High Performance

in

Surgery

Alex Almoudaris BSc (Hons.), MBBS, MRCS

Division of Surgery and Cancer, Imperial College London

Thesis submitted to Imperial College London for a Doctorate in Philosophy (PhD)
2015

ABSTRACT

The national identification of high performing providers in surgery is of prime importance to patients, surgeons and commissioners of healthcare. This thesis explores how high performance is identified, defined and measured nationally and attempts to identify the factors that underlie high performance in colorectal cancer surgery during the peri-operative period.

An introduction into the determinants of high performance in surgery as well as defining quality as it pertains to surgery is then undertaken. Identification of available national data sources and metrics for national performance are then identified. Comparison is made between voluntary and compulsory reporting systems highlighting greater capture of peri-operative mortality in compulsory reporting datasets.

A novel marker that reflects outcome following complication management is developed. This marker is based on re-operations and is derived from compulsory reporting datasets. The use of non-operative re-interventions is then assessed in oesophago-gastric cancer resections as proof of concept. An appraisal of all colorectal cancer units in England is then undertaken using a panel of metrics demonstrating that analysis on a single marker alone may be too simplistic.

Identifying factors that pertain to high performance beyond those available from routinely available datasets using a novel methodological approach called HiPer (High Performance) is performed. The interview based methodology identified rich qualitative factors in a group of colorectal cancer units worldwide that may be causal in their performance status.

Finally, results from the interview study were related to hard outcome data from each unit which demonstrated some correlation between the HiPer methodology and the outcome data in the final section of the feasibility study. The implications of this may be that a dual approach of analysing routinely collected data with a more qualitative HiPer style methodology may help us better understand how high performing units achieve their results.

DEDICATION

To my father

ACKNOWLEDGEMENTS

Firstly I must thank my supervisors Charles and Omar for their complete belief and support. I am grateful for their time and guidance both academically and personally.

Secondly my colleagues Elaine Burns, Ravi Mamidanna, Nick Symons, Hutan Ashrafian, Danilo Miskovic and Anna Pinto are acknowledged for their individual help and assistance during the whole process.

Finally I must thank my family for their unwavering support during this time.

DECLARATION

I hereby declare that I am the sole author of this thesis and that all the work within this is my own, except where it is referenced or carried out in collaboration with others who are appropriately credited.

Signed: Date:

Alex Almoudaris BSc(Hons.), MBBS, MRCS

PEER REVIEWED PUBLICATIONS (AS THESIS CHAPTERS)

Failure to rescue patients following re-intervention in gastric and oesophageal cancer surgery in England
AM Almoudaris, R Mamidanna, A Bottle, P Aylin, C Vincent, O Faiz, G Hanna
JAMA Surg. 2013 Mar;148(3):272-6. doi: 10.1001/jamasurg.2013.791.
PMID: 23553312

Single measures of performance do not reflect overall institutional quality in colorectal cancer surgery
AM Almoudaris, E Burns, R Mamidanna, A Bottle, P Aylin, C Vincent, O Faiz
Gut. 2013 Mar;62(3):423-9. doi: 10.1136/gutjnl-2011-301489.
PMID: 22345658

Value of failure to rescue as a marker of the standard of care following reoperation for complications after colorectal resection
AM Almoudaris , E Burns , R Mamidanna, A Bottle , P Aylin , C Vincent, O Faiz
Br J Surg. 2011 Dec;98(12):1775-83. doi: 10.1002/bjs.7648. Epub 2011 Aug 25.
PMID: 22034183

A colorectal perspective on voluntary submission of outcome data to clinical registries.
AM Almoudaris, E. Burns, A. Bottle, P. Aylin, A. Darzi, Omar Faiz
Br J Surg. 2011 Jan;98(1):132-9. doi: 10.1002/bjs.7301. Epub 2010 Oct 26.
PMID: 21136567

Establishing Quality Care in Colorectal Surgery- Review Article
AM Almoudaris, S. Clark, C. Vincent, O. Faiz
Colorectal Dis. 2011 Sep;13(9):961-73. doi: 10.1111/j.1463-1318.2010.02355.x. Epub 2010 Jun 28. Review.
PMID: 2059420
Awarded Editors feature paper at ESCP conference 2011

PEER REVIEWED PUBLICATIONS (ADDITIONAL PUBLICATIONS)

A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery.
Johnston MJ, Arora S, King D, Bouras G, Almoudaris AM, Davis R, Darzi A.
Surgery. 2015 Apr;157(4):752-63. doi: 10.1016/j.surg.2014.10.017.
PMID: 25794627

Failure to rescue in trauma patients: operative interventions must be considered.
Almoudaris AM, Mamidanna R, Faiz O.
Ann Surg. 2014 Jun;259(6):e85. doi: 10.1097/SLA.0000000000000538.
PMID: 24441807

Hierarchical multilevel analysis of increased caseload volume and postoperative outcome after elective colorectal surgery.
Burns EM, Bottle A, Almoudaris AM, Mamidanna R, Aylin P, Darzi A, Nicholls RJ, Faiz OD.
Br J Surg. 2013 Oct;100(11):1531-8. doi: 10.1002/bjs.9264.
PMID: 24037577

Mortality in high-risk emergency general surgical admissions.
Symons N, Moorthy K, Almoudaris AM, Bottle A, Aylin P, Vincent C, Faiz O
Br J Surg. 2013 Sep;100(10):1318-25. doi: 10.1002/bjs.9208. Epub 2013 Jul 17.
PMID: 23864490

National outcomes and uptake of laparoscopic gastrectomy for cancer in England.
Mamidanna R, Almoudaris AM, Bottle A, Aylin P, Faiz O, Hanna GB.
Surg Endosc. 2013 Sep;27(9):3348-58. doi: 10.1007/s00464-013-2916-5.
PMID: 23612763

An observational study of the frequency, severity, and etiology of failures in postoperative care after major elective general surgery.
Symons NR, Almoudaris AM, Nagpal K, Vincent CA, Moorthy K.
Ann Surg. 2013 Jan;257(1):1-5. doi: 10.1097/SLA.0b013e31826d859b.
PMID: 2304478

Poor 1-year survival in elderly patients undergoing nonelective colorectal resection.
Mamidanna R, Eid-Arimoku L, Almoudaris AM, Burns EM, Bottle A, Aylin P, Hanna GB, Faiz O.
Dis Colon Rectum, 2012Jul;55(7):788-96. doi:10.1097/DCR.0b013e3182585a35
PMID: 2270613

PRESENTATIONS- INVITED

Royal Society of Medicine (RSM)
Keeping Czech on quality in healthcare: Lessons from an RSM Fellowship in Prague
Royal Society of Medicine, London, June 2012
Invited talk at the annual pioneer lecture - Sir Muir Gray, RSM,

British Association of Surgical Oncologists (BASO)
Failure to Rescue- Surgical (FTR-S): Defining an important metric in surgery
Royal College of Surgeons, England, November 2011

PRESENTATIONS-INTERNATIONAL

Failure to rescue following re-intervention in gastric and oesophageal cancer surgery in England
United European Gastroenterology Week Conference, Amsterdam October 2012
AM Almoudaris, R Mamidanna, A Bottle , P Aylin , C Vincent, O Faiz, G Hanna

How do postoperative mortality 'outlier' institutions for colorectal cancer surgery perform on other measures of quality?
European Society of Coloproctologists, Copenhagen, September 2011
AM Almoudaris, E Burns, A Bottle, P. Aylin, C Vincent and O Faiz

Failure To Rescue-Surgical (FTR-S): Defining an important metric in surgery
Tripartite Colorectal meeting of the Associations of Coloproctologists of Great Britain & Ireland, United States and Australasia, Cairns- Australia July 2011
AM Almoudaris, E Burns, A Bottle, P. Aylin, C Vincent and O Faiz

PRESENTATIONS -NATIONAL

Failure To Rescue-Surgical (FTR-S): Defining an important metric in surgery
Association of Coloproctologists Great Britain & Ireland,
Birmingham, May 2011
AM Almoudaris, E Burns, A Bottle, P. Aylin, C Vincent and O Faiz
WINNER of the British Journal of Surgery prize

Should submission to the National Bowel Cancer Audit be mandated?
Association of Coloproctologists Great Britain & Ireland,
Bournmouth, May 2010
AM Almoudaris, E Burns, A Bottle, P. Aylin, A Darzi and O Faiz
Shortlisted for the British Journal of Surgery prize

Do units that submit data to a voluntary colorectal clinical registry have better outcomes than those that do not?
Association of Surgeons Great Britain & Ireland,
Liverpool, April 2010
AM Almoudaris, E Burns, A Bottle, A Darzi and O Faiz

PRESENTATIONS -REGIONAL

Failure to rescue following re-intervention in gastric and oesophageal cancer surgery in England
London's 6th Surgical Symposium 2012, Imperial College, September
AM Almoudaris , R Mamidanna, A Bottle , P Aylin , C Vincent, O Faiz, G Hanna

How do postoperative mortality 'outlier' institutions for colorectal cancer surgery perform on other measures of quality?
London's 5th Surgical Symposium 2010, Imperial College, September
AM Almoudaris, E Burns, A Bottle, P. Aylin, C Vincent and O Faiz

Re-operation Failure To Rescue Rate (r-FTR) - A novel metric in colorectal cancer surgery?
London's 4th Surgical Symposium 2010, Imperial College, September
AM Almoudaris, E Burns, A Bottle, P Aylin, C Vincent and O Faiz

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACPGBI	Association of ColoProctology of Great Britain and Ireland
APE	Abdomino-Perineal Excision
APER	Abdomino-Perineal Excision Rate
AR	Anterior resection
ASA	American Society of Anaesthesiology
ASCO	American Society of Clinical Oncology
CABG	Coronary Artery Bypass Graft
CCT	Controlled Clinical Trials
COST	Clinical Outcomes of Surgical Therapy
CPEX	CardioPulmonary Exercise Testing
CQC	Care Quality Comission
CRAB	Cancer Recurrence and Blood Transfusion
CRM	Circumferential Resection Margin
CT	Computed Tomography
DVT	Deep Vein Thrombosis
ERP's	Enhanced Recovery Programe
EWS	Early Warning Systems
FACT-C	Functional Assessment of Cancer Therapy – Colorectal Scale
FCE	Finished Consultant Episode
FTR	Failure To Rescue
FTR-S	Failure To Rescue-Surgical
GMC	General Medical Council
GMT	Greenwich Mean Time
HCFA	Health Care Financing Administration
HDU	High Dependency Unit
HES	Hospital Episodes Statistics
HiPer	HiPerformance
HMO	High Mortality Outlier
HMQ	High Mortality Qunitile
HRQoL	Health Related Quality of Life
IC	Information Commission
ICD-10	International Classification of Disease- Version 10
ICU	Intensive Care Unit
IOM	Institute Of Medicine
IRB	Institutional Review Board
JAMA	Journal of the American Medical Association
LMO	Low Mortality Outlier
LMQ	Low Mortality Quintile
LOS	Length Of Stay
MDT	Multi-Disciplinary Meeting
MeSH	Medical Subject Headings
MEWS	Modified Early Warning Scores
MGH	Massachusetts General Hospital
MIA	Minimally Invasive Approach
NBOCAP	National BOwel Cancer Audit Programme
NCASP	National Clinical Audit Support Unit
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NHS	National Health Service

NICE	National Institute of Clinical Excellence
NSQIP	National Surgical Quality Improvement Programe
OPCS-4	Office of Population Censuses and Surveys- Version 4
OR	Odds Ratio
PAS	Patient Administration Systems
PE	Pulmonary Embolus
PET	Positron Emission Tomography
POA	Present On Admission
POSSUM	Physiological and Operative Severity (Score)
PROMS	Patient Reported Outcome Measures
PSI’s	Patient Safety Indicators
QoL	Quality of Life
RCT	Randomised Control Trial
SD	Standard Deviation
SF-36	Short Form-36
SHMI	Summary Hospital-level Mortality Indicator
SMR	Standardised Mortality Ratio
SPSS	Statistical Package for the Social Sciences
SRR	Standardized Rate Ratios
TME	Total Mesorectal Excsion
US	United States
VPN	Virtual Private Network
VSQIP	Vascular Surgical Quality Improvement Program
X-rays	X-radiation

1.0 WHY SEEK OUT HIGH PERFORMANCE IN SURGERY?

1.1 Chapter overview

This chapter sets the context for the thesis by identifying why it is of value and necessary to identify high performance in surgery. Previously demonstrated variation in surgical outcome across different specialities and the consequences of such variation are discussed. The definition of high performance in the context of the quality of surgical practice then follows.

1.2 High performance in surgery

In healthcare it is paramount to ensure firstly no harm is done, but secondly, that any interventions undertaken serve our patients well. By identifying outcome differences following surgical intervention, it may be possible to understand the causes of such variability. Where the variation is undesired, recognition could lead to efforts to improve it. When taken into the context of the estimated 234 million major surgical procedures undertaken each year world-wide (Weiser et al., 2008) any marginal improvements in outcome that are adopted, may potentially benefit large numbers of patients.

1.4 Defining High Quality in surgery

The need to achieve high quality service provision within healthcare is a global desire. The term quality will be attributed to more qualitative aspects of care and performance will be used to reflect those aspects of care that are traditionally more easily measured- so called quantitative aspects. Prioritisation of 'quality' as a central theme in future healthcare planning in the United Kingdom has been demonstrated recently by an influential government report entitled 'High Quality Care for All' (Crown, 2002). Universal healthcare improvement initiatives, such as the development of a pre-operative surgical checklist by the World Health Organization, further reflect the contemporary global commitment to prioritising high quality care within surgery (Haynes et al., 2009).

Having a clear definition of what quality in surgery means as well as how to measure it, are pre-requisites for improving surgical standards. Although it is accepted that service providers should strive to improve quality, what exactly does quality mean to surgeons and their patients? Previous attempts have been made to define quality markers in surgery and specifically in colorectal cancer surgery by using qualitative research methodologies (McGory et al., 2006).

Delivery of quality care

Having understood what the elements of quality assessment may be, it is natural to try to understand how such quality is delivered in practice. Maxwell in his influential paper: "Quality Assessment in health", outlines six dimensions of health care quality. Recognition and measurement of his proposed six dimensions are necessary to prove to stakeholders whether a service represents quality care or not according to him.

