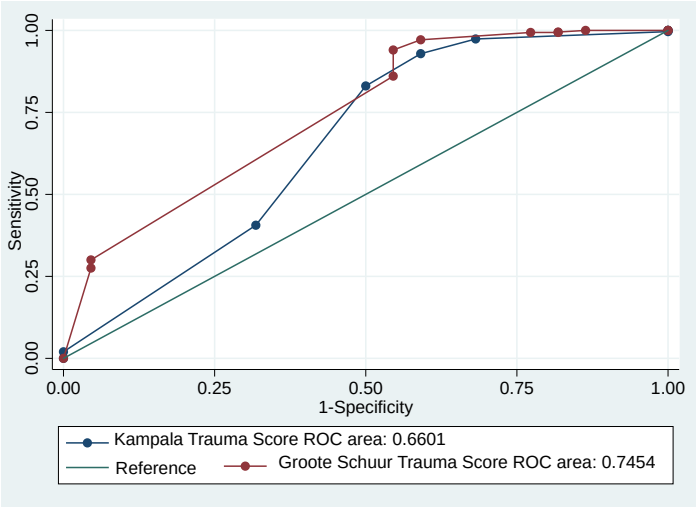
Figure 4.5a Comparison of KTS and GSTS by ROC curves in the Severely injured validation dataset



P = 0.0087

The ROC for the GSTS was 0.9033 (95% CI 0.0.8692 – 0.9375) compared to 0.8581 (95% CI 0.81344 – 0.90269) for the KTS. This difference was statistically significant (p = 0.0087). The GOF for the GSTS was 3.16 (p = 0.2057) in the-severely injured validation and the GOF for the KTS was 1.75 (p = 0.1853).

Figure 4.6a Comparison of KTS and GSTS by ROC curves in the Operative validation dataset



P = 0.1566

The ROC for the GSTS was 0.7454 (95% CI 0.64234 – 0.84843) compared to 0.6601 (95% CI 0.51792 – 0.80225) for the KTS. This difference was not statistically significant (p = 0.1566). The GOF for the GSTS was 0.09 (p = 0.7661) in the operative validation dataset and the GOF for the KTS was 4.03, which was statistically significant (p = 0.0447).

Validation against the Trauma Injury-Severity Score

Applying the TRISS co-efficients in turn to each dataset, a TRISS probability of survival was generated for each patient with the required completed data. The following

table compares the TRISS probability of survival estimates with those generated using the GSH Injury severity prediction rule.

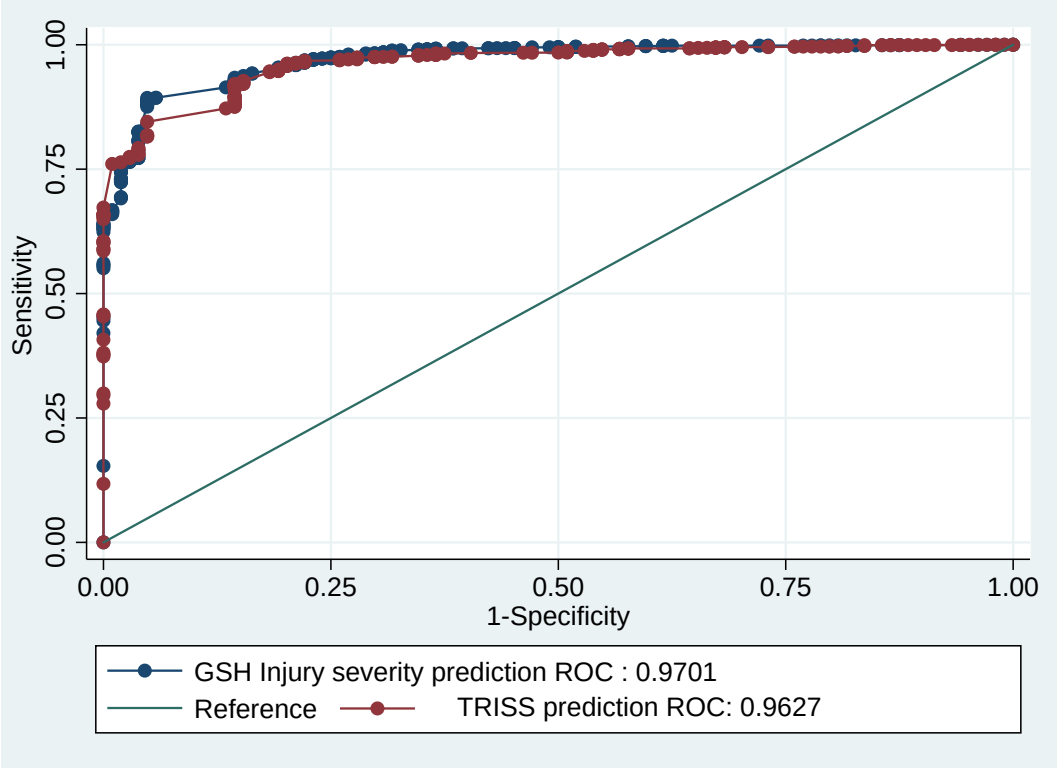
Table 4.7a Summary statistics for the TRISS probability of survival and Groote Schuur Injury severity probability of survival estimates in the Trauma Surgery derivation and validation dataset

Score	Frequency (%)	Median estimated probability of survival	IQR	Estimated probability of survival range	Missing (%)	P-Value*
TRISS Derivation	5,581 (74.89)	0.9873	0.9779 - 0.9941	0.0044 - 0.9946	1,879 (25.19)	
GSH Prediction Derivation	6,082 (81.53)	0.9993	0.9975 - 0.99941	0.0092 - 0.9994	1,378 (18.47)	<0.0001
TRISS Prediction Severely injured	1,641 (92.71)	0.9747	0.9422 - 0.9788	0.0044 - 0.9885	129 (7.29)	
GSH prediction Severely injured	1,693 (95.65)	0.9878	0.9767 - 0.9988	0.1636 - 0.9990	77 (4.35)	<0.0001
TRISS Operative	771 (81.16)	0.9788	0.9649 - 0.9846	0.1693 - 0.9946	179 (18.84)	
GSH Prediction Operative	784 (82.53)	0.9878	0.9762 - 0.9984	0.1693 - 0.9994	166 (17.47)	0.03

P-value derived after probability proportion test comparing rates of missing data*

As an indicator of data burden, the proportion of missing data for each set of predictions was compared. There was a significantly higher proportion of missing TRISS predictions in all three datasets. The ability of the predictions to discriminate between survivors and non-survivors in each dataset are compared in the Figures 4.7a–4.9a.

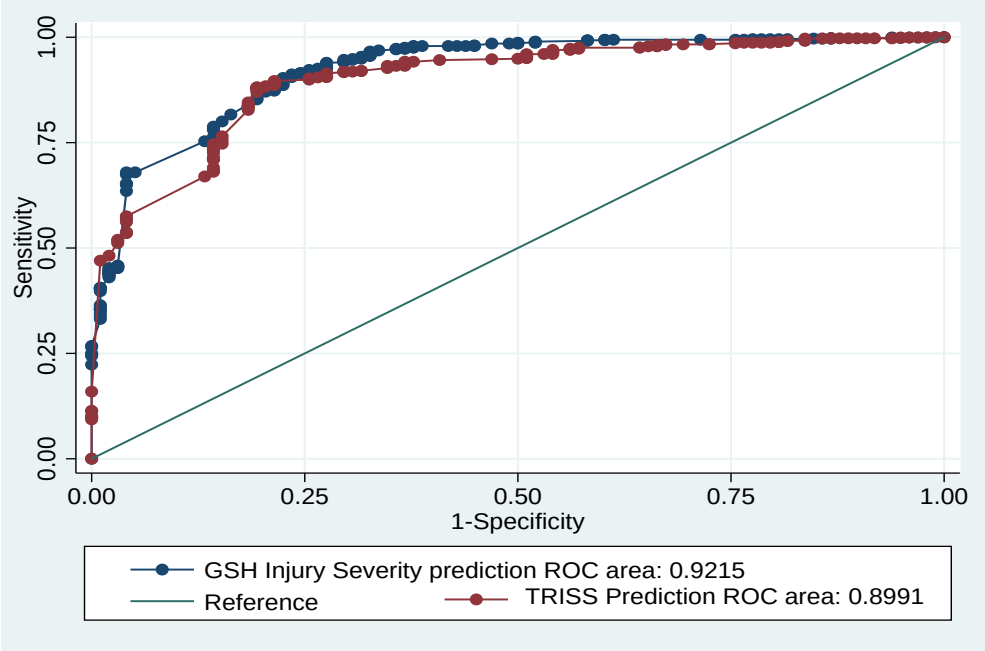
Figure 4.7a Comparison of TRISS and GSH Injury Severity predictions by ROC curves in the derivation dataset



P-Value =0.1609

The ROC for the GSH injury severity prediction was 0.9701 (95% CI 0.9582 – 0.9823) compared to 0.9627 (95% CI 0.9508 – 0.9746) for the TRISS prediction. This difference was not statistically significant (p = 0.1609). The GOF for the GSH Injury Severity prediction was 32.35 (p <0.0001), and the GOF for the TRISS prediction was 54.12 (p<0.0001).