Table 1- Table summarising Maxwell's Six dimensions

- *Access to services*
- *Relevance to need (for the whole community)*
- *Effectiveness (for individual patients)*
- *Equity (fairness)*
- *Social acceptability*
- *Efficiency and economy*

Maxwell's Six dimensions from Quality in Health, RJ Maxwell, BMJ 12/5/1984

Maxwell when critiquing the delivery of healthcare asked the questions of whether patients had access to care wherever and whenever they needed it. Do all patients with similar needs get the same care? Plus ensuring that the care they receive is acceptable to them. Furthermore, does the care that is provided to them do what it is intended on doing and at a reasonable cost with minimal waste? Finally, is the care received and delivered sensitive and appropriate when considering the need of the whole community? Not only does an appreciation of Maxwell's dimensions help us understand delivery of healthcare but it also allows us to understand variability in such delivery. If a quality standard is set, variability beyond certain thresholds from this may represent exceptional or sub-optimal care.

The Institute Of Medicine (IOM) produced six domains of quality care that should be sought in an ideal healthcare system. These six domains are listed below-

Table 2- The Institute of Medicine domains of quality

Safety	Minimizing medical errors and adverse events
Effectiveness	Maximizing intended health outcomes
Patient-centeredness	Focusing on patient and family comprehension, preferences, goals and priorities in making treatment decisions
Timeliness	Minimizing delay between onset of illness and initiation of treatment
Efficiency	Providing maximally cost-effective care
Equity	Providing care of equal quality regardless of gender, ethnicity, region, socioeconomic status or insurance coverage

With respect to surgery, each of these domains is indeed very relevant to the pursuit of high quality care. Ensuring that patients that require operations receive the correct operation at the correct time, irrespective of location or socioeconomic status, and without complication is vital to ensuring the best outcome possible.

Therefore Maxwell's domains in part are just as relevant to surgery as they are to wider healthcare systems.

1.4.1 Defining and measuring quality in surgery

The National Institute of Medicine in the United States (US) defines quality as the '*degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*' (Institute of Medicine, 1990). Importantly, this definition recognises the importance of the structural factors and surgical service processes that may enhance the likelihood of achieving positive outcomes. Furthermore, emphasis upon contemporary practice is echoed within the definition.

In the United States (US) the Agency for Healthcare Research and Quality defines a quality measure as '*a mechanism that enables the user to quantify the quality of a selected aspect of care by comparing it to a criterion*' (Center for Health Policy Studies, 1995). According to Bergman and colleagues an ideal quality metric should be universally relevant to clinicians, decision-makers and patients (Bergman et al., 2006). Moreover, as suggested by Mayer and co-workers (Mayer et al., 2009b), a structured quality framework should incorporate both clinical pathway measures and patient reported outcomes. The former objective measures are arguably of greater relevance to clinicians whilst the latter subjective measures perhaps better reflect the patient perspective. Furthermore, metrics must be able to discriminate between good and poor quality and should also be resilient to ‘gaming’ - i.e. achieving apparent quality improvement through chasing targets rather than through genuine quality improvement (Smith, 1995).

Thus defining quality in surgery would acknowledge each of these factors. Demonstrable timely assessment/intervention using evidence based foundations that improve a patient’s health status in a manner acceptable to patients would represent a generalised proposed definition of quality in surgery. High quality would be the demonstration of the above factors reproducibly and with minimal variation given presenting case-mix.

1.3 Variation in practice and surgical outcome

Variation in practice and outcome occurs throughout healthcare. Taking the most commonly performed procedure in emergency surgery- the appendicectomy, this seems to be the case. Worldwide circa 5-10% of patients undergoing emergency open appendicectomy may develop a superficial wound infection (Shaikh et al., 2009). The practice of open removal of the appendix has been fairly standardized from its first description in 1735 by Claudius Amyand. This being the case, why is there a twofold difference in this outcome? If outcome can differ by such a magnitude for a procedure that most would agree is performed in a standardized fashion almost universally, what is expected from more complex procedures? An attempt to understand how and why such variation occurs is thus a critical precursor to understanding high performance in surgery.

1.3.1 Variability in tonsillectomy rates in England

The first published description of surgical variation has been attributed to and demonstrated by James Allison Glover (1874-1963). In 1938 he demonstrated a 10 fold variation in tonsillectomy rates in England, United Kingdom (Glover, 1938). It has been reported that it was only when Glover formally published his findings that variability in surgical practice was formally recognized in the literature. More striking than the procedural variation was the 8 fold difference in risk of death that Glover also identified across the nation.

Glover noted that despite similar age and sex demographics, children in more affluent areas were three times more likely to have undergone tonsillectomy. The findings also uncovered that diametrically opposite rates often occurred in neighboring counties which were unexplainable at the time. However what Glover postulated was that the differing rates were more likely to be a function of physicians' decision making and beliefs than actual clinical need. Whilst some degree of variation is expected due to differing populations, presentation and natural disease course, the notion of unwarranted variation is central to understanding why variation in outcome is important to improving patient outcomes. In the former example, unnecessary procedures would have also subjected the children to the associated morbidity of surgery and risk of death, without apparent clinical benefit. Such an example highlights why the study of unwarranted variation is important.

1.3.2 Mortality in US Cardiac centers

In 1987 hospital level coronary artery by-pass graft rates (CABG) were published by the Health Care Financing Administration (HCFA), the then administrator of the national Medicare program in the United States of America (USA). This document highlighted wide variation in outcome from CABG procedures (Publication, 1986). This led a group of cardiac surgeons and epidemiologists from New England (USA) to undertake a prospective audit of their results as they were convinced their outcomes (which were worse than the national median) were attributable to them operating on a more co-morbid population that had not been properly accounted or 'adjusted' for by the HCFA report. In 1991 the results of the participating 5 hospitals were published in the Journal of the American Medical Association (JAMA) and the authors concluded-

"the observed differences in in-hospital mortality rates among institutions and among surgeons in northern New England are not solely the result of differences in case mix as described by these variables and may reflect differences in currently unknown aspects of patient care" (O'Connor et al., 1991).

In other words the authors are describing a scenario where they believe their outcomes are not expected given the patients and the pathology that presented to their service. Their outcome lay outside the expected control limits and hence required further investigation. This seminal piece of work identified that the study of populations can reveal unwarranted variations in surgical outcome.

1.3.3 Dartmouth Atlas

At the same time, in the United States what is now known as the Dartmouth Atlas of Healthcare was being devised with the arrival of John E. Wennberg to Dartmouth College. The Dartmouth Institute as it later became known began to publish outcome data by common medical and surgical conditions and by region of the United States. The purpose of this undertaking was to unmask any variation in practice and outcome in the country (College., 1996) in healthcare in general. Broadly the findings of the atlas demonstrated wide variation both in the procedures performed for matched pathologies but also in hard outcome measures across most specialties and procedures examined. More specifically, in its first edition several surgical procedures were analyzed. In particular with respect to breast surgery the atlas uncovered significant variation in the use of breast sparing surgery versus more radical surgery. It was discovered that the proportion of women undergoing breast sparing surgery varied by a factor of 33 across the United States from the examined Medicare patients. Looking at another surgical procedure from the same report it reported a greater than ten-fold difference in the rates of radical prostatectomies (0.6 to 7.0 per thousand men) after adjusting for differences in age and race. As Glover surmised, the differences in such rates in the US were not attributable to any conceivable factors at the time. It was becoming apparent that even in advanced healthcare systems great variability occurred without apparent causality given the presenting populations and pathologies. For this reason a pressing need developed to understand why such variability existed.

1.3.4 Variation in Abdomino-perineal excision rates in England

A contemporary English study by Morris and colleagues looked at the differences in Abdomino-perineal excision rates (APER) for rectal cancer in England using routinely collected data (Morris et al., 2008). Rectal cancer can be treated using two main surgical methods. Excision of the tumour and maintaining intestinal continuity at the same or subsequent procedure is termed an anterior resection. Excision of the tumour and anus and leaving an end stoma is termed an Abdomino-perineal excision (APE). Tumours that are too close to the anus are not amenable to salvaging the anus to maintain oncological clearance. The study demonstrated variation in the use of APE resections ranging from 8.5% to 52.6% by hospital Trust from 1998 to 2004 (Morris et al., 2008). The recommended target figure by national bodies is 30% (ACPGBI). Although this initial study did not take into the height of the tumour above the anal margin (a key consideration in determining operative approach) or case-load, the findings were still of relevance to questioning contemporary colorectal practice. Subsequent analysis by Morris on a sub-group (only Trusts in Northern & Yorkshire Regions) demonstrated that units with higher APE rates were also those operating on tumours higher up in the rectum (with a correlation co-efficient of 0.71). Clinically it would be expected that lower tumours would be considered for the more radical APE procedure. The sub-group analysis goes against expected and recommended practice and supported initials calls that further investigation was warranted (Morris et al., 2008).

1.3.5 Atlas of variation England

More recently in England, routinely collected data have been the basis of a department of health initiative entitled The Atlas of Variation (Care., 2010). This report first published in November 2010 described its purpose as demonstrating "*variation and the utilization of health care services that cannot be explained by variation in patient or patient preferences*" in England. The atlas uncovered some 38 fold variation in the use of bariatric surgical procedures across the country. Some two fold difference in the undertaking of cataract extraction surgery and 30 fold differences in primary hip replacement; all standardized for the catchment populations they represented in England. These figures and the magnitudes of differences are clearly undesirable in an equitable health care system. However, although the variation in processes in the examples given may be worrying, apart from the cardiac surgery audit from New England, outcomes have not been demonstrably affected by similar magnitudes. This asks the question of whether variability in the number and type of procedures undertaken is relevant if seemingly little outcome differences exist nationally.

1.3.6 Bristol Paediatric Cardiac Surgery 1984-1995

Between 1991-1995 the post-operative mortality rate for children undergoing cardiac surgery aged less than one was twice the national average with a greater difference observed for babies under the age of 30 days at Bristol Royal Infirmary. Criticism has come from the delay in recognizing such variation in outcome. The number of procedures and volumes in themselves were not worrying. In the official report to the events at Bristol, it was noted that nationally comparable data existed in the form of Hospital Episodes Statistics (HES) from 1990 and "*it was not* *recognized as a valuable tool for analyzing the performance of hospitals. It is now, belatedly*" (Health.). This is just one example whereby variation in outcome if recognized should have led to further investigation into practice and may have prevented unnecessary deaths. This takes us on to consider whether outcome variation is indeed a relevant marker of quality. It is seen from the subsequent enquiry into the failings at the Bristol Royal Infirmary several recommendations were set out. Of relevance these included-

1. *There must be a single, coherent, co-ordinated set of generic standards: that is, standards relating to the patient's experience and the systems for ensuring that care is safe and of good quality.*
2. *The monitoring of clinical performance at a national level should be brought together and co-ordinated.*
3. *The new system should provide a mechanism for surveillance whereby patterns of performance in the NHS which may warrant further scrutiny can be identified as early as possible.*
4. *At national level, the indicators of performance should be comprehensible to the public as well as to healthcare professionals. They should be fewer and of high quality, rather than numerous but of questionable or variable quality.*
5. *The Hospital Episode Statistics database should be supported as a major national resource which can be used reliably, with care, to undertake the monitoring of a range of healthcare outcomes (Copyright, 2001).*

The points highlighted from the enquiry into the failings at Bristol are very salient to understanding why measurement of variation in healthcare and specifically surgery is important. The final report identified that national measures of performance should be used to identify any unwanted variation in practice.

Furthermore they identify the HES database as a potential tool for taking this forward.

In order to identify variation there must be some appreciation of a high performing unit/department with which to set benchmarks. How or where are such units identified? Whilst it is seen that variation in surgery is inevitable, each patient is unique, as is the pathology, when case-mix differences are accounted for it would be expected that outcome should be predictable. For example, a unit that only operates on recurrent rectal cancers would, it is expected, have differing outcome form those operating on primary early cancers. However, when the aforementioned units are compared to similar units as themselves undertaking similar case-load and case-mix, it would be expected that variation between such similar units be minimal if there is equity in the care delivered. Thus, the extent of acceptable variation may be different for each type of unit dependent on the presenting workload.

However, in surgical practice variation in outcome is not necessarily inevitable. There are centers of surgical excellence around the world and better understanding how they deliver their care and achieve their results would be a natural starting point.

1.4 Excellence in practice and surgical outcome

Whilst it may seem that given inevitable variation in patients presenting health status- for example body mass index, co-morbidities and presenting degree of severity of illness, outcome variation is inevitable too. However there are centers that despite such variation report consistent high quality outcome. By looking further at an example of such a unit it may be possible to further elucidate how such results are achieved.

1.4.1 The Shouldice Hospital, Canada

One unit that is unequivocally regarded as a center of excellence and high performance is the Shouldice Hospital in Canada. This hospital was founded in 1945 and is the unquestionable world leader in primary inguinal hernia repair in terms of outcome. The hospital employs 150 people and has a greater than 99.5% success rate in repair. They report having the lowest recurrence rates in the world for the past 65 years (Hospital, 2012). Operating surgeons are required to undertake >700 cases per year. Patients are offered diet, weight loss advice and massage therapy to optimize their outcome. All patients have the same assessment, local anaesthetic operations without a mesh and where possible undertaken in exactly the same fashion every time. They are immediately given an exercise routine and asked to mobilise and encouraged to do so. The care is not limited to the operative procedure itself but Shouldice pride themselves on a standardized care package where the post-operative care is as important as the pre- or inter-operative care. In terms of cost, a surgeon and healthcare commentator Atul Gawande reported that most hernia repairs cost in the region of $4000 in the US compared to the $2000 at the Shouldice hospital. The key to the success he reported was the automation associated with dealing with unfamiliar situations. Most general surgeons will not do as many hernia repairs in their life-time compared to one to two years as a Shouldice surgeon. As a result Gawande argues, "*Shouldice surgeons are more familiar with how things should look and any deviations are immediately apparent to them*" (Atul Gawande, 1998).

Perhaps what favours the Shouldice clinic is not only the volume of similar work undertaken but also the whole patient pathway relies upon standardization of care. Whilst this level of standardization and outcome is immediately attractive, some degree of patient selection is likely to be undertaken; but this is supposition as such data are not publically available. Arguably, although there is no published evidence, patients may be selected for their likelihood of success. What is known is that obese patients are counseled and asked to lose weight before being offered surgery and more complex procedures are not undertaken by the unit. This reflects to some degree the selection process that is known about with respect to the Shouldice hospital.

Whether such a level of standardization is possible in a national health service is debatable. An obese patient that is reportedly house bound and unable to engage in exercise because of hernia associated pain is a difficult patient to refuse despite the increased risk of recurrence and perioperative morbidity. Furthermore, whilst using the Shouldice clinic as the exemplar of high performance in the setting of elective hernia repair, comparative centers of excellence are arguably more difficult to identify when moving to more complex and varied procedures.

1.5 Summary of chapter

In this chapter it has been demonstrated how variation in both procedures and practice occurs in surgery. This is not always attributable to the presenting demographics of the population. By analysing those units of excellence and high performance and learning from their methods and systems, it may possible to reduce unwanted variation in healthcare systems if these prove to be applicable more widely. However, what remains to be identified is how to identify such units and what measures should be used for the comparisons? This is the next fundamental question and what follows is an appraisal of what exactly does surgical quality and high performance mean in the context of what is measureable.

2.0 MEASURING QUALITY/HIGH PERFORMANCE IN SURGERY

2.1 Chapter overview

In this chapter conceptual frameworks for assessing and measuring quality in healthcare are visited. These will firstly consider Donabedian's structure-process-outcome paradigm. A detailed analysis of each of these components is generically made with respect to its application to surgical practice. Following on from this, it is appreciated that the Donabedian paradigm does not capture some of the more qualitative elements that may be important to consider, such as team-work and leadership. These types of factors are considered by introducing Lilford's concept of intervening variables, which highlights factors such as teamwork and leadership which are potentially important determinants of surgical outcome.

2.2 Donabedian Structure Process Outcome

A paradigm that serves as a useful means of categorising and understanding quality in healthcare is the Avendis Donabedian 1919-2000, structure-process-outcome approach (Donabedian). An Armenian General Practitioner in Jerusalem who subsequently moved to the United States to take up a post as a Public Health Professor is credited with providing a framework for understanding quality appraisal in healthcare systems. Shortly after joining the University of Michigan, Donabedian was given the role of revising and clarifying contemporary writings on quality assessment. By summarizing every article in press at the time on small index cards, Donabedian realized some common themes. These common themes provided the framework around which the quality of medical care could be understood and his proposition that quality of care could be considered using the mantra of structure-process-outcome. His seminal work entitled: "Evaluating the Quality of medical care" was published in 1966 in the Milbank Memorial Fund quarterly and to this date is one of the most cited scientific articles of its century (Donabedian, 1966). He proposed a methodology for systematically assessing a healthcare system and its constituent components for the purpose of better understanding it. This approach has given a framework for many social scientists to develop a way of understanding and attempting to measure quality in practice within a healthcare setting.

2.3 Structural measures

Structural measures are relatively straightforward to measure and quantify. They tend to relate to the physical resources of an institution, be they human- as in staff numbers and experience and specialty- to availability of equipment and equipment for investigations. They tend to change very little with short to medium timeframes and thus measurements can be discretely performed rather than continuously. Structural measures can include anything from submission of data to outcome registries to the volume of cases undertaken by surgeons and units - these specific factors will be considered in greater detail in the subsequent chapter.

2.4 Process measures

Process measures reflect the actions by the medical teams that engage the patient directly. Donabedian in his summary regarding process measures concluded that in their description they may be better suited to answering the question of "whether medicine is properly practiced". These measures are now common place in use as quality indicators. Measures include the use of thrombophrophylaxis in surgical patients and antibiotics for wound prophylaxis at surgical induction and these are also substantiated with a large body of clinical evidence including many randomised clinical trials and meta-analyses (Gomez-Outes et al., 2012, Nelson et al., 2009, Sanabria et al., 2007). The attraction of using process measures is in part due to the relative ease in which interventions can be implemented or omitted and duly any outcome differences be purportedly related to the change. There is also evidence whereby implementation of process measures has led to quality and outcome improvement specifically in the use of perioperative beta-blockers. In the highly cited study perioperative cardiac outcome of patients was improved with appropriate instigation of blood pressure control versus those that did not have this process instigated (Mangano et al., 1996). This caused a paradigm shift at the time in anaesthetics.

2.5 Outcome measures

Outcome measures are the least challenging to understand and many identify their measurement as robust and transparent irrefutable evidence of quality. This notion is usually too simplistic. Traditionally outcome measures have been the measures used to understand performance of hospitals and compare institutions.

Outcome measures can be thought of as what actually happens to the health of the patient - the outcome - as a result of the treatment and care they receive. The recent government white paper 'Liberating the NHS: Transparency in outcomes – a framework for the NHS' states that "at a national level the focus and accountability should, as far as possible, be centred around the outcomes of care" (Health, 2010). The report states the benefits of outcome reporting-

1. will allow **accountability** of healthcare commissioning and ultimately "*a mechanism by which the Secretary of State can hold the new NHS Commissioning Board to account for securing improved health outcomes for patients through the commissioning process*".

2. are intended to "*act as a catalyst for **driving up quality** across all NHS services"* and not for punitive performance comparisons.

Now whilst clearly each of the individual components of the Donabedian triad are important factors they each have individual qualities that are unique to them. By better understanding these qualities it allows us to see how they might be applied to a surgical system.

2.6 Characteristics of structure, process and outcome measures

What follows is a more in depth appraisal of the characteristics of the Donabedian triad of proposed measures. Each of these measures has a unique place for assessing and appraising healthcare systems and a greater understanding of these follows with specific reflection on their applicability to surgery.

2.6.1 Characteristics of structural measures

Structural measures are the readily calculated and collated factors, such as number of nurses and nurse-patient ratios or the availability of resources, such as CT scanners and operating theatres. These measures share common characteristics.

- **Efficiency**: one structural measure may be shown to impact on numerous outcome measures. This may be true across surgical specialties as well. This may make the argument for investing in improving such a measure as having widespread benefit. For example, the number of operating theatres can give an idea of the capacity of a unit and potential throughput. This would apply across many surgical specialties if the theatre usage is shared, as is commonly the case. The specialist skill set within an organization and specifically within a speciality can give insight into the potential standard of care received for that individual speciality. For example, a colorectal unit with a dedicated specialist nurse in colorectal cancer and specific multi-disciplinary cancer meetings with dedicated oncologists, radiologists and pathologists would reflect intention of delivering a high quality service (if nothing else). Thus structural factors can give insight into wider surgical capability across an institution as well as more refined aspects within a certain specialty and the slight temporal change makes analyzing such factors attractive.

- **Expediency**: many of these measures can be very quickly assessed and measured due to the fact they change infrequently. Furthermore such data are usually held by institutions or centrally without the need for extra resources to calculate.

- **Actionability**: few structural measures are readily and easily changed. This may make their use of academic importance without being practically actionable to improve results. For example, a chief executive looking to improve outcome in one department will no doubt have difficulty appointing a slew of new specialist staff and equipment and realistically expect immediate returns.
- **Discrimination:** crudely measuring structural measures may not discriminate between individual surgeons or teams. In other words there may be several departments in the hospital that share the same structural measure e.g interventional radiology. However, where one team does not appropriately make use of this facility or indeed uses the facility too late in the patient pathway, simply the presence and availability of this structural resource may not impact on the performance of the unit. Thus simply the presence of a structural measure such as intensive care beds or out-of hours endoscopy services may not reflect good care if they are inappropriately utilized. Importantly, structural measures tend to be the most difficult to influence and change. It has been noted though that structural measures are useful for setting minimum standards (Cooperberg et al., 2009).

Structural measures can be informative and be useful as predictors of future outcome. In the example given below structural measures have been shown to be more predictive of a unit's outcome in the future than relying on historic outcome data. The graph below shows units undertaking pancreatic (top graph) and oesophageal resections (bottom graph) in the USA. It shows in both instances previous risk adjusted mortality (2005-2006) poorly predicted which hospital will go on to to perform well in subsequent years (2007-2008). However in both instances hospital volume showed a monotonic effect in predicting performance. Graphs redraw using data and graphs taken from *Performance management in Surgery for ACS Surgery, Dimick & Birkmeyer [book chapter]*.

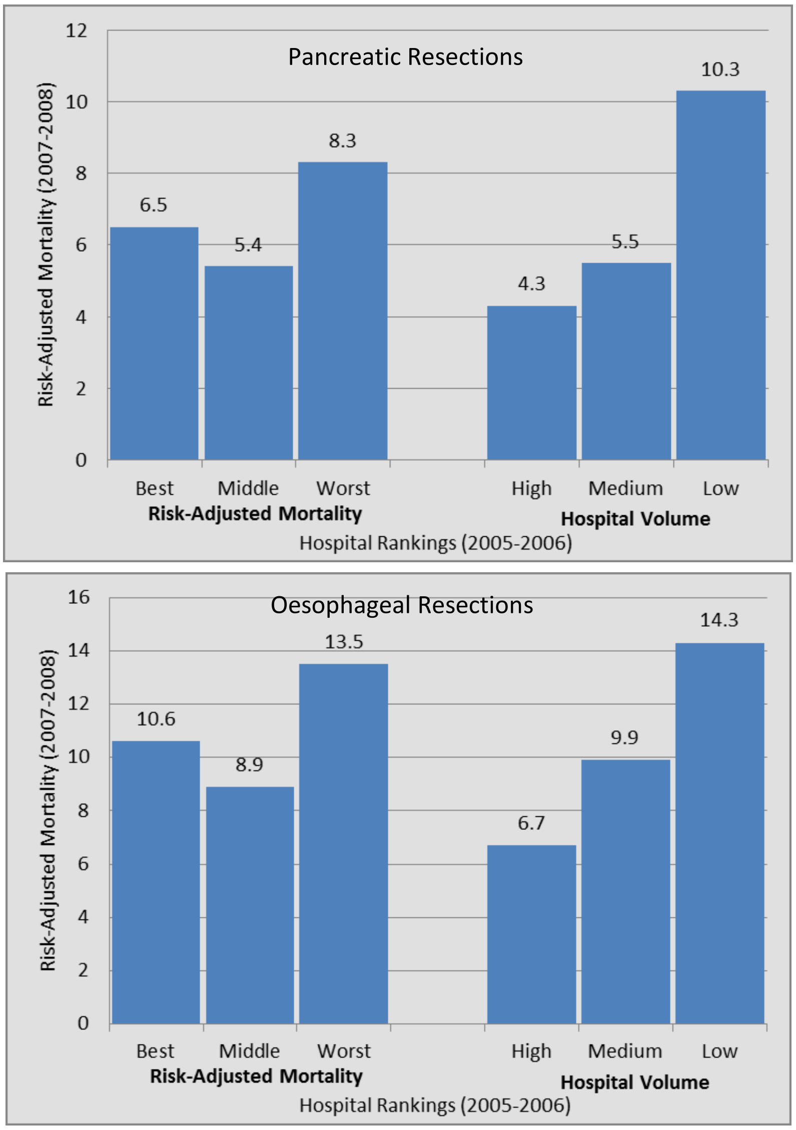


Figure 1 Relative ability of historical (2005–2006) measures of hospital volume and risk-adjusted mortality to predict subsequent (2007–2008) risk-adjusted mortality in US Medicare patients. (a) Pancreatic resection. (b) Oesophageal resection percentages shown.

If used in this manner they can be extremely useful in guiding what criteria, from minimum volume of cases required to re-certify a specialist to the minimum levels of staffing required to operate a safe emergency surgical service.

2.6.2 Characteristics of process measures

Process measures are the procedures that are actually undertaken by a unit or institution to achieve its results. These can reflect almost any aspect of a healthcare interaction from pre-operative optimization of patients to the way patients are intra-operatively monitored. As with structural measures, these also share some common characteristics.

- **Relationship**: their use may be easily related to outcome measures. When attempting to identify the impact of use, samples of patients can undergo the intervention process and another control group can be compared. Assuming all other factors are equal, the benefit/impact of the process intervention should be relatively easy to interpret, as in the example given. In other words, it is arguably easier directly to attribute the influence of process measures to outcome by comparing similar groups that do and do not undergo the process measures.

- **Actionable**: if certain process measures are found to be of use, they are, in comparison to structural measures generally easier to implement. Process measures tend to be more 'material' in the sense that they usually involve a physical intervention/action e.g. the giving of a certain medication/substance or undertaking a procedure in a novel or refined way e.g. using ultrasound guidance to facilitate central venous line placement. These tend to be more binary or discrete events whose undertaking is usually easily recorded and thus outcome followed up.

- **Abstractable**: process measures are usually more easily abstracted from data and information already held. This allows for comparisons from already collected data such as electronic patient records. Given the fact the events are usually discretely separable from 'normal' procedures they are usually easily recorded. The most notable advances in pre- and peri-operative optimization of patients and reductions in length of stay have been demonstrated and directly attributable to institution of certain process measures, such as fluid restriction and early mobilization (Khoo et al., 2007, Noblett et al., 2006).

- **Validity**: to institute a process measure into clinical care there must be good evidence that its use has an impact on outcome. This may require pooling of results, as in the case of meta-analyses, before validity can be demonstrated (Varadhan et al., 2010, Gouvas et al., 2009a).

- **Identification**: it may be difficult to identify which process measures in particular are having the beneficial effects if they are part of a group of measures established in practice. Identifying the individual impact of each process measure may be very difficult to establish in true clinical practice. Such has been the case with the so called enhanced recovery pathways, where at the outset over 12 different facets of care were instituted. More recent research has attempted to focus in on the specific facets that are of importance to outcome (Fearon et al., 2005).

Process measures however are not the panacea of quality measurement. Implementation of a process measure is no guarantee of preferential outcome, for example an ultrasound guided central venous insertion by a medical student is likely to have worse outcome than 'traditional' non-ultrasound aided placement by an expert. This hypothetical example highlights that use of a process measure needs to be appropriate for the clinical situation and patient. The application of a process measure needs to appropriate for that individual patient and not simply be undertaken to tick an audit box. Pre-operative β-blockade in a bradycardic patient would be dangerous, although many studies have shown outcome benefit if β-blockade is appropriately used (Talati et al., 2009).

2.6.3 Characteristics of outcome measures

There has been a drive towards the reporting of outcome measures in England. With the publication of national outcome from vascular and colorectal surgery, mortality and re-operation rates have been reported. These allow for rapid cross-unit and in some instances individual surgeon comparisons. The characteristics of outcome measures can also be summarised by grouping their qualities.

- **Focus**: a focus on outcomes directs attention towards the patient (rather than the service) and helps nurture a 'whole system' perspective. Whilst it is thought that many individual processes are important into achieving goals, the ultimate outcome reflects all of these interactions. Thus it seems sensible to consider the ultimate outcome before attempting to understand the individual processes that make up the outcome. In other words, even if the best goal-directed therapy is used and the most up-to-date processes are employed if the patient does not survive this encounter or the resulting functional outcome is unacceptable to the patient this is of importance and must be recognised. This is the domain of outcome measures.