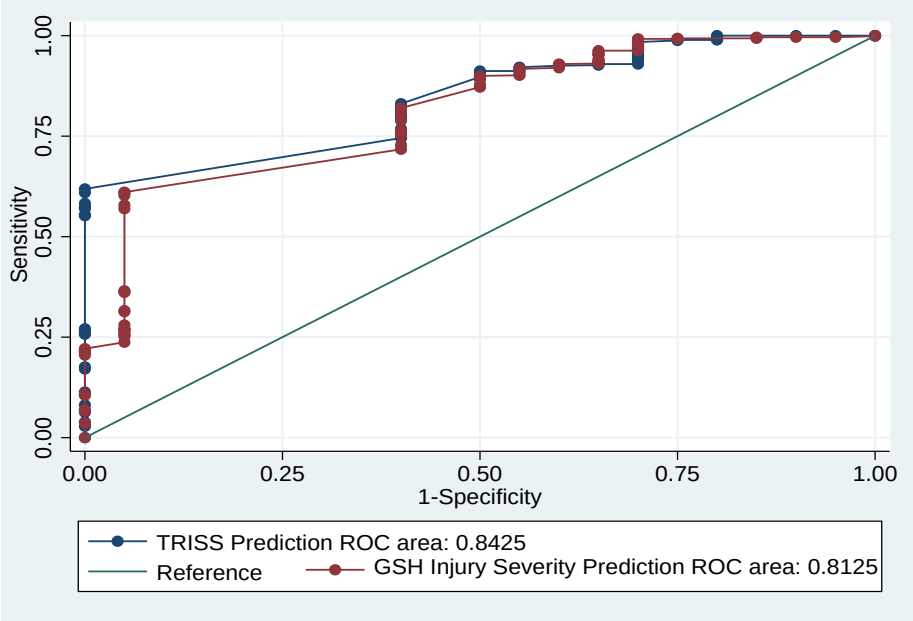
Figure 4.8a Comparison of TRISS and GSH Injury Severity predictions by ROC curves in the severely injured validation dataset.



P = 0.0506

The ROC for the GSH Injury Prediction was 0.9215 (95% CI 0.8938 – 0.9492) compared to 0.8991 (95% CI 0.8694 – 0.9287) for the TRISS. This difference was only of borderline significance (p = 0.0506). The GOF for the GSH Injury Severity prediction was 1.96 (p = 0.1619) in the severely injured-validation and 3.49 (p = 0.0616) for the TRISS prediction.

Figure 4.9a Comparison of TRISS and GSH Injury Severity predictions by ROC curves in the operative validation dataset

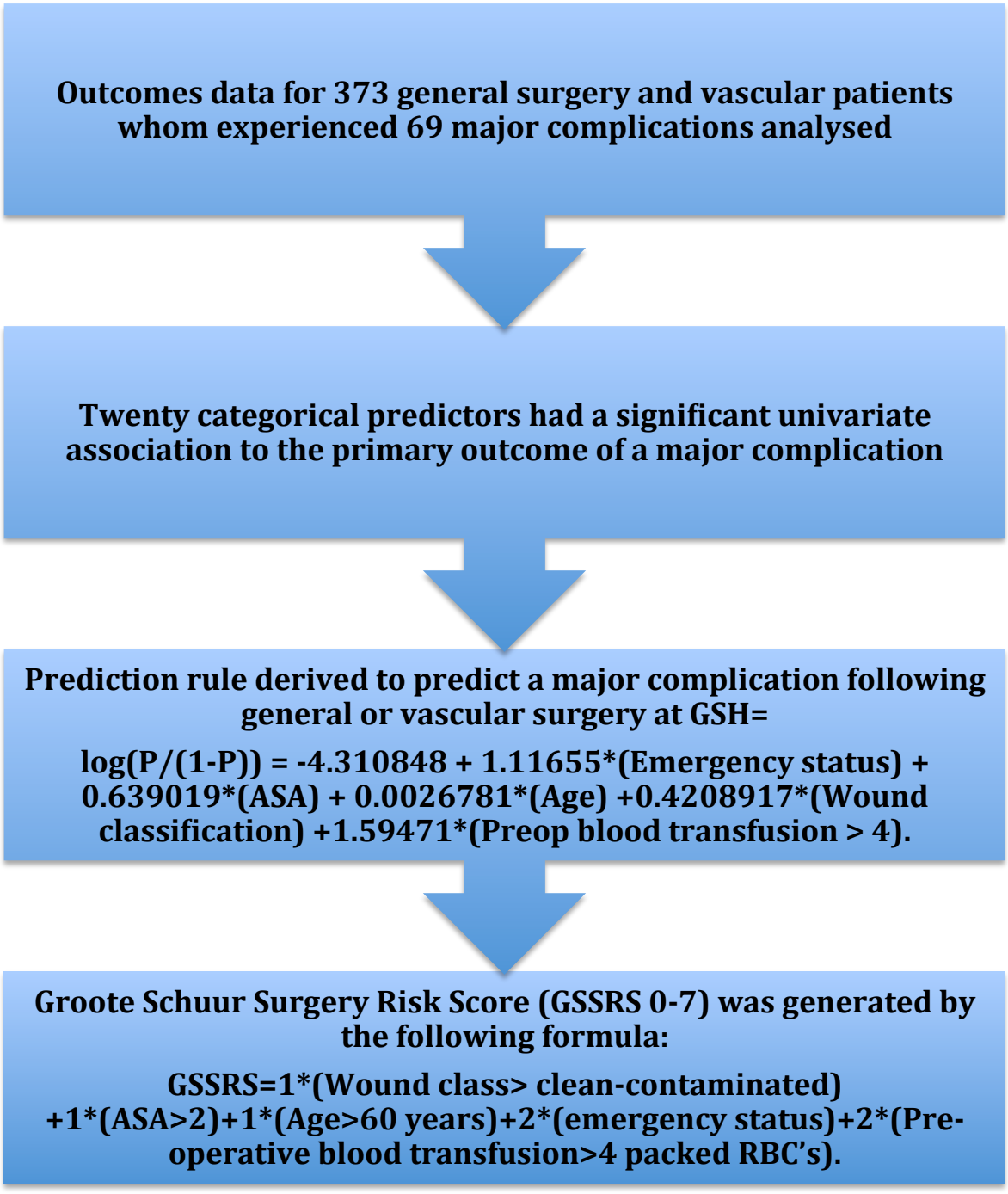


P = 0.1911

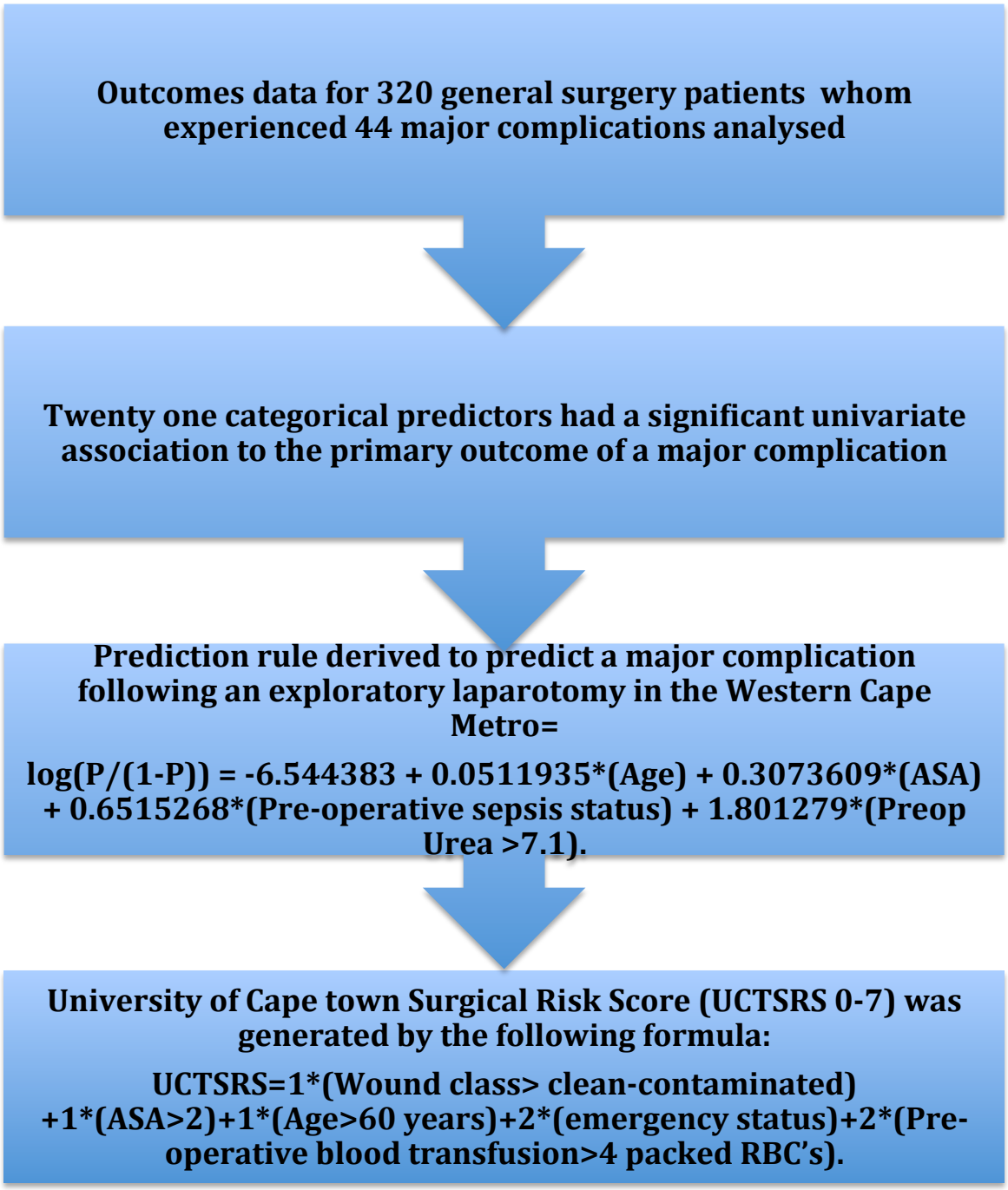
The ROC for the GSH Injury Severity prediction was 0.8125 (95% CI 0.7239 – 0.9012) compared to 0.8425 (95% CI 0.7788 – 0.9062) for the TRISS prediction. This difference was not statistically significant (p = 0.1911). The GOF for the GSH Injury Severity prediction was 10.74 (p = 0.0011) in the Operative validation dataset, and 6.16 for the TRISS prediction**.** This was also not significant (p = 0.0131).