- **Goals**: health outcome measures more often clearly represent the goals of care and the NHS. Although the phrase ‘targets’ is becoming synonymous with negative connotations, many healthcare systems are target or goal driven. This gives an identifiable number/value that people can work towards. For example in the context of surgery the so called “Two week wait rule” attempted to improve the diagnosis of patients with suspected cancer by expediting diagnostics and primary referral. Whilst units are assessed on how well they achieve these targets, this outcome measure has not necessarily reflected in better identification of patients with suspected colorectal cancer (Thorne et al., 2006).

- **Meaningful**: outcome measures tend to be more meaningful to potential users of indicators and may be more easily understood as measures of patient safety than other perhaps more specific methods of measuring safety. Once justified, end targets may be easier to understand than other measures of quality. For example, a unit with low post-operative 30 day mortality is likely to be viewed with higher regard in terms of quality than one with all imaging modalities such as PET/CT scanners on site (structural measures).

- **Innovation**: a focus on outcomes means providers are encouraged to experiment with new modes of delivery to improve patient care, safety and experience. As with process measures, novel ways of managing patients, or patient flow may impact on outcome. For example, recently outcome benefits in terms of length of stay and re-admissions were given as the benefits of introducing so called enhanced recovery pathways in colorectal surgery. The panel of different interventions in some centres has been shown to improve some outcome end points, such as length of stay (Khoo et al., 2007).

- **Far sighted**: focusing on outcomes allows for providers to adopt a long-term strategy, such as health promotion, which may realise longer term benefits and improve safety more generally. For example, in cancer surgery long-term outcome such as functional outcome and cancer recurrence rates may drive up quality to a greater extent than focusing on short term goals, such as volume of cases (Almoudaris et al., 2011c). Short term goals tend to be driven by short term cost efficiencies. Concentrating on longer term outcome may balance this focus.

- **Manipulation**: outcomes are less open to manipulation or 'gaming' than process measures / indicators. However outcomes can be influenced if risk-adjustment models are exaggerated/ upstaged. Some outcomes such as crude mortality are not open to gaming. However, when adjustment is made for confounding variables such as palliative patients and patients with many co-morbidities, these figures may be open to potential gaming (Klugman et al., 2010).

- **Measurement definition**: while some aspects of measuring outcomes are relatively easy to measure and unambiguous (e.g deaths) others are not (e.g wound infections). This is true not only between units but also within units, where one clinician may describe a wound infection as anything that prolonged hospitalisation or required further therapy such as antibiotics or opening up of the wound, versus another clinician who may only describe wound infections as those requiring operative intervention. Although extreme examples are given here, this emphasises the potential problem of definition. However, it is acknowledged that standardised systems, such as the Clavien classification, exist for such scenarios- yet these are not routinely implemented in clinical practice (Clavien et al., 1992).

- **Attribution**: outcomes may be influenced by many factors that are outside the control of the health care organisation (e.g length of stay and availability of nursing homes/community care placement). Almost all outcome measures are influenced by other factors. Even on-table deaths in theatre may not be attributable to surgical misadventure. This may be related to unexpected anaesthetic complications or unpredicted patient factors, such as underlying cardiac pathology unrecognised at the time of surgery. Whilst this metric may be reflective of the wider unit and patient selection, it is not necessarily (as one may first assume) directly attributable to surgeon error.

- **Sample size**: outcome assessment requires large sample sizes to detect a statistically significant effect. This is especially true for the assessment of complications. By their definition complications and thus their rates occur less often than routine events and are undesirable. The difficulty here is that some complications (especially the most undesirable) or so-called ‘never events’ may require hundreds if not thousands of data points till they are observed/occur. For these to gain statistical significance large samples need to be analysed. Appreciating this fact is why they are termed ‘never events’ and even one occurrence is viewed as a system failure, hence overcoming the effect of requiring a large sample size to detect. This is considered in greater detail in the subsequent chapter.

- **Timing**: outcome may take a long period of time to observe. Most outcome measures are usually collectable after a patient care episode has finished and thus can be thought of as lagging indicators. In other words, outcome measures are reflective indicators rather than reactive indicators. They reflect what *has* happened. This means that this information can only be used to improve care for the next cohort of patients rather than for the assessed cohort, thereby explaining the term 'lagging indicators'.

- **Interpretation**: observed measures of safety outcomes may be difficult to interpret if the processes that produced the outcome are complex or occurred distant to the observed outcome. There may be many interactions that lead to the observed outcome that may not be easily discernible by measuring just an endpoint/outcome. For example, in the previously used example of fast track surgery or enhanced recovery, the observed reductions in length of stay are difficult to interpret when so many processes are introduced that may synergistically combine to infer benefit. However, without further probing it may be that only one or two processes are salient in achieving the observed outcome. In other words, it may be difficult to assess what factors in particular translate into the desired outcome.

- **Ambiguity**: good outcomes can often be achieved despite poor processes of care and levels of safety. This follows Jim Reasons model (Reason, 2000) that failures in care or safety are usually the product of many failings that align to result in a poor outcome. The opposite is true that many failures in safety go unnoticed as they do not align to produce a poor outcome.

Outcome measures are useful but can be more complex than initially thought. Measuring mortality as an outcome measure for elective groin hernia repairs is on the whole a meaningless exercise. This exceptionally infrequent occurrence would be a meaningless outcome measure in this cohort of patients as almost all surgeons would have a zero, if not near zero, mortality rate, meaning that comparisons would be non-discriminatory. Outcome measures must also take into account the complexity of individual cases and the so called case-mix adjustment. Without allowing for concurrent morbidities and the ages of patients and the other surrounding factors many outcome measures are of limited value, requiring complex statistical manipulation to case-mix adjust.

However the benefit of considering outcomes is they can be specialty specific and informative. For example, a unit's cancer recurrence rates at 1,3 and 5 years are a very important metric for appraisal. Increasingly functional outcome measures, such as the UK NHS's Patient Reported Outcome Measures (PROMS), are being used. These measures attempt to consider the surgical interventions actual benefit to the patient. Whilst these are still largely confined to a few clinical scenarios (joint replacement, groin hernias), patient functional outcome measures are also hugely important in appraising a service's level of quality.

These will be further considered in general in chapter 3.0 Contemporary methods of appraising national surgical performance' and with particular emphasis on colorectal surgery in subsequent chapters.

2.6 Intervening Variables

Whilst the structure-process-outcome framework for assessing a healthcare system has many conceptual benefits, one aspect it does not fully account for are the more subtle factors that may be influential in surgical quality. These factors include how teams function, institutional culture, morale and leadership which may be important determinant of surgical quality. The augmentation of Donabedian's model was undertaken by Lilford in 2004 (Figure 2) to incorporate these other aspects(Lilford et al., 2004). This has been refined more recently to demonstrate a more stepwise schematic by Vincent, 2011 (Figure 3). To date however there is still little direct evidence that improvements in these domains directly and reproducibly influences outcome.

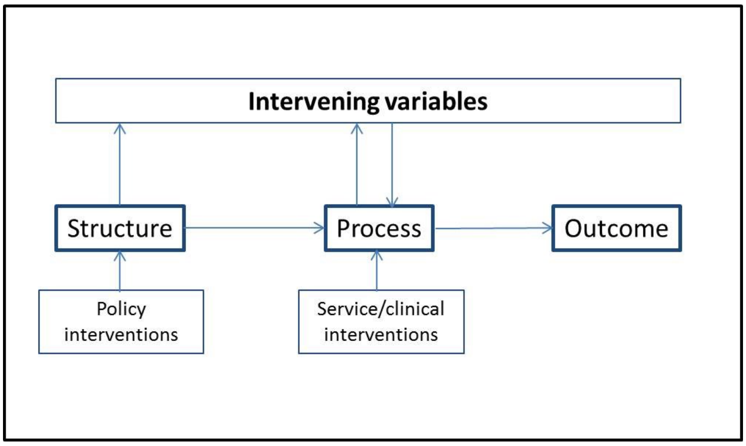


Figure 2- Schematic of Lilfords Intervening variables and their relationship to other variables. Adapted from 'Patient Safety' 2nd Edition Wiley, Vincent, 2011

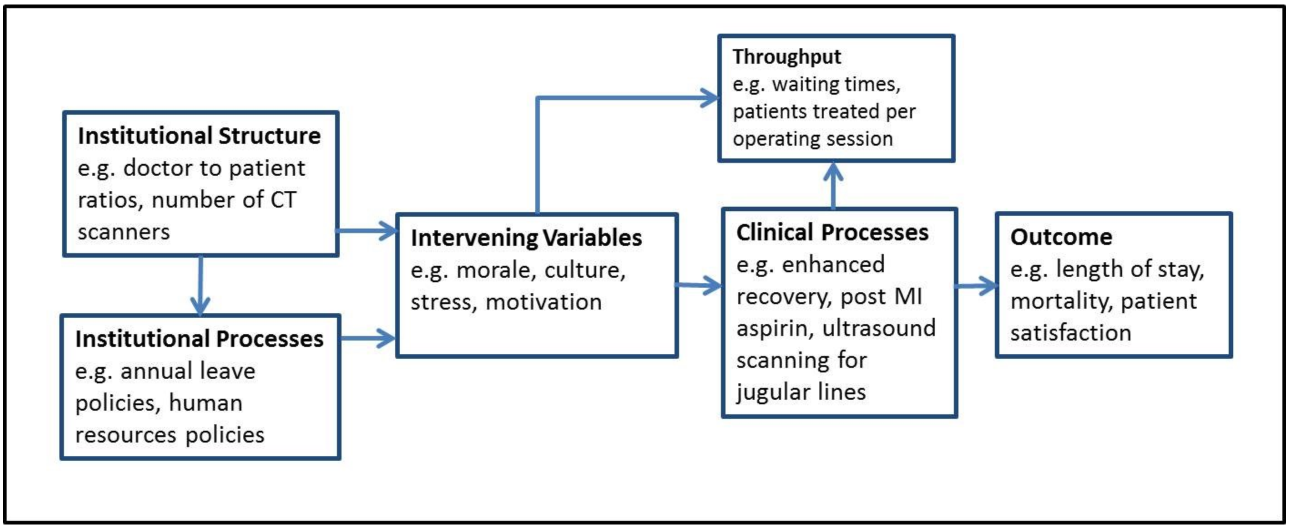


Figure 3- Adaptation of Lilford et al 2004 schematic adapted from 'Patient Safety' 2nd Edition Wiley, Vincent, 2011

There are many demonstrably important factors that Lilford has identified that have direct relevance to surgery. The table below describes some of these factors where evidence exists.

Table 3 Expansion of intervening variables

Intervening Variable	Component of interest	Example
Team Work	*Communication hierarchy* (Sutcliffe et al., 2004)	There should be the ability for members of the team to cross professional boundaries and be free to contact other members without fear of hierarchical boundaries- so called 'speaking up'. This is also extremely important in considering complication identification and management. Aberration in vital signs are usually the forewarning that a complication is developing. Nurses are usually the first recipients of this information and must feel at will to pass this information on without fear.
	Availability	For a high quality service all members of the team (nursing staff, junior doctors and allied healthcare professionals) should be contactable. Predefined criteria should exist with respect to escalation and whom to contact if initial contacts are unavailable.
	Information exchange (Sanfey et al., 2011)	Handover (hand-off in the USA) is a recognized area where information degradation occurs and mistakes occur. Handovers are increasing due to shift pattern styles of work and thus there is greater likelihood of information degradation.
Leadership	*Emotional competence/ Awareness* (Patel et al., 2010)	By understanding one's own and others' circumstances and motives, consultants can effectively manage their team and patients. Such understanding may avoid conflict and create a more harmonious and effective work environment. This extends to

		relationships with management as well as within the clinical setting.
	Senior review	In surgery little can make up for an experienced observation or assessment of a situation/patient. As a result demonstration by senior clinicians of this fosters a high quality service.
Culture	*Performance feedback* (Lytle et al., 2007)	A supportive environment where performance in analytically considered and appraised is vital to create a quality service. This extends beyond clinicians analyzing their results to wider engagement by management level staff in continual performance appraisal, feedback and constructive review.

2.7 Conclusions and implications

This chapter has reviewed the parameters of assessing surgical quality. It can be seen that ideally, it would be possible to assess a number of different aspects of a surgical service. Structural measures might include the nurse to patient ratios and the availability of specialist teams. Process measures may include the use of laparoscopy or enhanced recovery pathways. Outcome variables of interest may include both peri-operative measures of complications and long-term cancer survival. Finally, intervening variables might consider the departmental culture and organisational management styles and their influence of the clinical teams. The next chapter considers how this might be done in practice.

3.0 CONTEMPORARY METHODS OF APPRAISING NATIONAL SURGICAL PERFORMANCE

3.1 Chapter overview

This chapter aims to explore how some of the different measures in current use for national surgical quality appraisal are reported and collated. This is with a view to understanding what method(s) would be most suitable for identifying high performing surgical and colorectal units on a national basis. Clinical registries are firstly considered including understanding their inception from the days of Codman. Such registries are also analysed for their contemporary use and how they have developed. Then the Hospital Episodes Statistics (HES) database is introduced to consider which metrics are derivable. This is a nationwide national data repository with administrative origins but this may have potential to identify performance, given the richness of the data held. Finally, an introduction to Patient Reported Outcome Measures (PROMS) is undertaken and how their use has entered clinical practice. Their potential uses as a performance tool are also considered. This chapter concludes with consideration of what method, measure, or combination of measures may be most suitable for the identification of the highest performing units from the measures already available.

3.2 Clinical registries

Dr Ernest Amory Codman MD, 1869-1940, is regarded as the modern father of healthcare quality assessment in the United States. He was the Chief of Surgical Services who, whilst at Massachusetts General Hospital (MGH) in the early part of the 1900's, was obsessed with the concept of the "End Result". Codman believed that by following up all patients who were operated upon a year later he would ultimately be better informed as to their outcome. This unpopular concept (at the time) led to his resignation in 1911 and to open his own hospital- "The End Result Hospital". The final straw was when Codman displayed a picture of an ostrich laying golden eggs with its head in the sand at a standing room only medical seminar at MGH. The ostrich was a metaphor for the hospital; the golden eggs were the vast wealth accrued by the MGH staff.

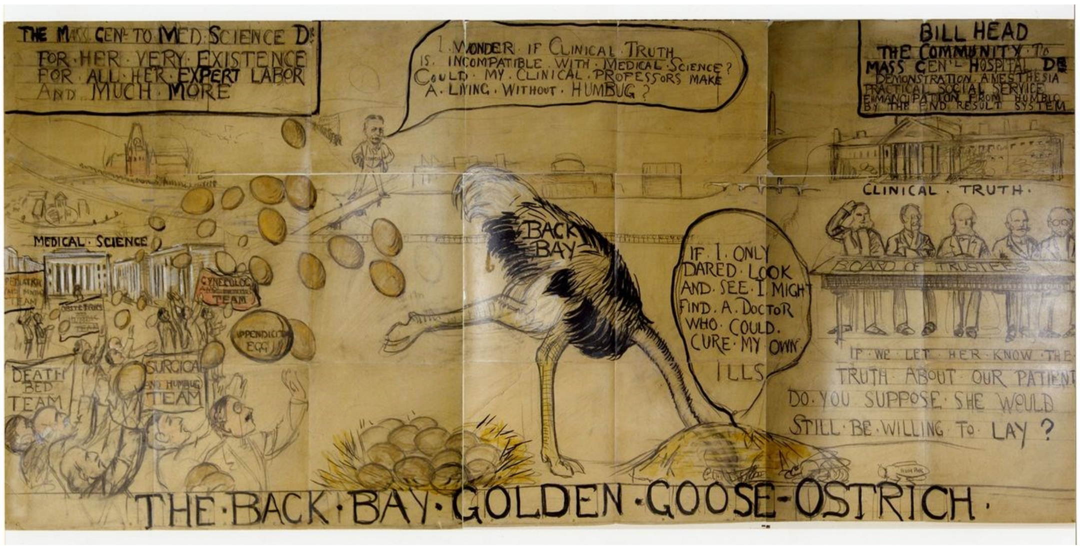


Figure 4 Digitization of Codman's original image- The Golden Goose-Ostrich with its head beneath the sand, taken from Internet Archive (https://archive.org) in accordance with Terms of Use, Privacy Policy, and Copyright Policy 31 December 2014.

Codman asked the question as to whether the goose-ostrich would still be laying golden eggs if patients knew the true outcomes from the hospital. His intention was that every patient operated upon should be followed-up. This was for the benefit of patient and surgeon alike. For the surgeon this process offers an opportunity to understand the out of hospital functional outcome and how the patient has progressed or otherwise as a result of the medical intervention. Furthermore, follow-up would allow the surgeon to assess any post-operative complications that may have occurred out of hospital that the surgeon may not otherwise be informed about. For the patient, Codman-style follow up allows reporting of functional outcomes and of any non-expected recovery events. Codman realized the importance of both interrogating outcomes but also of learning from them. Furthermore, Codman believed that patients' presenting complaints, procedures and outcome should be publicly reported annually.

There was general and widespread dissatisfaction amongst the medical establishment. Surgeons were unhappy with the implied potential criticism. In support of his theory, Codman created the first clinical registry.

In 1920 he attempted to catalogue and track nationally all bone sarcoma patients with the intention of improving care. He noted that despite all advances during his lifetime, including the discovery of X-radiation (X-rays) and subsequent radiographs, amputations were still the favoured cure for most bone tumours. Furthermore, anecdotally at least, patients seemed to have differing outcomes after these procedures in different centers. By cataloging the variability in outcomes of all sarcoma patients he intended to embarrass the establishment into reconsidering its record and recall practices. In six years he had only catalogued 17 patients onto his registry from an organization of over 7,000 members. The essence of his beliefs are reflected well in this quotation-

‘‘*The common sense notion that every hospital should follow every patient it treats, long enough to determine whether or not the treatment has been successful, and then to inquire, ‘If not, why not?’ with a view to preventing similar failures in the future’*

Codman c.1910

Codman realized that only by analyzing one’s results would one be able to reflect on the appropriateness and success of any intervention. In a primitive form, Codman touches upon setting quality standards and self-audit and evidence based practice as we know them today. These processes are undoubtedly crucial in the process of defining quality. What he did attempt to do was create the first recorded surgical registry.

Presently, with respect to surgical pathologies there are five national cancer-related registries and one benign-related registries registered with the Information Commission (IC) and the National Clinical Audit Support Programme (NCASP) [http://www.ic.nhs.uk/services/national-clinical-audit-support-programme-ncasp]. Two other notable registries exist in the form of the Vascular and Cardiac registries that are independently resourced.

Table 4- Table of National audits and registries of surgical outcome

Pathology	Name
Bowel Cancer	The Bowel Cancer Audit
Head and Neck Cancer	The National Head & Neck Cancer Audit
Breast Cancer	National Mastectomy and Breast Reconstruction Audit
Oesophago-Gastric Cancer	National Oesophago-Gastric Cancer Audit
Vascular	National Vascular Database
Adult Cardiac Surgery	The Adult Cardiac Surgery Audit

The value of self-reporting or so-called voluntary reporting systems has been previously questioned. In one study of a national clinical registry, after retrospective validation checks were made, up to 31% of cases were not reported to the registry over a 2 year period by one unit. Within these cases, 22% had some form of complication or adverse event (Dreisler et al., 2001). The submitting format is arguably the most important factor when considering using registries for performance assessment. Without 100% (or mandatory) submission, there is the real risk that cases are not recorded or submitted. This not only skews the unit's apparent performance but also that of the entire cohort.

The benefit of clinical audits is the level of detail that is contained within the audits. The audits are designed to record very detailed patient-specific information usually including pre and post-operative outcome as well as the pre-operative and in some cases intraoperative processes undertaken. Such information is vital for the true understanding of the pathologies and how they are treated on a national level. However with these benefits comes the limitation that they are labour intensive to complete and, on the face of it, do not necessarily confer any direct benefits for the individual submitter.

Most recently the government in England invited all vascular surgeons to be part of a national drive openly to report their outcome data on a basket of high risk procedures, including open and endovascular aortic aneurysm repair and carotid surgery. There was a 99% response rate and the figures were publically and openly reported in July 2013 (http://www.vsqip.org.uk/surgeon-level-public-reporting) in the Vascular Services Quality Improvement Programme (VSQIP).

Some such as the National Mastectomy and Breast Reconstruction Audit have heavy emphasis on PROMS, as patients undergoing elective breast surgery tend to be younger, fitter and mostly undergo planned procedures, making this form of assessment more meaningful. Use of PROMS shall be considered and use for performance measurement after assessment of the HES dataset will be considered.

3.3 Hospital Episodes Statistics (HES) – the modern era

The HES database is a nationally collated data warehouse containing details of admissions of NHS patients in England since 1986. The data are taken from each hospital's Patient Administration System (PAS) for every inpatient and, since 2003–2004, every outpatient Finished Consultant Episode (FCE). Demographic and any procedural/diagnostic information within that admission period is captured against a unique patient identifier. The patient identifier allows identification of previous or subsequent admissions, or procedural data pertaining to that patient. Each record (at the time of writing) also contains up to 20 International Classification of Diseases, tenth revision (ICD-10), secondary diagnoses and up to 24 procedural interventions recorded using the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures 4th Revision OPCS-4 codes (Faiz et al., 2008a).

With regards to in-patient data, most simply HES data is accessible in three formats. Firstly it is available nationally aggregated at primary or main procedure level annually (HESonline, 2010). This groups procedures such as right-hemicolectomies and reports the number performed nationally. The second format is the reporting of all Finished Consultant Episodes (FCEs) - which is defined as the total care a patient receives under a named consultant, per provider. This is reported by generic specialty grouping. In this format a group of ICD-10 codes are aggregated (e.g code C00-C048 for all neoplasms). Again these reports are annually produced. So for example it would be possible to find out how many procedures were performed by each Trust. These first two formats are freely available on-line (www.hesonline.nhs.uk). In this format researchers are limited to looking at trends in national volume of procedures with no further information being extractable. The final format available is anonymised patient level patient data reported per FCE per provider. This format allows for more complex analyses however separate application and approvals are required to access this data. Such application will also report the full patient dataset that includes all diagnostic and procedural fields, date of admission and discharge, as well as age at admission. Directly reported or derivable from HES, are outcome measures that are useful in appraising surgical performance. What follows are details of the outcome measures available and critique of them for use as performance indicators that are applicable to the appraisal of surgical performance.

In particular the native HES dataset can be interrogated and analysed to produce metrics and information that are of use specifically for the appraisal of surgical performance. These metrics can be calculated from the administrative dataset.

3.3.1 HES derivable outcome measure -Length of stay

This is usually calculated as the complete number of days from admission to discharge. Its attractiveness as a measure lies in that it is readily calculated and not open to gaming. It is easily derived from administrative datasets and commonly quoted as the average LOS. When considering an average length of stay, the mean maybe skewed by outliers, thus may misrepresent the population as a whole. Furthermore, the mean would likely misrepresent any ‘improvements’ over time that may occur in a unit, as historical outliers will still affect more contemporary data (Fisher and Altaffer, 1992). The median however, reflecting the central tendency of LOS, is not as influenced by outliers. The median informs us that half the population, have a stay in hospital below a certain figure. It however gives us little information about the distribution of the rest of the population. To some extent this can be inferred from the population range. One statistical approach is to define an acceptable upper limit of LOS. In the methodology the 75th percentile population LOS for this limit was chosen. This approach identifies all patients that have a LOS above the 75th percentile of the whole population as outliers, and this methodology has been described previously (Cohen et al., 2009a). Whilst not a perfect approach, this method acknowledges that the population mean is less likely to be influenced by outliers than an individual unit mean, and the 75th percentile as a cut off reduces the probability of a LOS occurring above this by chance even further.

3.3.1.1 Influences on Length of stay

It must be noted however that LOS measurements may be influenced by non-clinical factors. Irrespective of the clinical course and subsequent outcome, a patient may have their discharge delayed for a variety of reasons. Within hospital this may be due to delays in organizing take home medications. Although seemingly trivial, a delay of one overnight stay may be proportionately relatively significant if the 'clinical' stay is under five days. One extra day would represent a 20% increase in the LOS. Similarly the influence of the day of the week a patient is admitted has been shown to have an influence on LOS. In patients whose planned discharge falls on a weekend, it has been shown they are more likely to be kept in until the following Monday, thus also influencing the LOS (Varnava et al., 2002). This may have implications for surgeons who have mid-week or end of week operating lists.

Outside hospital factors are also important with the ultimate discharge destination having a significant impact upon LOS. It has been shown that patients were significantly more likely to have an increased LOS if they required discharge to nursing homes or other institutions (Brasel et al., 2007). These findings are also reflected in a National Audit from Scotland published in 2010 which since 2000 has coded for delayed discharges in their administrative datasets (National Statistics Publication, 2010). Furthermore, a recent Cochrane review has concluded that the benefits of a structured discharge plan can be reflected in reduced length of stay (Shepperd et al., 2010). In colorectal surgery the adoption of enhanced recovery programs is likely to have had influence in reducing LOS (King et al., 2006a). This may be reflected in the coding as operations being performed laparoscopically. However, one key concern is that not all operations that are commenced in the laparoscopic fashion are completed laparoscopically. It is not discernible from administrative datasets which operations are completed in this fashion, thus caution is warranted when attempting to afford reductions in LOS due solely to laparoscopy, from these datasets.

In the future, novel ways of assessing LOS may be used. These may include longitudinal analysis of subsequent admissions and aggregated LOS. Index LOS may be too simplistic a measure if patients are subsequently re-admitted for prolonged periods. A native index LOS measure would not truly reflect the care received. Alternatives may be the calculation of total length of stay if subsequent admissions can accurately be assigned to the index procedure (Mamidanna et al., 2010). This methodology may be a more reflective assessment of LOS if non-clinical factors can be taken into account as well.

3.3.2 HES derivable outcome measure -In-hospital mortality

Most simply, this is calculated as the number of deaths that occur in-hospital - otherwise known as the crude mortality. This can be presented as a rate usually expressed as the number of deaths per 1000 hospital discharges. The strength of measuring in-hospital mortality as a crude rate is that it is a binomial outcome that is well-defined and not subject to variability of interpretation. Deaths within hospital, especially after surgery are affected by 'non-clinical' factors to a very small degree, as compared with length of stay. Rates are commonly cited in the literature and thus should make for easy comparison of units. However the problems encountered with reporting of crude rates is that the crude in-hospital mortality rate makes no attempt to take into account case-mix (Elsevier, 2010). Such case-mix differences may potentially underlie to some degree the reported disparity and variation in outcome between units and rates (Rigby et al., 2001, Brunelli et al., 2006, Mohil et al., 2008). There is evidence that suggests hospitals undertaking major surgical procedures can fall within control limits after case-mix adjustment as compared to before adjustment (Ansari et al., 1999). In certain populations, however, risk-adjustment has been shown to have little effect (Dimick and Birkmeyer, 2008). This said, it would be fair to assume a higher death rate in a tertiary center operating on the most complex patients as well as emergencies, as compared to a small district unit only undertaking elective 'routine' resections.

3.3.2.1 Accounting for case-mix and mortality

A development from the crude mortality rate was the concept of standardization (Armstrong, 1995) which can either be direct or indirect dependent upon the origin of the reference used to weight the strata-specific rates. Standardized mortality ratios (SMR) are constructed by taking the observed/expected counts and multiplying by 100. This method can be useful when considering factors such as age, sex and race, however, the process becomes difficult when attempting to standardize for multiple factors. Its use is only stable when studying a population with small numbers of deaths. Critics have questioned the reliability of such a measure, citing Standardized Rate Ratios (SRR) which are the ratio of the age-adjusted mortality rate for a study population to the age-adjusted mortality rate for a reference population, as being preferable. This view is not supported when applied analytically to a population dataset (Goldman and Brender, 2000). Such methods, however, whilst they make some appreciation of case-mix, do not account for more complex variation, such as types of operation or presenting stage of disease.

More complex methods of accounting for case-mix differences have been devised using advanced statistical methods. In using logistic regression, taking mortality as the dependent, the development of a linear equation for the log of a positive outcome is possible. The co-variates influence the odds of the dependent occurring and these can be applied to create a predicted probability of the dependent occurring in an individual patient which can be aggregated per unit. Such risk adjustment has become the method of choice over others for large scale quality improvement initiatives (Cohen et al., 2009b). How and which variables to include in risk-adjustment models has been the subject of significant debate (Dimick et al., 2010a). However, what appears to be true is that by including clinical measures, such as individual laboratory results, these have been shown to enhance the predictive power of the models (Escobar et al., 2008). It is clear however that different methodologies of risk assessment can yield both differing (Steinberg et al., 2008, Fedeli et al., 2007, Atherly et al., 2004, Shahian et al., 2010) and similar results even when analyzing the same dataset (Aylin et al., 2007a). Thus the outcome of risk-adjustment is dependent upon the method and co-variates used.