Appendix 5 Summary Flowcharts of all 3 Programmes

Essentials programme



Procedure-targeted programme



Trauma Quality Improvement Programme

Outcomes data for 7,460 trauma patients, including 950 operatively managed patients and 1,770 severely-injured patients analysed

Twenty seven categorical predictors had a significant univariate association to the primary outcome of in-hospital death

Prediction rule derived to predict in-hospital death following trauma in the Western Cape Metro=

log(P/(1-P)) = -2.528225 + 1.025246*(Airway status) +1.316493*(Presence of hypotension) + 3.094101*(Assessed as life threatening injury by admitting physician) -0.7618551*(GCS motor score) + 1.079287*(Age>54 years).

Groote Schuur Hosital Trauma Score (GSTS 1-12) was generated by the following scoring rubric:

Component	Score
Age	
<55	1
≥55	0
Trauma triage colour	
Green/Yellow	3
Red	0
Intubation status	
No	1
Yes	0
Systolic blood pressure (mmHg)	
≥90	1
<90	0
GCS motor	
None	1
Extension to pain	2
Flexion to pain	3
Withdraws from pain	4
Localises pain	5
Obeys commands	6
Total score	1-12

References

(1) Dare AJ, Grimes CE, Gillies R, Greenberg SL, Hagander L, Meara JG, *et al* Global surgery: defining an emerging global health field. Lancet 2014 Dec 20;384(9961):2245-2247.

(2) Bickler SW, Spiegel DA. Global surgery-defining a research agenda. Lancet 2008 Jul 12;372(9633):90-92.

(3) Farmer PE, Kim JY. Surgery and global health: a view from beyond the OR. World J Surg 2008 Apr;32(4):533-536.

(4) Weiser TG, Semel ME, Simon AE, Lipsitz SR, Haynes AB, Funk LM, *et al* In-hospital death following inpatient surgical procedures in the United States, 1996-2006. World J Surg 2011 Sep;35(9):1950-1956.

(5) Abdullah F, Troedsson H, Cherian M. The World Health Organisation program for emergency surgical, obstetric, and anesthetic care: from Mongolia to the future. Arch Surg 2011 May;146(5):620-623.

(6) Meara JG, Leather AJ, Hagander L, Alkire BC, Alonso N, Ameh EA, *et al* Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development. Surgery 2015 May 15.

(7) World Health Organisation. The World Health Organisation program for emergency surgical, obstetric, and anesthetic care. 2015; Available at: http://www.who.int/surgery/en/. Accessed June, 2015, 2015.

(8) Rose J, Weiser TG, Hider P, Wilson L, Gruen RL, Bickler SW. Estimated need for surgery worldwide based on prevalence of diseases: a modelling strategy for the WHO Global Health Estimate. Lancet Glob Health 2015 Apr 27;3 Suppl 2:S13-20.

(9) Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. 1971. Milbank Q 2005;83(4):731-757.

(10) Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, *et al* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2095-2128.

(11) Christianson A. March of dimes, global report on birth defects. The hidden toll of dying and disabled children. 2009; Available at: http://www.marchofdimes.org/materials/global-report-on-birth-defects-the-hidden-toll-of-dying-and-disabled-children-executive-summary.pdf. Accessed April/11, 2015.

(12) Houle B, Clark SJ, Gomez-Olive FX, Kahn K, Tollman SM. The unfolding counter-transition in rural South Africa: mortality and cause of death, 1994-2009. PLOS One 2014 Jun 24;9(6):e100420.

(13) 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 Jan 29;360(5):491-499.

(14) Pearse RM, Rhodes A, Moreno R, Pelosi P, Spies C, Vallet B, *et al*. EuSOS: European surgical outcomes study. Eur J Anaesthesiol 2011 Jun;28(6):454-456.

(15) Biccard BM, Madiba TE *et al*. The South African Surgical Outcomes Study : a seven-day prospective observational cohort study. S Afr Med J. 2015 Jun;105(6):465-75.

(16) 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 Jul 12;372(9633):139-144.

(17) Statistics South Africa. 2013; Available at: http://www.statssa.gov.za/. Accessed June 24, 2015.

(18) Ko CY, Maggard M, Agustin M. Quality in surgery: current issues for the future. World J Surg 2005 Oct;29(10):1204-1209.

(19) The World Health Organisation. Guidelines for trauma quality improvement programs. 2014; Available at: http://whqlibdoc.who.int/publications/2009/9789241597746_eng.pdf. Accessed June 24, 2015.

(20) Mock C. WHO releases Guidelines for trauma quality improvement programs. Inj Prev 2009 Oct;15(5):359.

(21) Juillard CJ, Mock C, Goosen J, Joshipura M, Civil I. Establishing the evidence base for trauma quality improvement: a collaborative WHO-IATSIC review. World J Surg 2009 May;33(5):1075-1086.

(22) Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729.

(23) Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? J Am Coll Surg 2004 Apr;198(4):626-632.

(24) Porter GA, Soskolne CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. Ann Surg 1998 Feb;227(2):157-167.

(25) Hannan EL, Popp AJ, Feustel P, Halm E, Bernardini G, Waldman J, *et al*. Association of surgical specialty and processes of care with patient outcomes for carotid endarterectomy. Stroke 2001 Dec 1;32(12):2890-2897.

(26) Hannan EL, Racz M, Holmes DR, King SB,3rd, Walford G, Ambrose JA, *et al*. Impact of completeness of percutaneous coronary intervention revascularization on long-term outcomes in the stent era. Circulation 2006 May 23;113(20):2406-2412.

(27) O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT, Nugent WC, *et al*. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. JAMA 1996 Mar 20;275(11):841-846.

(28) 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. Ann Surg 2002 Sep;236(3):344-53; discussion 353-4.

(29) Arora S, Sevdalis N, Ahmed M, Wong H, Moorthy K, Vincent C. Safety skills training for surgeons: A half-day intervention improves knowledge, attitudes and awareness of patient safety. Surgery 2012 Jul;152(1):26-31.

(30) Ayres-De-Campos D, Cruz J, Medeiros-Borges C, Costa-Santos C, Vicente L. Lowered national cesarean section rates after a concerted action. Acta Obstet Gynecol Scand 2015 Apr;94(4):391-398.

(31) Balachandran R, Kappanayil M, Sen AC, Sudhakar A, Nair SG, Sunil GS, *et al*. Impact of the International Quality Improvement Collaborative on outcomes after congenital heart surgery: a single center experience in a developing economy. Ann Card Anaesth 2015 Jan-Mar;18(1):52-57.

(32) Berenguer CM, Ochsner MG,Jr, Lord SA, Senkowski CK. Improving surgical site infections: using National Surgical Quality Improvement Program data to institute Surgical Care Improvement Project protocols in improving surgical outcomes. J Am Coll Surg 2010 May;210(5):737-41, 741-3.

(33) Berrien K, Devente J, French A, Cochran KM, McCaffrey M, Horton BJ, *et al*. The perinatal quality collaborative of North Carolina's 39 weeks project: a quality improvement program to decrease elective deliveries before 39 weeks of gestation. N C Med J 2014 May-Jun;75(3):169-176.

(34) Bliss LA, Ross-Richardson CB, Sanzari LJ, Shapiro DS, Lukianoff AE, Bernstein BA, *et al*. Thirty-day outcomes support implementation of a surgical safety checklist. J Am Coll Surg 2012 Dec;215(6):766-776.

(35) Camp SL, Stamou SC, Stiegel RM, Reames MK, Skipper ER, Madjarov J, *et al*. Quality improvement program increases early tracheal extubation rate and decreases pulmonary complications and resource utilisation after cardiac surgery. J Card Surg 2009 Jul-Aug;24(4):414-423.

(36) Campbell DA,Jr, Englesbe MJ, Kubus JJ, Phillips LR, Shanley CJ, Velanovich V, *et al*. Accelerating the pace of surgical quality improvement: the power of hospital collaboration. Arch Surg 2010 Oct;145(10):985-991.

(37) Cassidy MR, Rosenkranz P, McCabe K, Rosen JE, McAneny CI. Cough: reducing post-operative pulmonary complications with a multidisciplinary patient care program. JAMA Surg 2013 Aug;148(8):740-745.