Finally, when considering case-mix, intention of treatment is commonly overlooked (Klugman et al., 2010). This is largely because such data are not readily available from administrative datasets. When treatment intention is considered in surgical populations, it has been shown that mortality rates are almost halved, when palliative and planned end of life procedures are discounted from analyses (Gillion, 2005). There has however been criticism of the rise in the number of patients being coded with 'palliative' codes (Hawkes, 2010). Overzealous inclusion of such codes may influence overall hospital mortality statistics and thus ironically call to question the robustness of what 'adjusted' mortality statistics tell us.

3.3.2.2 When should mortality be measured?

Although in-hospital mortality is the most commonly cited mortality measure in the literature, some have questioned its relevance in certain populations and for certain procedures (Jarman et al., 2002, Seagroatt and Goldacre, 1994). Although in-hospital mortality may be a more sensitive measure of the clinical course, it may not reflect the true outcome of operating on groups such as the elderly. In the literature, mortality rates have been quoted at different times post-operatively (30days, 90days, 1 year); each is likely to be sensitive to different factors (Mamidanna et al., 2012). Mortality after 30 days of an operation, be it in or out of hospital gives a fair indication of outcome from any post-operative complications. Thirty day mortality is considered by some to be a more accurate measure of hospital performance as it is less dependent on hospital discharge policies (Borzecki et al., 2010). Consider a hypothetical scenario of a unit unsafely discharging patients very soon after their procedures. Such a unit would have very low in-hospital mortality rates; however 30-day mortality would be a more reflective indicator of the unit's true performance if these patients then go on to die. Furthermore, in certain conditions and populations, such as cancer resections in the elderly, a longer term measure of mortality may be more appropriate. It has been shown that in elderly patients undergoing resections for colorectal cancer 3.7%-12.9% (depending on age band analyzed) of patients are dead after 30-days of surgery compared to 14.1%-36.1% at one year (Faiz et al., 2010a).

In summary, mortality measures are numerous and complex. Choosing the correct measure is likely to influence the interpreted outcome from such use. How severity and co-morbidities are considered is also a pivotal factor in the meaning of such measures. The Department of Health in late 2010 after consultation is introducing a universal measure to be used for comparison of all acute NHS Trust hospitals. The Summary Hospital-level Mortality Indicator (SHMI) will be reported in context of a units workload and is based around existing Hospital Standardised Mortality Ratio (Deaprtment of Health, 2010). It remains to be seen whether the technical changes made to this measure are more widely accepted.

It must be noted that non-cause specific in-hospital death is recorded by HES. For deaths occurring after discharge, linkage to the Office of National Statistics mortality dataset is required.

3.3.3 HES derivable outcome measure - Return to theatre rates

Return to theatre rates have been suggested to be good surrogates for serious surgical complications (Ansari and Collopy, 1996, Birkmeyer et al., 2001a, Morris et al., 2007a). The crux of defining the relevance of such a measure is whether returns to theatre are planned or unplanned. From administrative datasets this is usually difficult to discern as coding fields for returns to theatre generally do not exist. However, to overcome this challenge, by selectively choosing specific codes for certain re-operations (for example post-operative bleeding, post-operative intra-abdominal abscess) this problem can be mitigated when compared to the dates of any subsequent operative procedures within the same admission. Such an approach has been used in one English series derived from HES that demonstrated significant variation in re-operation rates in colorectal surgery (Burns et al., 2011b). It has been shown that in patients that require returns to theatre, they not only have longer lengths of stay, but also are more likely to be dead at 30 days (Morris et al., 2007a). Some have described the factors underlying such a metric, specifically that in up to 70% of returned cases surgeon error (Kroon et al., 2007), or technical factors are causal (Birkmeyer et al., 2001a) whereas others have cited specific patient factors (Ploeg et al., 2008) such as concurrent morbidities.

It is likely to be true that both patient factors and surgeon factors are responsible for a patient requiring an emergent return to theatre. Furthermore an unexpected return to theatre is not the same as a planned return to theatre, although not discernible from coding. At present, only by record review, can surgeons' intention be considered. In one series it was found that in 50% of cases, surgeons had planned to have laparoscopic re-looks on the same admission due to the complexity of the surgery. This, though fully planned, would have been coded as an unplanned return to theatre (Birkmeyer et al., 2001a).

The future of using unplanned returns to theatre as a metric for performance and quality control will be in ensuring it is being faithfully derived from whatever dataset it is calculated from. Crucially, surgeons' intention and planning will need to be taken into account fully to integrate this metric into acceptable widespread use.

3.3.4 Emergency Re-admission within 28 days of discharge rates

Re-admission to hospital emergently within 28 days of discharge is considered to be undesirable for patients and clinicians alike. This metric is derivable from HES by using linkage with subsequent admissions. Simplistically this may represent a poorly planned initial discharge or a missed complication or evolution thereof. It has been estimated that one third occur within one month, and half within 90 days (Zook and Moore, 1980, Corrigan and Kazandjian, 1991). Most preventable re-admissions have been observed to occur within the first 28 days of discharge, validating this time period for common use (Sibbritt, 1995). Using re-admissions further in time post discharge, may, in fact reflect disease progression and a continuum of the original disease, but not necessarily a poor discharge following the index admission. It must be noted that not all re-admissions are attributable to the 'index' admission within 28days (Courtney et al., 2003). For example, a patient discharged after an uneventful routine laparoscopic cholecystectomy, who is subsequently re-admitted a week later with a fractured neck of femur, would on administrative datasets be classed as an emergency re-admission. This is clearly correct but potentially misinforming, if used to appraise the quality of the original surgery. Such scenarios are reflective of how datasets should be carefully interrogated and understood before deriving assumptions. This being said, the 28 day re-admission rate is a useful reflection of preventable returns to hospital with rates from 12-75% quoted as being preventable in mixed populations of patients (Benbassat and Taragin, 2000). Re-admissions to hospital within 28 days of discharge are clearly multi-factorial. In a study of 186,000 colorectal patients, independent predictors of 28-day readmission included: distal bowel resection, benign diagnosis, young age, worse social deprivation and high provider unit volume status (Faiz et al., 2010b). This mixed panel of predictors in this population demonstrates how the 28 day re-admission rate is complex to understand. It appears that whether the admission was related to the index procedure and whether it was preventable are key considerations for this marker. It is thus clear to see that with further refinement and correct patient level data linkage re-admission data can be a strong marker of a unit's performance.

Table 5- Table summarising the strengths and potential weaknesses of different measures of outcome

Measure	Strengths	Weakness
Length of stay	Easily calculated. Commonly cited.	Does not account for any subsequent re-admission stay. Influenced by external factors (social care setup at home, discharge destination).

In-hospital mortality	Easily calculated. Commonly cited. Less influenced by 'non-clinical' factors.	Crude values may not reflect case-mix. Varied Risk-adjustment methods may lead to different interpretations. May occur infrequently for certain conditions/operations to be discriminatory. Will miss deaths occurring shortly after discharge.
Re-admission rate	Easily calculated. Commonly cited.	Complex to calculate re-admission cause from administrative datasets. Preventability difficult to derive.
Return to theatre rate	Reflects in-hospital care. Not affected by non-clinical factors.	Complex to calculate. More dependent upon accurate coding. May reflect intra-operative processes more closely.

3.3.5 Use of HES for surgical performance

Undoubtedly the information contained within the dataset is of importance in appraising surgical performance. Limited advances would be made with the freely available information due to the aggregation and the inability to perform meaningful risk-adjustment. For these reasons, to pursue use of this dataset, patient level information would be required. There have been criticism over the use of the dataset for outcome analysis (Williams and Mann, 2002). These largely relate to potential inaccuracies of the data due to the fact clinical information is translated by non-clinical coders into hospitals PAS systems. Within this translational step, errors may be made. Recent attempts to quantify the accuracy of the HES dataset have shown coding of primary diagnoses with an accuracy of 96.0% (Burns et al., 2012). In other words, when analyzing the primary reason the patient is admitted to a unit, there is excellent accuracy of HES with retrospective reviews. Furthermore HES is being used by clinical registries to cross validate their submissions, reflecting the greater accuracy of the dataset (www.ic.nhs.uk/bowel). A recent systematic review has identified HES being increasingly used for health-care outcome assessment. In 1994, 2 papers were identified by the review as compared to 26 in 2010 (Williams and Mann, 2002).

3.4 Patient Reported Outcome Measures (PROMS)

PROMS were developed by the NHS to assess the effectiveness of care delivered from the patient perspective. After piloting, national rollout occurred in 2009 for four elective surgical procedures: groin hernia surgery, hip replacement, knee replacement and varicose vein surgery. The methodology captures patients' health status before and after the operative interventions (London School of Hygeine and Tropical Medicine, 2005). The aim is scientifically to assess the interventional impact from the patients' perspective. The recorded information briefly comprises four elements before and after the procedure:

1. A generic measure of health status,
2. Condition specific measures that are designed to be sensitive to change in health status for that condition,
3. An assessment of the patients living arrangements to adjust and contextualize results,
4. An assessment of whether the patient feels he or she has any disability and whether any assistance was sought to complete the questionnaires.

Procedure	Condition-specific	Generic
Unilateral Hip Replacement	Oxford Hip Score	EQ-5D*
Unilateral Knee Replacement	Oxford Knee Score	EQ-5D*
Groin Hernia Surgery	None	EQ-5D*
Varicose Vein Surgery	Aberdeen Varicose Vein Questionnaire	EQ-5D*
*: EQ-5D™ is a trademark of the EuroQol Group (www.euroqol.org)		

Figure 5- taken from NHS guidance on the routine collection of Patient Reported Outcome Measures (PROMS). For the NHS in England 2009/10

Other PROMS measurements are undertaken as in the case of the National Mastectomy and Breast Reconstruction Audit. However, these are independently managed by the respective specialist societies. From the most recently available finalised national PROMS data (April 2010 to March 2011), pre-operative uptake stood at 69.9% with subsequent post-operative uptake at 81.0% (Comission., 2012). Uptake rates are considered as a percentage of eligible procedures, as recorded from the corresponding time periods on HES.

Participation in PROMS is voluntary for the patient. As a result uptake rates are very variable by procedure. The graph below demonstrates pre-operative participation rates with monthly linkage to HES by procedure.

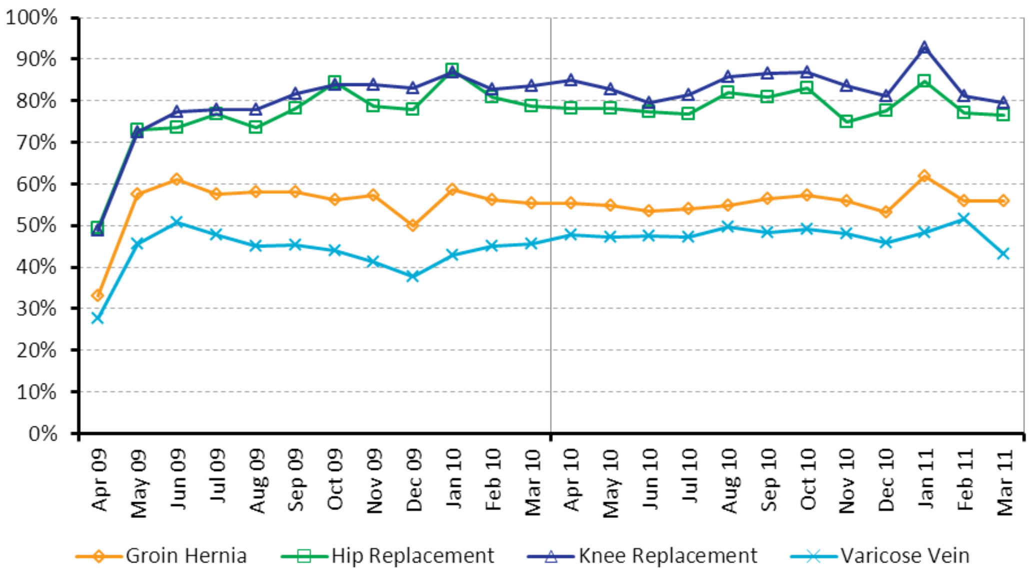


Figure 6- Graph demonstrating patient participation rates in PROMS taken from NHS guidance on the routine collection of Patient Reported Outcome Measures (PROMS). For the NHS in England 2009/10

It is clear that the more major procedures (joint replacements) have much higher consistent uptake rates. This is likely to be multi-factorial but it almost certainly reflects the fact that these patients reside in hospital for several days and thus are more likely to be captured. Groin and vein surgery are mostly done as day cases and thus patients spend less time in the department to be captured and/or have the process of completing the PROMs questionnaire explained to them.

As with clinical audits, without complete capture of all events it would be difficult to infer with reasonable rigor national differences. However on an individual unit level, for example, trends may appear in patients' experiences between differing surgeons. For these reasons PROMS are a valuable tool but limited in use for national benchmarking with participation rates as they are. Furthermore the national program only considers elective procedures and has not been validated for use in more major surgery such as emergency abdominal surgery. One attempt to reconcile the lack of clear reportable measure was a national initiative called the Better Metrics Project.

3.5 Others methods of measuring surgical performance

In England a national 'Better Metrics Project' was launched to identify evidence-based metrics that may be used to measure and benchmark performance across a broad array of medical and surgical subspecialties (Care Quality Commission, 2009a). With regard to surgery, colorectal cancer surgery was chosen and after wide professional and expert consultation, three metrics were decided upon to be reflective measures of quality. These were: the proportion of patients operated on for rectal cancer with a permanent stoma, the surgeon-specific and institutional caseload for colorectal surgery, and participation in a national colorectal cancer audit (Care Quality Commission, 2009a). At the time of writing the project has been discontinued with no metric refreshing undertaken since 2009. Possible reasons include few to no publications demonstrating the validity of the chosen metrics in promoting quality care.

THESIS AIMS

What is currently known on national surgical performance is that many different metrics may be used, each with its own strengths but equally all have specific characteristics that pertain to different aspects of the care given. There is a need to review and appraise these metrics to identify potential metrics that can best reflect high performance in surgical practice. Colorectal cancer surgery has been chosen for the on-going analyses in this thesis.

Current national surgical performance measurement predominantly uses clinical registries. At the time of writing many were voluntarily subscribed and submitted to. The value of clinical registries, especially if they are voluntarily subscribed to, may misinform on the actual performance of a unit or even affect national outcome statistics. It is necessary to identify whether differences exist between using voluntarily and mandatorily submitted data where possible in colorectal cancer surgical performance appraisal. The peri-operative period has been identified for further analysis.

Whilst administrative datasets such as HES are mandatory in data collection, their current use is limited to the information that is currently reported by them. To maximise the utility of such datasets it is necessary to investigate whether clinically relevant information can be derived from such datasets. Specifically no appraisal of surgical performance would be complete without an assessment of the management of complications. It remains to be seen if such an important metric is derivable for the first time from nationally collected mandatory collected administrative data.

Whilst hard-outcome data and potentially derivable metrics are important in appraising surgical performance, there is an important aspect that cannot be discerned from analysis of such datasets. Organisational and teamwork factors are associated with high performance and it would be important to incorporate an assessment of these traditionally very difficult to appraise factors in any comprehensive overview of performance.

Finally it would be informative to try and relate the findings of the study of the organisational and teamwork factors to the hard outcome measures previously identified.

The aims of this Thesis are to:

1. Identify and review what metrics are important to consider in appraising performance in colorectal cancer surgery.

2. Identify whether any outcome differences exist between using current methods of performance appraisal (e.g. clinical registries) and mandatory data sources (e.g. HES) using comparable end-points and time periods.

3. Acknowledging the limitations of administrative datasets it is necessary to appraise whether it is possible to creating meaningful novel metrics, such as the outcome from the management of complications, from these datasets.

4. Determine whether it is possible and feasible to appraise the national performance of all colorectal cancer units on a number of selected important metrics.

5. Design and pilot a methodology for capturing organisational and teamwork factors associated with high performance from clinical units.

6. Finally, to examine the association between organisational and teamwork factors and surgical outcomes from these units.

4.0 ESTABLISHING QUALITY IN SURGERY- A REVIEW OF THE LITERATURE

4.1 Chapter overview

This chapter consists of a review of the literature on what may constitute quality in colorectal surgery. The purpose of this review is to offer a contemporary perspective and highlight potentially important markers of quality in colorectal surgery in current use. An introduction to the review precedes the search strategy implemented. Results are offered according to the structure, process, and outcome framework. Other measures are also considered, including patient satisfaction ratings and functional outcome assessment.

Structural and process metrics, as well as clinical and patient reported outcome, are reported with a view to support quality appraisal in colorectal surgery. A clear appreciation of the scope of individual metrics for quality appraisal purposes is demonstrated if they are to be used meaningfully for performance benchmarking. Further work is debated including the requirement to understand the role of public and internal reporting of performance measures in colorectal surgery as drivers of quality improvement.

4.2 What do we mean by quality in colorectal surgery?

Clear definition and measurement of quality within surgery are pre-requisites for improving standards. Although it is accepted that service providers should strive to improve quality, what exactly does quality mean to colorectal surgeons and their patients? Previous attempts have been made to define quality markers in colorectal cancer surgery by using qualitative research methodologies (McGory et al., 2006).

It is necessary to explore current definitions and describe the methods available for quality measurement in colorectal surgery.

4.3 Aims

The aim of this chapter is to review the current literature to understand what quality in colorectal surgery is. Surgical and patient related outcome are considered to give a balanced view.

4.4 Methods

4.4.1 Search Strategy

The search terms used included the Medical Subject Heading terms and Boolean characters: 'colon' OR 'colorectal', OR 'rectal' OR 'rectum' AND 'Quality Indicator$' OR 'Quality Assurance' OR 'Quality of healthcare' OR 'Reference Standard$' OR 'Quality' plus a variable floating term. A two person independent review was undertaken from resulting citations and their consequent reference lists. The search was limited to citations from 2000 to 2010, humans and the English language. The most recent search was performed on the 7th February 2010 Duplicates were removed at all levels.

For colorectal specific indicators the following Medical Search Headings (MeSH) and Boolean terms were employed:

1. ‘colon’ OR ‘colorectal, OR ‘rectal’ OR ‘rectum’ AND

2. 'Quality indicator$' OR 'quality of healthcare' OR 'quality assurance' OR 'reference standard$' OR 'quality' AND

3. a floating keyword(s) according to the indicator in question (see table 6 – the indicators selected for searching are described according to the structure, process, outcome framework).

<table>
<tr><th>Static term</th><th></th><th>Variable 'floating' term</th><th></th><th>Citations returned</th><th>Included in review</th></tr>
<tr><td rowspan="7">'colon' OR 'colorectal, OR 'rectal' OR 'rectum'

AND

'Quality Indicator$' OR 'Quality Assurance' OR 'Quality of healthcare' OR 'Reference Standard$' OR 'Quality'</td><td rowspan="7">AND</td><td>Operative caseload
'Surgeon' OR 'Hospital' AND 'Volume'</td><td>structure</td><td>58</td><td>16</td></tr>
<tr><td>Technical factors
'TME' OR 'total mesenteric excision'
OR
'APER' OR 'abdominoperineal resection'</td><td rowspan="2">process</td><td>236</td><td>11</td></tr>
<tr><td>Pathology measures
Lymph node$ AND 'pathology'</td><td>78</td><td>14</td></tr>
<tr><td>Intraoperative measures
'surgical blood loss'</td><td rowspan="4">outcome</td><td>2</td><td>2</td></tr>
<tr><td>Perioperative surgical morbidity measures
'anastomotic leak' OR 'anastomotic dehiscence' OR 're-operation'</td><td>330</td><td>11</td></tr>
<tr><td>Perioperative mortality measures
'operation' OR 'surgery' AND 'mortality' AND 'outcome$'
'mortality' AND 'outcome$'</td><td>97</td><td>7</td></tr>
<tr><td>Oncological Outcomes
'survival' AND 'local recurrence' OR 'loco-regional recurrence'</td><td>131</td><td>8</td></tr>
</table>

Table 6- Table showing use of static and variable floating terms used in search

4.4.2 Eligibility Criteria/ Inclusion Criteria/Floating terms/Study Selection

Inclusion Criteria

Meta-analyses, Randomised controlled trials (RCT), controlled clinical trials (CCT), cohort and case-control studies as well as review articles were considered. Studies were included if they reported on colorectal surgery with respect to improving outcome or defining standards of quality care. Studies from 2000-2010 were evaluated for a contemporary perspective in the initial search. Articles in press and published online in English were considered. Reference lists from cited articles were also examined and cross-referenced articles were included where relevant.

Exclusion Criteria

Studies were excluded if they did not answer the study question being examined in the relevant part of the review or if they did not describe or critique standards or measures that could be used to define quality, or bench-marking, in colorectal surgery. Studies were excluded if they did not pass the quality assessment. Correspondences and letters were also excluded.

Floating terms

The floating terms used in the search were chosen by the senior authors. These were chosen to complement the Donabedian search theme and were selected after consensus was reached.

Study Selection

Two reviewers independently performed the searches with concordance of all references. Selection occurred with two reviewers independently screening the titles and full abstracts for relevance and excluding those that did not meet the inclusion criteria. Discordant decisions were arbitrated by a third author. From this, full text articles were retrieved both electronically and manually. Studies were included if they fulfilled the inclusion criteria and passed the quality assessment and were relevant to the discussion point of the article as agreed by two reviewers.

4.4.3 Information sources

OVIDSP incorporating the MEDLINE, EMBASE, PsycInfo databases as well as the Cochrane database were electronically searched. Government and specialist society guidelines (e.g.,ACPGBI, ASCRS), cancer networks (e.g., ASCO, NCCN), and organisations that monitor quality of medical and surgical care (e.g., NICE, CQC) were considered to identify currently available contemporary clinical practice guidelines via printed documents and their websites.

4.4.4 Quality assessment

Quality assessment of the included studies was carried out using the Newcastle-Ottawa Scale for the non-randomised studies where appropriate e.g case-control studies, cohort studies. At least one 'star' per category on the Newcastle-Ottawa scale of the assessed nonrandomised studies was achieved per relevant study. Further details of the scale and its application can be found at http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

4.5 Results

4.5.1 Structural factors of quality in Colorectal Surgery

4.5.1.1 Operative Caseload

Of the 16 articles included that related specifically to the volume-outcome relationship in colorectal surgery, poor outcome and low surgeon, or low hospital, caseload have been demonstrated in at least five of these studies (Bentrem and Brennan, 2005, Borowski et al., 2007, Luft et al., 1979, Schrag et al., 2000, Schrag et al., 2002). Conversely, reduction in perioperative mortality and increased survival have been observed amongst adult patients (Schrag et al., 2000) (Ko et al., 2002) and elderly (Dimick et al., 2003) patients undergoing colon cancer surgery at high volume institutions. Furthermore, longer term studies have shown significant survival advantage at 5 years amongst patients undergoing colorectal cancer resection in high volume hospitals (Rogers et al., 2006). No such relationship with either survival or local recurrence rates (Engel et al., 2005) was however observed amongst 884 patients treated for rectal cancer in one study. Significant variability between high volume surgeons has also been shown when comparing anastomotic leak rates in one series of 556 patients following bowel resections (small and large bowel), with a six-fold difference in leak rate between the extremes (Hyman et al., 2009). The patients in the latter study were however neither randomised nor case-mix adjusted for.

A positive association between either institution, or surgeon, caseload and perioperative mortality or intestinal continuity rates in elective rectal cancer surgery was not identified in one large North American retrospective population-based cohort study (Schrag et al., 2002). The latter trial did however confirm a two-year survival advantage to patients operated upon by high volume surgeons. A systematic review by Salz and colleagues investigating the impact of operative volume in rectal cancer surgery observed that high volume centres and surgeons perform more sphincter-saving procedures and are associated with lower postoperative mortality. However, the complication rates, local recurrence rates and overall survival did not show benefit in higher volume centres (Salz and Sandler, 2008). In a recent meta-analysis of thirty-five studies, postoperative morbidity following treatment for colonic cancer was associated with surgeon caseload and education. Moreover, reduced postoperative mortality was also associated with high hospital, and surgeon, caseload (Iversen et al., 2007). These findings are reflected in an English population-based audit where high volume surgeons (deemed as those with operative caseloads>18.5 cases per year) and colorectal specialists were observed to be more likely to perform sphincter saving procedures for rectal cancer (Borowski et al., 2007). Moreover, a prospective multi-centre observational study of 1557 patients identified that sphincter conservation was more likely to occur in higher volume centres (Ptok et al., 2007a). Schrag and colleagues also demonstrated hospital and surgeon volumes as important predictors of stoma rates in patients undergoing a primary resection for colon cancer (Schrag et al., 2003). In another study Harling and co-workers from the Danish National Colorectal Group demonstrated, from a national database comprising over five-thousand patients, a significantly higher risk of a permanent stoma amongst patients undergoing rectal cancer procedures in low volume institutions (Harling et al., 2005). They did not however identify any differences in anastomotic leak rates, 30-day mortality or 5-year survival rates. Considerable controversy exists regarding the relationship between surgical caseload and outcome and its consequent impact on colorectal service provision worldwide. In the United Kingdom (UK) it is now recommended that a colorectal surgeon perform at least 20 colorectal cancer procedures per year (NICE, 2004).

4.5.2 Process factors of quality in Colorectal Surgery

4.5.2.1 Technical factors

Abdomino-perineal excision (APE) is a procedure that may be inherently associated with poor oncological outcomes (Marr et al., 2005) in addition to the obvious need for a permanent stoma. Process measures like APE rates and the quality of surgical excision, as evidenced by the quality of the plane of rectal dissection evaluated on histological examination of the mesorectal fascia (Quirke, 2003), may be used to benchmark aspects of surgical quality in rectal cancer treatment.

Grading of mesorectal quality into three categories according to the completeness of excision, can distinguish between the technical proficiency of surgeons. In an analysis of 100 patients undergoing total mesenteric excision (TME) surgery for rectal cancer anatomical factors were more likely to predict inadequate mesorectal quality than clinical factors (Hyuk Baik et al., 2008). Quality of mesorectal excision may potentially be of use as an assessment tool for evaluating technical proficiency in rectal cancer surgery. The relationship of this process measure to outcome is however as yet uncertain.

The drive to raise standards has involved efforts to decrease non-restorative operations in favour of sphincter-saving procedures. This has been echoed by the

‘Better Metrics Project’ in the United Kingdom (Care Quality Commission, 2009a). The reason for poorer oncological outcomes in the lower rectum may relate to the more advanced nature of rectal cancer in this location (greater proportion of T4/R1/circumferential resection margin (CRM) positive cancers with higher rates of tumour or bowel perforation) (Marr et al., 2005); (Wibe et al., 2004), and/or, to deficient surgical technique leading to ‘waisting’ of the rectal specimen rather than achieving a cylindrical excision (West et al., 2008a). It is important to note that there is not universal agreement that APE is necessarily independently associated with poorer outcomes (Chuwa and Seow-Choen, 2006). Morris and co-workers, using NHS routinely-collected data linked to the cancer registry observed wide variation in the use of APE across colorectal service providers in England (Morris et al., 2008). The Association of Coloproctology of Great Britain and Ireland (ACPGBI) have, however, recommended that the proportion of resectable rectal cancers treated by APE should be less than 30% of total rectal cancer excisions (Association of Coloproctology of Great Britain and Ireland, 2007).

Functional outcome after rectal cancer surgery is a key factor when considering the most appropriate procedure for an individual patient. Surgeon-related factors are as important in preservation of function as in long-term local oncological control (Moriya, 2006). Furthermore evidence exists that the surgical approach (i.e. laparoscopic versus open) seems to confer neither benefit, nor disadvantage, functionally when performing TME surgery (Morino et al., 2009). Poor functional outcome has been shown to be associated with preoperative factors too. In patients undergoing low anterior resection and TME, differences in faecal incontinence rates were observed between patients that underwent pre-operative radiotherapy and those that did not, when assessed five years' post treatment (Lange et al., 2007). Functional outcome following rectal cancer treatment is therefore multifactorial relating to patient, surgical and adjuvant treatment factors, and is an important marker of operative success and quality.

4.5.2.2 Pathology measures

The American Society of Clinical Oncology (ASCO) have derived a range of surgical quality process metrics that evaluate both the surgical and adjuvant management of colorectal cancer (Desch et al., 2008). These metrics include: administration of adjuvant treatments for colonic and rectal cancer and lymph node yield (excision and examination of > 12 lymph nodes in colon cancer) (Desch et al., 2008). The relationship between oncological outcomes, such as survival, and lymphadenectomy yield is however complex. In terms of patients being restaged due to lymph node yield, it has been shown that, at low lymph node yields, the likelihood of detecting a positive lymph node increases up to 5-6 nodes. Above this value, it is less likely that >7 lymph nodes will change a patient's stage (Baxter et al., 2010). Average lymph node yields in excess of 12 nodes are more likely amongst surgeons with higher caseload and those who are sub-specialised (Dillman et al., 2009). Poor 5-year survival rates correlate with both adverse tumour characteristics as well as with low lymphadenectomy yields for Stage II and III cancers (Morris et al., 2007c, Chang et al., 2007). Conversely, node negativity has been associated with significantly improved survival (Wong et al., 2002).