(38) Ceppa EP, Pitt HA, House MG, Kilbane EM, Nakeeb A, Schmidt CM, *et al*. Reducing surgical site infections in hepatopancreatobiliary surgery. HPB (Oxford) 2013 May;15(5):384-391.

(39) Christelis N, Wallace S, Sage CE, Babitu U, Liew S, Dugal J, *et al*. An enhanced recovery after surgery program for hip and knee arthroplasty. Med J Aust 2015 Apr 20;202(7):363-368.

(40) Cima R, Dankbar E, Lovely J, Pendlimari R, Aronhalt K, Nehring S, *et al*. Colorectal surgery surgical site infection reduction program: a national surgical quality improvement program-driven multidisciplinary single-institution experience. J Am Coll Surg 2013 Jan;216(1):23-33.

(41) Coe SG, Crook JE, Diehl NN, Wallace MB. An endoscopic quality improvement program improves detection of colorectal adenomas. Am J Gastroenterol 2013 Feb;108(2):219-26; quiz 227.

(42) Cohen ME, Liu Y, Ko CY, Hall BL. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. Ann Surg 2015 Feb 26.

(43) Collinge CA, McWilliam-Ross K, Kelly KC, Dombroski D. Substantial improvement in prophylactic antibiotic administration for open fracture patients: results of a performance improvement program. J Orthop Trauma 2014 Nov;28(11):620-625.

(44) Compoginis JM, Katz SG. American College of Surgeons National Surgical Quality Improvement Program as a quality improvement tool: a single institution's experience with vascular surgical site infections. Am Surg 2013 Mar;79(3):274-278.

(45) Crawford RS, Cambria RP, Abularrage CJ, Conrad MF, Lancaster RT, Watkins MT, *et al*. Pre-operative functional status predicts perioperative outcomes after infrainguinal bypass surgery. J Vasc Surg 2010 Feb;51(2):351-8; discussion 358-9.

(46) Cronin J, Livhits M, Mercado C, Chen F, Foster N, Chandler C, *et al*. Quality improvement pilot program for vulnerable elderly surgical patients. Am Surg 2011 Oct;77(10):1305-1308.

(47) da Silva PS, de Aguiar VE, Neto HM, de Carvalho WB. Unplanned extubation in a paediatric intensive care unit: impact of a quality improvement program. Anaesthesia 2008 Nov;63(11):1209-1216.

(48) Dinh MM, Bein KJ, Gabbe BJ, Byrne CM, Petchell J, Lo S, *et al*. A trauma quality improvement program associated with improved patient outcomes: 21 years of experience at an Australian Major Trauma Center. Injury 2014 May;45(5):830-834.

(49) Dua A, Desai SS, Seabrook GR, Brown KR, Lewis BD, Rossi PJ, *et al*. The effect of Surgical Care Improvement Project measures on national trends on surgical site infections in open vascular procedures. J Vasc Surg 2014 Dec;60(6):1635-1639.

(50) Dubose J, Teixeira PG, Inaba K, Lam L, Talving P, Putty B, *et al*. Measurable outcomes of quality improvement using a daily quality rounds checklist: one-year analysis in a trauma intensive care unit with sustained ventilator-associated pneumonia reduction. J Trauma 2010 Oct;69(4):855-860.

(51) Dumon KR, Edelson PK, Raper SE, Foster-Kilgarriff K, Williams NN. Implementation of designated bariatric surgery program leads to improved clinical outcomes. Surg Obes Relat Dis 2011 May-Jun;7(3):271-276.

(52) Durham R, Shapiro D, Flint L. In-house trauma attendings: is there a difference? Am J Surg 2005 Dec;190(6):960-966.

(53) Esposito TJ, Sanddal TL, Reynolds SA, Sanddal ND. Effect of a voluntary trauma system on preventable death and inappropriate care in a rural state. J Trauma 2003 Apr;54(4):663-9; discussion 669-70.

(54) Etzioni DA, Wasif N, Dueck AC, Cima RR, Hohmann SF, Naessens JM, *et al*. Association of hospital participation in a surgical outcomes monitoring program with inpatient complications and mortality. JAMA 2015 Feb 3;313(5):505-511.

(55) Evans KH, Daines W, Tsui J, Strehlow M, Maggio P, Shieh L. Septris: a novel, mobile, online, simulation game that improves sepsis recognition and management. Acad Med 2015 Feb;90(2):180-184.

(56) Ferrario CM, Joyner J, Colby C, Exuzides A, Moore M, Simmons D, *et al*. The COSEHC Global Vascular Risk Management quality improvement program: first follow-up report. Vasc Health Risk Manag 2013;9:391-400.

(57) Findlay J, Keogh M. Simple multidisciplinary education of junior doctors and nurses improves prescription of venous thromboembolism prophylaxis. J Perioper Pract 2011 Jan;21(1):28-32.

(58) Fox CJ, Gillespie DL, Cox ED, Mehta SG, Kragh JF,Jr, Salinas J, *et al*. The effectiveness of a damage control resuscitation strategy for vascular injury in a combat support hospital: results of a case control study. J Trauma 2008 Feb;64(2 Suppl):S99-106; discussion S106-7.

(59) Gabbe BJ, Biostat GD, Lecky FE, Bouamra O, Woodford M, Jenks T, *et al*. The effect of an organised trauma system on mortality in major trauma involving serious head injury: a comparison of the United Kingdom and Victoria, Australia. Ann Surg 2011 Jan;253(1):138-143.

(60) Ganz DA, Kim SB, Zingmond DS, Ramirez KD, Roth CP, Jennings LA, *et al*. Effect of a falls quality improvement program on serious fall-related injuries. J Am Geriatr Soc 2015 Jan;63(1):63-70.

(61) Garrouste-Orgeas M, Soufir L, Tabah A, Schwebel C, Vesin A, Adrie C, *et al*. A multifaceted program for improving quality of care in intensive care units: IATROREF study. Crit Care Med 2012 Feb;40(2):468-476.

(62) Goodney PP, Eldrup-Jorgensen J, Nolan BW, Bertges DJ, Likosky DS, Cronenwett JL, *et al*. A regional quality improvement effort to increase beta blocker administration before vascular surgery. J Vasc Surg 2011 May;53(5):1316-1328.e1; discussion 1327-8.

(63) Graham LA, Singletary BA, Richman JS, Maddox TM, Itani KM, Hawn MT. Improved adverse post-operative outcomes with revised American College of Cardiology/American Heart Association guidelines for patients with cardiac stents. JAMA Surg 2014 Nov;149(11):1113-1120.

(64) Greco M, Capretti G, Beretta L, Gemma M, Pecorelli N, Braga M. Enhanced recovery program in colorectal surgery: a meta-analysis of randomized controlled trials. World J Surg 2014 Jun;38(6):1531-1541.

(65) Gruen RL, Jurkovich GJ, McIntyre LK, Foy HM, Maier RV. Patterns of errors contributing to trauma mortality: lessons learned from 2,594 deaths. Ann Surg 2006 Sep;244(3):371-380.

(66) Guillamondegui OD, Gunter OL, Hines L, Martin BJ, Gibson W, Clarke PC, *et al*. Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to improve surgical outcomes. J Am Coll Surg 2012 Apr;214(4):709-14; discussion 714-6.

(67) Gupta PK, Ramanan B, Lynch TG, Sundaram A, MacTaggart JN, Gupta H, *et al*. Development and validation of a risk calculator for prediction of mortality after infrainguinal bypass surgery. J Vasc Surg 2012 Aug;56(2):372-379.

(68) Gupta RT, Sexton JB, Milne J, Frush DP. Practice and quality improvement: successful implementation of TeamSTEPPS tools into an academic interventional ultrasound practice. AJR Am J Roentgenol 2015 Jan;204(1):105-110.

(69) Halpern MT, Spain P, Holden DJ, Stewart A, McNamara EJ, Gay G, *et al*. Improving quality of cancer care at community hospitals: impact of the National Cancer Institute Community Cancer Centers Program pilot. J Oncol Pract 2013 Nov;9(6):e298-304.

(70) Han SJ, Saigal R, Rolston JD, Cheng JS, Lau CY, Mistry RI, *et al*. Targeted reduction in neurosurgical laboratory utilization: resident-led effort at a single academic institution. J Neurosurg 2014 Jan;120(1):173-177.

(71) Hannan EL, Sarrazin MS, Doran DR, Rosenthal GE. Provider profiling and quality improvement efforts in coronary artery bypass graft surgery: the effect on short-term mortality among Medicare beneficiaries. Med Care 2003 Oct;41(10):1164-1172.

(72) Harter P, Muallem ZM, Buhrmann C, Lorenz D, Kaub C, Hils R, *et al*. Impact of a structured quality management program on surgical outcome in primary advanced ovarian cancer. Gynecol Oncol 2011 Jun 1;121(3):615-619.

(73) Hashmi ZG, Haider AH, Zafar SN, Kisat M, Moosa A, Siddiqui F, *et al*. Hospital-based trauma quality improvement initiatives: first step toward improving trauma outcomes in the developing world. J Trauma Acute Care Surg 2013 Jul;75(1):60-8; discussion 68.

(74) Hawn MT, Vick CC, Richman J, Holman W, Deierhoi RJ, Graham LA, *et al*. Surgical site infection prevention: time to move beyond the surgical care improvement program. Ann Surg 2011 Sep;254(3):494-9; discussion 499-501.

(75) 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 Jan 29;360(5):491-499.

(76) Hemmila MR, Cain-Nielsen AH, Wahl WL, Vander Kolk WE, Jakubus JL, Mikhail JN, *et al*. Regional collaborative quality improvement for trauma reduces complications and costs. J Trauma Acute Care Surg 2015 Jan;78(1):78-85; discussion 85-7.

(77) Henry JA, Orgoi S, Govind S, Price RR, Lundeg G, Kehrer B. Strengthening surgical services at the soum (first-referral) hospital: the WHO emergency and essential surgical care (EESC) program in Mongolia. World J Surg 2012 Oct;36(10):2359-2370.

(78) Ilan R, Doan J, Cload B, Squires M, Day A. Removing nonessential central venous catheters: evaluation of a quality improvement intervention. Can J Anaesth 2012 Dec;59(12):1102-1110.

(79) Inaba K, Hauch A, Branco BC, Cohn S, Teixeira PG, Recinos G, *et al*. The impact of in-house attending surgeon supervision on the rates of preventable and potentially preventable complications and death at the start of the new academic year. Am Surg 2013 Nov;79(11):1134-1139.

(80) Jeffries HE, Mason W, Brewer M, Oakes KL, Munoz EI, Gornick W, *et al*. Prevention of central venous catheter-associated bloodstream infections in pediatric intensive care units: a performance improvement collaborative. Infect Control Hosp Epidemiol 2009 Jul;30(7):645-651.

(81) Kagan RJ, Neely AN, Rieman MT, Hardy A, Warner P, Bailey JK, *et al*. A performance improvement initiative to determine the impact of increasing the time interval between changing centrally placed intravascular catheters. J Burn Care Res 2014 Mar-Apr;35(2):143-147.

(82) Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB, *et al*. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg 1998 Oct;228(4):491-507.

(83) Kim DS, Madden A, Alexander C, Zois M, Ogunyemi D, Azziz R. Implementing a teaching service for voluntary faculty in obstetrics and gynecology: long-term results. J Reprod Med 2013 Sep-Oct;58(9-10):371-376.

(84) Klein AL, Brown CV, Aydelotte J, Ali S, Clark A, Coopwood B. Implementation of a surgical intensive care unit service is associated with improved outcomes for trauma patients. J Trauma Acute Care Surg 2014 Dec;77(6):964-7; discussion 967-8.

(85) Kogan A, Preisman S, Berkenstadt H, Segal E, Kassif Y, Sternik L, *et al*. Evaluation of the impact of a quality improvement program and intensivist-directed ICU team on mortality after cardiac surgery. J Cardiothorac Vasc Anesth 2013 Dec;27(6):1194-1200.

(86) Konstantinidis A, Fogel S, Jones J, Gilliam B, Kundzins J, Baker C. Use of the national surgical quality improvement program to reduce surgical mortality: implementation of intensive pre-operative screening and intervention. Am Surg 2014 Sep;80(9):896-900.

(87) Kresowik TF, Hemann RA, Grund SL, Hendel ME, Brenton M, Wiblin RT, *et al*. Improving the outcomes of carotid endarterectomy: results of a statewide quality improvement project. J Vasc Surg 2000 May;31(5):918-926.

(88) Lancaster RT, Conrad MF, Patel VI, Cambria RP, LaMuraglia GM. Predictors of early graft failure after infrainguinal bypass surgery: a risk-adjusted analysis from the NSQIP. Eur J Vasc Endovasc Surg 2012 May;43(5):549-555.

(89) Larochelle M, Hyman N, Gruppi L, Osler T. Diminishing surgical site infections after colorectal surgery with surgical care improvement project: is it time to move on? Dis Colon Rectum 2011 Apr;54(4):394-400.

(90) Lehmkuhl D, Meissner W, Neugebauer EA. Evaluation of the "initiative pain-free clinic" for quality improvement in post-operative pain management. A prospective controlled study. Schmerz 2011 Sep;25(5):508-515.

(91) Lin DM, Weeks K, Bauer L, Combes JR, George CT, Goeschel CA, *et al*. Eradicating central line-associated bloodstream infections statewide: the Hawaii experience. Am J Med Qual 2012 Mar-Apr;27(2):124-129.

(92) Lolom I, Deblangy C, Capelle A, Guerinot W, Bouvet E, Barry B, *et al*. Effect of a long-term quality improvement program on the risk of infection related to peripheral venous catheters. Presse Med 2009 Jan;38(1):34-42.

(93) Lucas DJ, Pawlik TM. Quality improvement in gastrointestinal surgical oncology with American College of Surgeons National Surgical Quality Improvement Program. Surgery 2014 Apr;155(4):593-601.

(94) Mainthia R, Lockney T, Zotov A, France DJ, Bennett M, St Jacques PJ, *et al*. Novel use of electronic whiteboard in the operating room increases surgical team compliance with pre-incision safety practices. Surgery 2012 May;151(5):660-666.

(95) Marsteller JA, Sexton JB, Hsu YJ, Hsiao CJ, Holzmueller CG, Pronovost PJ, *et al*. A multicenter, phased, cluster-randomised controlled trial to reduce central line-associated bloodstream infections in intensive care units*. Crit Care Med 2012 Nov;40(11):2933-2939.

(96) McDermott FT, Cordner SM, Cooper DJ, Winship VC, Consultative Committee on Road Traffic Fatalities in Victoria. Management deficiencies and death preventability of road traffic fatalities before and after a new trauma care system in Victoria, Australia. J Trauma 2007 Aug;63(2):331-338.

(97) McNelis J, Castaldi M. "The National Surgery Quality Improvement Project" (NSQIP): a new tool to increase patient safety and cost efficiency in a surgical intensive care unit. Patient Saf Surg 2014 Apr 24;8:19-9493-8-19. eCollection 2014.

(98) McPeake J, Cantwell S, Booth MG, Daniel M. Central line insertion bundle: experiences and challenges in an adult ICU. Nurs Crit Care 2012 May-Jun;17(3):123-129.

(99) Mirski MA, Pandian V, Bhatti N, Haut E, Feller-Kopman D, Morad A, *et al.* Safety, efficiency, and cost-effectiveness of a multidisciplinary percutaneous tracheostomy program. Crit Care Med 2012 Jun;40(6):1827-1834.

(100) Munday GS, Deveaux P, Roberts H, Fry DE, Polk HC. Impact of implementation of the Surgical Care Improvement Project and future strategies for improving quality in surgery. Am J Surg 2014 Nov;208(5):835-840.

(101) Murphy LM, Murphy SO, Hastings MA, Olberding A. Are Interprofessional Roundtable Debriefings Useful in Decreasing ED Fall Rates? Findings From a Quality-Improvement Project. J Emerg Nurs 2015 Apr 11.

(102) Navarro S, Koo M, Orrego C, Munoz-Vives JM, Rivero M, Montmany S, *et al.* Study on the improvement of trauma patient care: TRAUMACAT project. Med Clin (Barc) 2014 Jul;143 Suppl 1:25-31.

(103) Neily J, Mills PD, Young-Xu Y, Carney BT, West P, Berger DH, *et al.* Association between implementation of a medical team training program and surgical mortality. JAMA 2010 Oct 20;304(15):1693-1700.

(104) Neumayer L, Mastin M, Vanderhoof L, Hinson D. Using the Veterans Administration National Surgical Quality Improvement Program to improve patient outcomes. J Surg Res 2000 Jan;88(1):58-61.

(105) Nguyen VD, Lawson L, Ledeen M, Treat L, Buss J, Barclay C, *et al.* Successful multidisciplinary interventions for arterio-venous fistula creation by the Pacific Northwest Renal Network 16 vascular access quality improvement program. J Vasc Access 2007 Jan-Mar;8(1):3-11.

(106) O'Reilly D, Mahendran K, West A, Shirley P, Walsh M, Tai N. Opportunities for improvement in the management of patients who die from haemorrhage after trauma. Br J Surg 2013 May;100(6):749-755.

(107) Osborne NH, Nicholas LH, Ryan AM, Thumma JR, Dimick JB. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for Medicare beneficiaries. JAMA 2015 Feb 3;313(5):496-504.

(108) Oshiro BT, Kowalewski L, Sappenfield W, Alter CC, Bettegowda VR, Russell R, *et al.* A multistate quality improvement program to decrease elective deliveries before 39 weeks of gestation. Obstet Gynecol 2013 May;121(5):1025-1031.

(109) Owen RM, Perez SD, Bornstein WA, Sweeney JF. Impact of surgical care improvement project inf-9 on post-operative urinary tract infections: do exemptions interfere with quality patient care? Arch Surg 2012 Oct;147(10):946-953.

(110) Owens P, McHugh S, Clarke-Moloney M, Healy D, Fitzpatrick F, McCormick P, *et al*. Improving surgical site infection prevention practices through a multifaceted educational intervention. Ir Med J 2015 Mar;108(3):78-81.

(111) Pemberton J, Rambaran M, Cameron BH. Evaluating the long-term impact of the Trauma Team Training course in Guyana: an explanatory mixed-methods approach. Am J Surg 2013 Feb;205(2):119-124.

(112) Perkins ZB, Maytham GD, Koers L, Bates P, Brohi K, Tai NR. Impact on outcome of a targeted performance improvement program in haemodynamically unstable patients with a pelvic fracture. Bone Joint J 2014 Aug;96-B(8):1090-1097.

(113) Pettker CM, Thung SF, Lipkind HS, Illuzzi JL, Buhimschi CS, Raab CA, *et al*. A comprehensive obstetric patient safety program reduces liability claims and payments. Am J Obstet Gynecol 2014 Oct;211(4):319-325.

(114) Pfeifer K, Slawski B, Manley AM, Nelson V, Haines M. Improving pre-operative medication compliance with standardised instructions. Minerva Anestesiol 2015 Apr 24.

(115) Pinto Thirukumaran C, Dolan JG, Reagan Webster P, Panzer RJ, Friedman B. The impact of electronic health record implementation and use on performance of the Surgical Care Improvement Project measures. Health Serv Res 2015 Feb;50(1):273-289.

(116) Pronovost P. Interventions to decrease catheter-related bloodstream infections in the ICU: the Keystone Intensive Care Unit Project. Am J Infect Control 2008 Dec;36(10):S171.e1-5.

(117) Reames BN, Krell RW, Campbell DA,Jr, Dimick JB. A checklist-based intervention to improve surgical outcomes in Michigan: evaluation of the Keystone Surgery program. JAMA Surg 2015 Mar 1;150(3):208-215.

(118) Reinke CE, Kelz RR, Baillie CA, Norris A, Schmidt S, Wingate N, *et al*. Timeliness and quality of surgical discharge summaries after the implementation of an electronic format. Am J Surg 2014 Jan;207(1):7-16.

(119) Renzi C, Sorge C, Fusco D, Agabiti N, Davoli M, Perucci CA. Reporting of quality indicators and improvement in hospital performance: the P.Re.Val.E. Regional Outcome Evaluation Program. Health Serv Res 2012 Oct;47(5):1880-1901.

(120) Roche-Nagle G, Bachynski K, Nathens AB, Angoulvant D, Rubin BB. Regionalisation of services improves access to emergency vascular surgical care. Vascular 2013 Apr;21(2):69-74.

(121) Roquilly A, Cinotti R, Jaber S, Vourc'h M, Pengam F, Mahe PJ, *et al*. Implementation of an evidence-based extubation readiness bundle in 499 brain-injured patients. a before-after evaluation of a quality improvement project. Am J Respir Crit Care Med 2013 Oct 15;188(8):958-966.

(122) Rowell KS, Turrentine FE, Hutter MM, Khuri SF, Henderson WG. Use of national surgical quality improvement program data as a catalyst for quality improvement. J Am Coll Surg 2007 Jun;204(6):1293-1300.

(123) Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120.

(124) Salazar A, Tyroch AH, Smead DG. Electronic trauma patient outcomes assessment tool: performance improvement in the trauma intensive care unit. J Trauma Nurs 2011 Oct-Dec;18(4):197-201.

(125) Salim A, Teixeira PG, Chan L, Oncel D, Inaba K, Brown C, *et al*. Impact of the 80-hour workweek on patient care at a level I trauma center. Arch Surg 2007 Aug;142(8):708-12; discussion 712-4.

(126) Sarkar B, Brunsvold ME, Cherry-Bukoweic JR, Hemmila MR, Park PK, Raghavendran K, *et al*. American College of Surgeons' Committee on Trauma Performance Improvement and Patient Safety program: maximal impact in a mature trauma center. J Trauma 2011 Nov;71(5):1447-53; discussion 1453-4.

(127) Scarborough JE, Pappas TN, Cox MW, Bennett KM, Shortell CK. Surgical trainee participation during infrainguinal bypass grafting procedures is associated with increased early post-operative graft failure. J Vasc Surg 2012 Mar;55(3):715-720.

(128) Schwann NM, Bretz KA, Eid S, Burger T, Fry D, Ackler F, *et al*. Point-of-care electronic prompts: an effective means of increasing compliance, demonstrating quality, and improving outcome. Anesth Analg 2011 Oct;113(4):869-876.

(129) Share DA, Campbell DA, Birkmeyer N, Prager RL, Gurm HS, Moscucci M, *et al*. How a regional collaborative of hospitals and physicians in Michigan cut costs and improved the quality of care. Health Aff (Millwood) 2011 Apr;30(4):636-645.

(130) Sheikh K, Bullock C, Preston SD. Evaluation of quality improvement interventions for reducing adverse outcomes of carotid endarterectomy. Med Care 2004 Jul;42(7):690-699.

(131) Shore BJ, Hutchinson S, Harris M, Bae DS, Kalish LA, Maxwell W,3rd, *et al*. Epidemiology and prevention of cast saw injuries: results of a quality improvement program at a single institution. J Bone Joint Surg Am 2014 Feb 19;96(4):e31.

(132) Singh H, Kadiyala H, Bhagwath G, Shethia A, El-Serag H, Walder A, *et al*. Using a multifaceted approach to improve the follow-up of positive fecal occult blood test results. Am J Gastroenterol 2009 Apr;104(4):942-952.

(133) Spijkers AT, Meylaerts SA, Leenen LP. Mortality decreases by implementing a level I trauma center in a Dutch hospital. J Trauma 2010 Nov;69(5):1138-1142.

(134) Stahl K, Palileo A, Schulman CI, Wilson K, Augenstein J, Kiffin C, *et al*. Enhancing patient safety in the trauma/surgical intensive care unit. J Trauma 2009 Sep;67(3):430-3; discussion 433-5.

(135) Stamou SC, Camp SL, Reames MK, Skipper E, Stiegel RM, Nussbaum M, *et al*. Continuous quality improvement program and major morbidity after cardiac surgery. Am J Cardiol 2008 Sep 15;102(6):772-777.

(136) Stanek JJ, Renslow MA, Kalliainen LK. The effect of an educational program on opioid prescription patterns in hand surgery: a quality improvement program. J Hand Surg Am 2015 Feb;40(2):341-346.

(137) Stein R, Neufeld D, Shwartz I, Erez I, Haas I, Magen A, *et al*. Assessment of surgical discharge summaries and evaluation of a new quality improvement model. Isr Med Assoc J 2014 Nov;16(11):714-717.

(138) Stelfox HT, Straus SE, Nathens A, Bobranska-Artiuch B. Evidence for quality indicators to evaluate adult trauma care: a systematic review. Crit Care Med 2011 Apr;39(4):846-859.

(139) Tadros RO, Faries PL, Malik R, Vouyouka AG, Ting W, Dunn A, *et al*. The effect of a hospitalist comanagement service on vascular surgery inpatients. J Vasc Surg 2015 Jun;61(6):1550-1555.

(140) Teixeira PG, Inaba K, Dubose J, Melo N, Bass M, Belzberg H, *et al*. Measurable outcomes of quality improvement using a daily quality rounds checklist: two-year prospective analysis of sustainability in a surgical intensive care unit. J Trauma Acute Care Surg 2013 Oct;75(4):717-721.

(141) Tillman M, Wehbe-Janek H, Hodges B, Smythe WR, Papaconstantinou HT. Surgical care improvement project and surgical site infections: can integration in the surgical safety checklist improve quality performance and clinical outcomes? J Surg Res 2013 Sep;184(1):150-156.

(142) Turek JW, Andersen ND, Lawson DS, Bonadonna D, Turley RS, Peters MA, *et al*. Outcomes before and after implementation of a pediatric rapid-response extracorporeal membrane oxygenation program. Ann Thorac Surg 2013 Jun;95(6):2140-6; discussion 2146-7.

(143) van der Veer SN, de Vos ML, van der Voort PH, Peek N, Abu-Hanna A, Westert GP, *et al*. Effect of a multifaceted performance feedback strategy on length of stay compared with benchmark reports alone: a cluster randomised trial in intensive care*. Crit Care Med 2013 Aug;41(8):1893-1904.

(144) Vernacchio L, Trudell EK, Hresko MT, Karlin LI, Risko W. A quality improvement program to reduce unnecessary referrals for adolescent scoliosis. Pediatrics 2013 Mar;131(3):e912-20.

(145) Weiner BK, Venarske J, Yu M, Mathis K. Towards the reduction of medication errors in orthopedics and spinal surgery: outcomes using a pharmacist-led approach. Spine (Phila Pa 1976) 2008 Jan 1;33(1):104-107.

(146) Wick EC, Hobson DB, Bennett JL, Demski R, Maragakis L, Gearhart SL, *et al*. Implementation of a surgical comprehensive unit-based safety program to reduce surgical site infections. J Am Coll Surg 2012 Aug;215(2):193-200.

(147) Witter FR, Lawson P, Ferrell J. Decreasing cesarean section surgical site infection: an ongoing comprehensive quality improvement program. Am J Infect Control 2014 Apr;42(4):429-431.

(148) Young-Xu Y, Neily J, Mills PD, Carney BT, West P, Berger DH, *et al*. Association between implementation of a medical team training program and surgical morbidity. Arch Surg 2011 Dec;146(12):1368-1373.

(149) Fung-Kee-Fung M, Watters J, Crossley C, Goubanova E, Abdulla A, Stern H, *et al*. Regional collaborations as a tool for quality improvements in surgery: a systematic review of the literature. Ann Surg 2009 Apr;249(4):565-572.

(150) Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. Ann Surg 2009 Sep;250(3):363-376.

(151) Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA,Jr, *et al*. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. Ann Surg 2008 Aug;248(2):329-336.

(152) Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA,Jr, *et al*. The patient safety in surgery study: background, study design, and patient populations. J Am Coll Surg 2007 Jun;204(6):1089-1102.

(153) Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg 2009 Nov;198(5 Suppl):S19-27.

(154) Stamou SC, Camp SL, Stiegel RM, Reames MK, Skipper E, Watts LT, *et al*. Quality improvement program decreases mortality after cardiac surgery. J Thorac Cardiovasc Surg 2008 Aug;136(2):494-499.e8.

(155) Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB, *et al*. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg 1998 Oct;228(4):491-507.

(156) Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA,Jr, *et al*. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. Ann Surg 2008 Aug;248(2):329-336.

(157) Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg 2009 Nov;198(5 Suppl):S19-27.

(158) Birkmeyer JD, Shahian DM, Dimick JB, Finlayson SR, Flum DR, Ko CY, *et al*. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. J Am Coll Surg 2008 Nov;207(5):777-782.

(159) Juillard CJ, Mock C, Goosen J, Joshipura M, Civil I. Establishing the evidence base for trauma quality improvement: a collaborative WHO-IATSIC review. World J Surg 2009 May;33(5):1075-1086.

(160) Newgard CD, Fildes JJ, Wu L, Hemmila MR, Burd RS, Neal M, *et al*. Methodology and analytic rationale for the American College of Surgeons Trauma Quality Improvement Program. J Am Coll Surg 2013 Jan;216(1):147-157.

(161) Champion HR, Copes WS, Sacco WJ, Lawnick MM, Keast SL, Bain LW,Jr, *et al*. The Major Trauma Outcome Study: establishing national norms for trauma care. J Trauma 1990 Nov;30(11):1356-1365.

(162) Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. J Trauma 1987 Apr;27(4):370-378.

(163) Hemmila MR, Cain-Nielsen AH, Wahl WL, Vander Kolk WE, Jakubus JL, Mikhail JN, *et al*. Regional collaborative quality improvement for trauma reduces complications and costs. J Trauma Acute Care Surg 2015 Jan;78(1):78-85; discussion 85-7.

(164) Hemmila MR, Nathens AB, Shafi S, Calland JF, Clark DE, Cryer HG, *et al*. The Trauma Quality Improvement Program: pilot study and initial demonstration of feasibility. J Trauma 2010 Feb;68(2):253-262.

(165) Hemmila MR, Cain-Nielsen AH, Wahl WL, Vander Kolk WE, Jakubus JL, Mikhail JN, *et al*. Regional collaborative quality improvement for trauma reduces complications and costs. J Trauma Acute Care Surg 2015 Jan;78(1):78-85; discussion 85-7.

(166) Fournier P, Dumont A, Tourigny C, Dunkley G, Drame S. Improved access to comprehensive emergency obstetric care and its effect on institutional maternal mortality in rural Mali. Bull World Health Organ 2009 Jan;87(1):30-38.

(167) Balachandran R, Kappanayil M, Sen AC, Sudhakar A, Nair SG, Sunil GS, *et al*. Impact of the International Quality Improvement Collaborative on outcomes after congenital heart surgery: a single center experience in a developing economy. Ann Card Anaesth 2015 Jan-Mar;18(1):52-57.

(168) 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 Jan 29;360(5):491-499.

(169) Haider AH. Playing darts without a dart board: why we need to create an International Trauma Data Bank (ITDB). Injury 2014 Jan;45(1):1-2.

(170) Haider AH, Hashmi ZG, Gupta S, Zafar SN, David JS, Efron DT, *et al*. Benchmarking of trauma care worldwide: the potential value of an International Trauma Data Bank (ITDB). World J Surg 2014 Aug;38(8):1882-1891.

(171) Anderson JE, Rose J, Noorbakhsh A, Talamini MA, Finlayson SR, Bickler SW, *et al*. An efficient risk adjustment model to predict inpatient adverse events after surgery. World J Surg 2014 Aug;38(8):1954-1960.

(172) Zargaran E, Schuurman N, Nicol AJ, Matzopoulos R, Cinnamon J, Taulu T, *et al*. The electronic Trauma Health Record: design and usability of a novel tablet-based tool for trauma care and injury surveillance in low resource settings. J Am Coll Surg 2014 Jan;218(1):41-50.

(173) Medhanyie AA, Moser A, Spigt M, Yebyo H, Little A, Dinant G, *et al*. Mobile health data collection at primary health care in Ethiopia: a feasible challenge. J Clin Epidemiol 2015 Jan;68(1):80-86.

(174) Project REDCap. 2015; Available at: http://www.project-redcap.org/. Accessed June, 24, 2015.

(175) Cook F. Analytical epidemiology lecture series. 2014.

(176) Chawda MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: which scoring system? Injury 2004 Apr;35(4):347-358.

(177) Ringdal KG, Skaga NO, Hestnes M, Steen PA, Roislien J, Rehn M, *et al*. Abbreviated Injury Scale: not a reliable basis for summation of injury severity in trauma facilities? Injury 2013 May;44(5):691-699.

(178) 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. Ann Surg 2009 Aug;250(2):187-196.

(179) Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. J Trauma 1997 Dec;43(6):922-5; discussion 925-6.

(180) Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. J Trauma 1989 May;29(5):623-629.

(181) Kobusingye OC, Lett RR. Hospital-based trauma registries in Uganda. J Trauma 2000 Mar;48(3):498-502.

(182) Haac B, Varela C, Geyer A, Cairns B, Charles A. The utility of the Kampala trauma score as a triage tool in a sub-Saharan African trauma cohort. World J Surg 2015 Feb;39(2):356-362.

(183) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310.

(184) Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363.

(185) Ringdal KG, Skaga NO, Hestnes M, Steen PA, Roislien J, Rehn M, *et al*. Abbreviated Injury Scale: not a reliable basis for summation of injury severity in trauma facilities? Injury 2013 May;44(5):691-699.

(186) Farmer DL. NSQIP lite: a potential tool for global comparative effectiveness evaluations. Arch Surg 2012 Sep;147(9):803-804.

(187) Cohen ME, Liu Y, Ko CY, Hall BL. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. Ann Surg 2015 Feb 26.

(188) Farmer PE, Kim JY. Surgery and global health: a view from beyond the OR. World J Surg 2008 Apr;32(4):533-536.

(189) Meara JG, Leather AJ, Hagander L, Alkire BC, Alonso N, Ameh EA, *et al*. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet 2015 Aug 8;386(9993):569-624.

(190) Haider AH, Saleem T, Leow JJ, Villegas CV, Kisat M, Schneider EB, *et al*. Influence of the National Trauma Data Bank on the study of trauma outcomes: is it time to set research best practices to further enhance its impact? J Am Coll Surg 2012 May;214(5):756-768.

(191) Ariyaratnam R, Palmqvist CL, Hider P, Laing GL, Stupart D, Wilson L, *et al*. Toward a standard approach to measurement and reporting of perioperative mortality rate as a global indicator for surgery. Surgery 2015 Jul;158(1):17-26.

(192) Saklad M. Grading of patients for surgical procedures. Anethesiology 1941;2:281-284.

(193) Ranta S, Hynynen M, Tammisto T. A survey of the ASA physical status classification: significant variation in allocation among Finnish anaesthesiologists. Acta Anaesthesiol Scand 1997 May;41(5):629-632.

(194) Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. Anaesthesia 1995 Mar;50(3):195-199.

(195) Parikh P, Shiloach M, Cohen ME, Bilimoria KY, Ko CY, Hall BL, *et al*. Pancreatectomy risk calculator: an ACS-NSQIP resource. HPB (Oxford) 2010 Sep;12(7):488-497.

(196) Andersen J, Lassiter R, Bickler S, Talamini M, Chang D. Brief tool to measure risk-adjusted surgical outcomes in resource-limited hospitals. 2012;147(9):798-803.

(197) Muller E, Barday Z, Kahn D. HIV-positive-to-HIV-positive kidney transplantation. N Engl J Med 2015 May 21;372(21):2070-2071.

(198) Muller E, Barday Z, Mendelson M, Kahn D. HIV-positive-to-HIV-positive kidney transplantation--results at 3 to 5 years. N Engl J Med 2015 Feb 12;372(7):613-620.

(199) Muller E. Transplantation in resource-limited setting: using HIV-positive donors for HIV-positive patients. Clin Nephrol 2015;83(7 Suppl 1):39-41.

(200) Muller E. Research Highlights. Transplantation 2015 Nov;99(11):2234.

(201) Muller E. Elmi Muller, MBChB, MMED: Head, Transplantation Service, Groote Schuur Hospital and Professor of Surgery, University of Cape Town, South Africa. Transplantation 2015 Dec;99(12):2440-2441.

(202) Gonzalez AA, Dimick JB, Birkmeyer JD, Ghaferi AA. Understanding the volume-outcome effect in cardiovascular surgery: the role of failure to rescue. JAMA Surg 2014 Feb;149(2):119-123.

(203) Dossett LA, Redhage LA, Sawyer RG, May AK. Revisiting the validity of APACHE II in the trauma ICU: improved risk stratification in critically injured adults. Injury 2009 Sep;40(9):993-998.

(204) 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 Apr;149(4):519-524.

(205) Bickler S, Ozgediz D, Gosselin R, Weiser T, Spiegel D, Hsia R, *et al*. Key concepts for estimating the burden of surgical conditions and the unmet need for surgical care. World J Surg 2010 Mar;34(3):374-380.

(206) 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 Sep 22;380(9847):1059-1065.

(207) Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and operative mortality in the modern era. Ann Surg 2014 Aug;260(2):244-251.

(208) McCulloch P, Nagendran M, Campbell WB, Price A, Jani A, Birkmeyer JD, *et al*. Strategies to reduce variation in the use of surgery. Lancet 2013 Sep 28;382(9898):1130-1139.

(209) Revels SL, Wong SL, Banerjee M, Yin H, Birkmeyer JD. Differences in perioperative care at low- and high-mortality hospitals with cancer surgery. Ann Surg Oncol 2014 Jul;21(7):2129-2135.

(210) Bhangu A. Determining universal processes related to best outcome in emergency abdominal surgery: a multicenter, international, prospective cohort study. BMJ open 2014;4(10):e006239.

(211) Gifford E, Kim DY, Nguyen A, Kaji AH, Nguyen V, Plurad DS, *et al*. The effect of residents as teaching assistants on operative time in laparoscopic cholecystectomy. Am J Surg 2015 Aug 12.

(212) Iannuzzi JC, Chandra A, Rickles AS, Kumar NG, Kelly KN, Gillespie DL, *et al*. Resident involvement is associated with worse outcomes after major lower extremity amputation. J Vasc Surg 2013 Sep;58(3):827-31.e1.

(213) Tsigonis AM, Landercasper J, Al-Hamadani M, Linebarger JH, Vang CA, Johnson JM, *et al.* Are Breast Cancer Outcomes Compromised by General Surgical Resident Participation in the Operation? J Surg Educ 2015 Jul 15.

(214) Meyerson SL, Teitelbaum EN, George BC, Schuller MC, DaRosa DA, Fryer JP. Defining the autonomy gap: when expectations do not meet reality in the operating room. J Surg Educ 2014 Nov-Dec;71(6):e64-72.

(215) Patel M, Bhullar JS, Subhas G, Mittal V. Present Status of Autonomy in Surgical Residency--a Program Director's Perspective. Am Surg 2015 Aug;81(8):786-790.

(216) Sachs TE, Pawlik TM. See one, do one, and teach none: resident experience as a teaching assistant. J Surg Res 2015 May 1;195(1):44-51.

(217) 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 Jan 29;360(5):491-499.

(218) Dalton JE, Kurz A, Turan A, Mascha EJ, Sessler DI, Saager L. Development and validation of a risk quantification index for 30-day post-operative mortality and morbidity in noncardiac surgical patients. Anesthesiology 2011 Jun;114(6):1336-1344.

(219) Ingraham AM, Richards KE, Hall BL, Ko CY. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. Adv Surg 2010;44:251-267.

(220) Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, *et al.* Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013 Nov;217(5):833-42.e1-3.

(221) Bohnen JD, Chang DC, Lillemoe KD. Reconceiving the Morbidity and Mortality Conference in an Era of Big Data: An 'Unexpected' Outcomes Approach. Ann Surg 2015 Nov 25.

(222) Hutter MM, Rowell KS, Devaney LA, Sokal SM, Warshaw AL, Abbott WM, *et al.* Identification of surgical complications and deaths: an assessment of the traditional surgical morbidity and mortality conference compared with the American College of Surgeons-National Surgical Quality Improvement Program. J Am Coll Surg 2006 Nov;203(5):618-624.

(223) Ioannidis JP. Are mortality differences detected by administrative data reliable and actionable? JAMA 2013 Apr 3;309(13):1410-1411.

(224) Management Sciences for Health MSH. Family Planning Management Development. Focusing on customer service. Fam Plan Manag 1996 Spring;5(1):1-18.

(225) Anderson JE, Lassiter R, Bickler SW, Talamini MA, Chang DC. Brief tool to measure risk-adjusted surgical outcomes in resource-limited hospitals. Arch Surg 2012 Sep;147(9):798-803.

(226) Chatterjee AS, Renganathan DN. POSSUM: A Scoring System for Perforative Peritonitis. J Clin Diagn Res 2015 Apr;9(4):PC05-9.

(227) 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 Apr;149(4):519-524.

(228) Weeks SR, Stevens KA, Haider AH, Efron DT, Haut ER, MacKenzie EJ, *et al*. A modified Kampala trauma score (KTS) effectively predicts mortality in trauma patients. Injury 2015 Jul 20.

(229) Weeks SR, Juillard CJ, Monono ME, Etoundi GA, Ngamby MK, Hyder AA, *et al*. Is the Kampala trauma score an effective predictor of mortality in low-resource settings? A comparison of multiple trauma severity scores. World J Surg 2014 Aug;38(8):1905-1911.

(230) Haac B, Varela C, Geyer A, Cairns B, Charles A. The utility of the Kampala trauma score as a triage tool in a sub-Saharan African trauma cohort. World J Surg 2015 Feb;39(2):356-362.

(231) Schluter PJ. Trauma and Injury Severity Score (TRISS): is it time for variable re-categorisations and re-characterisations? Injury 2011 Jan;42(1):83-89.

(232) Demetriades D, Chan L, Velmanos GV, Sava J, Preston C, Gruzinski G, *et al*. TRISS methodology: an inappropriate tool for comparing outcomes between trauma centers. J Am Coll Surg 2001 Sep;193(3):250-254.

(233) Hoffmann M, Lefering R, Rueger JM, Kolb JP, Izbicki JR, Ruecker AH, *et al*. Pupil evaluation in addition to Glasgow Coma Scale components in prediction of traumatic brain injury and mortality. Br J Surg 2012 Jan;99 Suppl 1:122-130.

(234) Neale R, Rokkas P, McClure RJ. Interrater reliability of injury coding in the Queensland Trauma Registry. Emerg Med (Fremantle) 2003 Feb;15(1):38-41.

(235) Gakwaya A, Kigula-Mugambe JB, Kavuma A, Luwaga A, Fualal J, Jombwe J, *et al*. Cancer of the breast: 5-year survival in a tertiary hospital in Uganda. Br J Cancer 2008 Jul 8;99(1):63-67.

(236) Bhangu A, Kolias AG, Pinkney T, Hall NJ, Fitzgerald JE. Surgical research collaboratives in the UK. Lancet 2013 Sep 28;382(9898):1091-1092.

(237) Pinkney TD, Calvert M, Bartlett DC, Gheorghe A, Redman V, Dowswell G, *et al*. Impact of wound edge protection devices on surgical site infection after laparotomy: multicenter randomised controlled trial (ROSSINI Trial). BMJ 2013 Jul 31;347:f4305.

(238) National Surgical Research Collaborative. Multicenter observational study of performance variation in provision and outcome of emergency appendicectomy. Br J Surg 2013 Aug;100(9):1240-1252.

(239) Bhangu A, Fitzgerald E, Harrison E. 2016; Available at: 2.

(240) 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 Jun;208(6):1009-1016.

(241) Paruch JL, Ko CY, Bilimoria KY. An opportunity to improve informed consent and shared decision making: the role of the ACS NSQIP Surgical Risk Calculator in oncology. Ann Surg Oncol 2014 Jan;21(1):5-7.

(242) Khavanin N, Kim JY, Davila AA, Ver Halen JP, Mlodinow AS, Bethke KP, *et al*. Abstract 46: The BRA Score: Creating a General Risk Calculator for Breast Reconstruction Outcomes. Plast Reconstr Surg 2014 Mar;133(3 Suppl):56-57.

(243) Neumayer L, Hosokawa P, Itani K, El-Tamer M, Henderson WG, Khuri SF. Multivariable predictors of post-operative surgical site infection after general and vascular surgery: results from the patient safety in surgery study. J Am Coll Surg 2007 Jun;204(6):1178-1187.

(244) Johnson RG, Arozullah AM, Neumayer L, Henderson WG, Hosokawa P, Khuri SF. Multivariable predictors of post-operative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. J Am Coll Surg 2007 Jun;204(6):1188-1198.

(245) Anderson JE, Chang DC. Using electronic health records for surgical quality improvement in the era of big data. JAMA Surg 2015 Jan;150(1):24-29.

(246) Cohen ME, Bilimoria KY, Ko CY, Richards K, Hall BL. Effect of subjective pre-operative variables on risk-adjusted assessment of hospital morbidity and mortality. Ann Surg 2009 Apr;249(4):682-689.

(247) Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol 2012 Sep 15;176(6):473-481.

(248) Bilimoria KY, Kmiecik TE, DaRosa DA, Halverson A, Eskandari MK, Bell RH,Jr, *et al*. Development of an online morbidity, mortality, and near-miss reporting system to identify patterns of adverse events in surgical patients. Arch Surg 2009 Apr;144(4):305-11; discussion 311.

(249) Meguid RA, Ahuja N, Chang DC. What constitutes a "high-volume" hospital for pancreatic resection? J Am Coll Surg 2008 Apr;206(4):622.e1-622.e9.

(250) Laing GL, Bruce JL, Skinner DL, Allorto NL, Clarke DL, Aldous C. Development, implementation, and evaluation of a hybrid electronic medical record system specifically designed for a developing world surgical service. World J Surg 2014 Jun;38(6):1388-1397.