In a systematic review of 61,371 patients it was found that a greater number of lymph nodes evaluated after surgical resection was positively associated with survival in patients with stage II and III colon cancer however the association is not clear cut (Chang et al., 2007). It has been demonstrated that specialist pathologists, as well as surgeons, are more likely to achieve adequate lymphadenectomy results (Morris et al., 2007c). However in a UK-based single-center audit lymph node harvest, following case-mix adjustment, was shown to vary according to the reporting pathologist but not the operating surgeon (Evans et al., 2008). Interestingly, patients with one or more positive nodes have greater nodal harvests than those with negative nodes (Evans et al., 2008, Johnson et al., 2002). In a further study the number of identified lymph node metastases increased continuously with increased total lymph node recovery in pT3 colon cancer (Tornroos et al., 2009). Variation in histo-pathological techniques may however substantially alter lymph node yield. Hernanz and colleagues have demonstrated that ‘fat clearing’ methods enhanced lymph node yield by an average of 10 nodes when examining fifty mesorectal specimens (Hernanz et al., 2009). As a result three patients in their study were upstaged. Other studies employing acetone preparation have similarly demonstrated efficacy at increasing lymph node yield using additional techniques (Vogel et al., 2008). Injection of methylene blue ex-vivo into the superior rectal artery of rectal cancer specimens has been shown to enhance lymph node detection by approximately 13 nodes in a small case-control study (Markl et al., 2007).

In rectal cancer, a prospective randomized multi-centre trial of 1227 patients found large variations in lymph node yield between pathologists and laboratories and also post neo-adjuvant radiotherapy as compared with those not undergoing neo-adjuvant therapy. In the same study patient age over 60 years and low invasion depth of the tumour were also associated with a lower yield. Recurrence free survival was shorter in patients who were node negative with fewer than seven lymph nodes retrieved as compared with node negative patients with more than eight nodes examined (Mekenkamp et al., 2009).

Recent validation of the Royal College of Pathologists minimum dataset for reporting colorectal specimens, using retrospective patient data from the UK, has shown the variables therein to be of prognostic significance and furthermore the failure to report on certain factors confers worse outcomes than clinical absence of the factors (Maughan et al., 2007). The collective efforts of the surgeon and the pathologist determine lymph node harvest. If this metric is to be used as a meaningful marker of quality of colon cancer care efforts to standardise surgery and histopathological techniques require consideration.

4.5.3 Outcome Metrics of quality in Colorectal Surgery

4.5.3.1 Intra-operative measures

Blood loss

Blood loss has been cited as a potential quality indicator in cancer surgery and although no recent randomised controlled trials have been published a narrative review of blood loss as a quality indicator in oncological surgery concludes the possible negative effects of transfusion in terms of immunosuppression, increased morbidity and potential long-term adverse effect upon oncologic outcomes.

Accurate measurement and reporting make this a potential marker that requires further evaluation (Dixon et al., 2009). Specifically, in colorectal surgery the receipt of blood transfusion was shown, amongst other factors, to be an independent predictor of anastomotic leak in patients undergoing elective anterior resection (Yeh et al., 2005). This is likely to be multi-factorial encompassing intra-operative as well as post-operative factors.

Perioperative surgical morbidity measures

The short-term problems surrounding anastomotic leaks are well known, but the potential long-term consequences are less well understood. Law and colleagues demonstrated significantly higher local and systemic recurrence rates as well as worse five year survival in a prospectively studied cohort of patients undergoing potentially curative resections for colorectal surgery in those patients that suffered anastomotic leak (Law et al., 2007). This may justify using leak rates as a quality measure in operations for colorectal cancer. In a study of 978 patients independent risk factors for anastomotic leakage included the use of irrigation-suction drains, blood transfusions, and an anastomotic level of 5 cm or less from the anal verge (Yeh et al., 2005). Interestingly, in rectal cancer surgery a review of 35 studies demonstrated no association between surgeon caseload and experience with respect to anastomotic leak rates (Iversen et al., 2007). A recent meta-analysis of the use of defunctioning stomas after low anterior resection has favoured the use of stomas in reducing clinical anastomotic leak rates and reoperation rates (Tan et al., 2009). The benefit of covering stomas has also been shown in other studies (Eberl et al., 2008, Den Dulk et al., 2009, Gastinger et al., 2005). In a smaller series employing hand-sewn anastomoses (Huh et al., 2007), defunctioning did not confer any advantage. A Cochrane review however has shown insufficient evidence to demonstrate benefit of hand-sewn over mechanical anastomosis in terms of complication rates (Lustosa et al., 2002).

As illustrated by Kingham and colleagues in their review, the causes of anastomotic leak may be multi-factorial and in some instances largely due to patient factors rather than technique as is highlighted by the bimodal distribution, with a significant proportion occurring whilst these patients are at home (Brisinda et al., 2009). Anastomotic leaks are a major cause of unplanned return to the operating theatre following colorectal surgery. Other causes in the perioperative period may include bleeding, wound dehiscence or stoma related complications. 'All-cause' unplanned return to theatre (i.e. re-interventions) has been cited as a potential quality marker in colorectal surgery. Morris and colleagues analysed 26,638 patients that underwent operations for colorectal cancer and found that amongst patients who underwent an unplanned intervention, a complication demanding intervention was associated with a significant increase in the likelihood of postoperative mortality (relative risk of 2.2). Moreover, the relative risk of mortality rose to 7.2 if further complications were encountered (Morris et al., 2007b). Significantly increased mortality rates have also been shown in other surgical specialities amongst patients requiring re-operation (Birkmeyer et al., 2001b). Merkow and colleagues found significant variability in re-operation rates between NSQIP hospitals in patients undergoing colorectal procedures and that ASA grade and being male (amongst others) were correlated highly with the likelihood of re-intervention (Merkow et al., 2009a).

4.5.3.2 Perioperative mortality measures

Perioperative mortality outcomes in colorectal surgery must reflect, at least in part, the performance of the surgeon. Consistently good outcomes are likely to represent a combination of strong pre-operative decision-making, appropriate case selection, the operative skill of the surgical team, as well as their synergy with intensivists and the wider hospital community when required. Postoperative mortality is an outcome measure that is easily collected, and can be validated and risk-adjusted (Keogh et al., 2004). It should be borne in mind however that postoperative mortality measures excludes patients who were declined for surgery. Unadjusted 30-day in-hospital mortality reported over a ten year period from the Hospital Episode Statistics database, a routinely-collected national dataset that encompasses the entire NHS in England, suggests that 30-day in-hospital mortality following elective colonic and rectal resection for cancer is 3.4% and 3.3% respectively (Faiz et al., 2009b). Investigators using a large US administrative database observed that between 1996 and 2003 30-day death rates in over 30,000 patients undergoing segmental colectomy for cancer were 6.7% at non-National Cancer Institute (NCI) designated centers and 3.2% at NCI centers. Mortality after proctectomy was 5.0% and 1.9% at these centers respectively (Paulson et al., 2008). In contrast, 30-day postoperative mortality following colorectal resection in Denmark was 9.9% as recorded by a prospective national audit over an eight month period (Nickelsen et al., 2005). Although seemingly high, the inclusion of urgent cases in the Danish study renders direct comparison to other countries difficult.

Interestingly, a national prospective audit of clinical registry data in the United Kingdom observed an unadjusted mortality risk of 7.5% following colorectal resection (Tekkis et al., 2003). Discrepancy between unadjusted HES mortality outcomes and those from the national clinical registry arise mostly due to the inclusion of urgent cases on the latter database. Furthermore, HES describes in-hospital mortality only, whereas the higher rates recorded within the clinical registry relate to in and out-of-hospital mortality. Such large differences in the measurement of unadjusted mortality rates highlight the difficulties encountered with apparently *straightforward* outcome metrics.

Critics of 30-day mortality highlight that a significant percentage of deaths occur after this 30-day window but within 90 days of surgery. In one study of patients undergoing colorectal surgery 90-day mortality rates were 4.1% and 28.9% for elective and emergency patients versus 1.4% and 15.5% for the same patient cohort at 30 days (Visser et al., 2009). A colorectal risk calculator devised using National Surgical Quality Improvement Programme (NSQIP) data allows surgeons to offer patients a 'hospital specific' preoperative mortality risk - which may offer a more meaningful insight into personal operative risk (Merkow et al., 2009a).

4.5.3.3 Oncological outcomes of quality in Colorectal surgery

Local recurrence following rectal cancer surgery is related to tumour factors that include: tumour site and its differentiation, the presence of lymph node metastases, lymphovascular tumour invasion, extramural vascular invasion, circumferential resection margin or serosal involvement (Dresen et al., 2009). Anastomotic leak requiring operative intervention has also demonstrated an association with loco-regional failure (Ptok et al., 2007b). However, what about surgery and oncological outcomes – does surgical proficiency matter in this regard? It been demonstrated by some (Garcia-Granero et al., 2009, Maslekar et al., 2007), but not all (Jeyarajah et al., 2007), investigators that in rectal cancer surgery, the quality of mesorectal excision is independently associated with oncological outcome metrics such as local recurrence.

But to what extent do the surgeon and the oncologist determine local disease failure? The Cooperative Investigators of the Dutch Colorectal Cancer group (Kapiteijn et al., 2002) compared outcomes from the Total Mesorectal Excision (TME) study that involved rectal surgery carried out by surgeons credentialed in TME, with outcomes from an older Cancer Recurrence and Blood Transfusion (CRAB) study where conventional surgery was performed without quality control. Local Recurrence (LR) rates were 16% in CRAB and 9% in the TME study suggesting that TME technique favours local control. The latter study offered a homogenous study population as none of the patients underwent radiotherapy and all were treated with curative intent. TME surgery appears to confer advantages over traditional rectal cancer surgery. Moreover, high quality TME, i.e. perfect mesorectal excision as evaluated by pathological examination, has been found by some investigators to be associated with improved local disease control (Garcia-Granero et al., 2009, Maslekar et al., 2007).

The impact of surgical intervention on loco-regional disease control is also influenced by the administration of adjuvant as well as surgical treatment. Dutch collaborators demonstrated the effect of short course preoperative radiotherapy (RT) followed by TME versus surgery alone. They reported LR rates of 2.4% in the RT+TME group versus 8.2% in the surgery only group at two years ($p<0.001$). They did not however identify a relationship with survival (Kapiteijn et al., 2001).

With regards to LR the magnitude of protection afforded by radiotherapy appears substantial. Moreover, significantly improved local disease control was also observed by the Dutch investigators when the study cohort was re-evaluated at six years (Peeters et al., 2007).

4.6 Discussion

In this section, the findings from the wider literature of surgical research, across various subspecialties are considered. This recognises that other non-colorectal specific works may have implications for colorectal surgery.

4.6.1 Structural metrics in general

Structural factors relate to any characteristics of hospital organizations' that may influence the quality of care delivered. Within healthcare various structural factors been shown to influence outcome. Although volume has been mentioned previously, specifically with reference to colorectal surgery, its use as a potential maker of quality has also been shown across a range of surgical procedures (Bentrem and Brennan, 2005, Luft et al., 1979, Birkmeyer et al., 2002, Begg et al., 1998, Sosa et al., 1998, Hannan et al., 2003, Wu et al., 2004, Katz et al., 2001, Killeen et al., 2007). Other structural factors that have been identified as relevant in health care investigations include: the number of doctors at a given institution (Jarman et al., 1999), their medical qualifications or degree of sub-specialisation (Chowdhury et al., 2007, Prystowsky et al., 2002), the nursing-to-patient ratio within institutions (Elixhauser and Halpern, 1999), the presence or absence of specialist equipment, the management culture of an institution (Glickman et al.,

2007) as well as the numbers of surgical beds within a surgical department (Brook et al., 1996). Structural factors represent easy quality metrics to assess but if they are to be used for these purposes it is essential that they accurately and consistently reflect the quality of the service being appraised.

Certainly, the evidence cited within the results section, drawn from the investigation of routinely-collected databases, offers support to the existence of a volume-outcome effect in colorectal surgery. Caution is perhaps warranted regarding the assumed validity of this association, and its consequent implications, for three reasons. Firstly, publication bias – resulting from the reluctance of journal editors to publish studies with apparent negative findings, may account for the paucity of reports of an insignificant relationship within the literature. Secondly, the volume-outcome relationship might confound the impact of other factors that genuinely influence outcomes. For example, in a 'free' referral system, surgeons deemed poor amongst their colleagues are unlikely to receive abundant referrals and therefore their volumes are likely to be low. Under these circumstances increasing surgical workload is unlikely to ameliorate poor performance. Thirdly, the volume-outcome relationship offers no insight into individual surgeon's, or institution's, ability to carry out a given procedure. Interestingly, a recently reported study, based on the data from Clinical Outcomes of Surgical Therapy (COST) study, demonstrated that 'credentialing' (i.e. establishing the proficiency of surgeons at a given technique) eradicated the volume-outcome effect in laparoscopic colorectal surgery (Larson et al., 2008). Perhaps such demonstration of proficiency by surgeons, or centers, represents a relevant, fair and accurate means of assuring surgical quality.

4.6.2 Process metrics in general

Process markers directly reflect the interactions that lead to effective care. Ideally, compliance with process measures should be reflected in subsequent good clinical outcomes. Often process metrics seek to compare current care with treatment standards and protocols and thereby highlight areas where quality may be lacking. Arriaga and colleagues, in their report from the 'Better Colectomy Project' have established a clear link between adherence to evidence based practice in colorectal surgery and good postoperative outcome (Arriaga et al., 2009). Specifically, they observed 40% non-adherence to a basket of thirty-seven evidence-based processes agreed by a consensus of expert opinion. In their retrospective review each consecutive failure to adhere to important processes increased the risk of postoperative complications by 60%.

Surgeons vary in their choice of surgical technique as well as their outcomes (McArdle and Hole, 1991). Although uniformity of practice is desirable on a societal level, benchmarking surgeons individually, according to their choice/use of surgical techniques may be hazardous. In colonic surgery use of the laparoscopic approach is not only dictated by surgeon choice but also by access to adequate training and facilities for minimal access surgery as well as appropriate case-mix. Evidence from a small study (n=72) investigating the place for minimal access rectal cancer surgery suggested that more complete macroscopic TME specimens were recovered following the laparoscopic approach – probably due to improved pelvic views (Gouvas et al., 2009b). However, complex inflammatory bowel disease, T4 stage colorectal cancer and re-do surgery represent circumstances where the evidence base for the laparoscopic approach is not clearly established at present. A single metric that evaluates laparoscopic usage as a quality marker may bias against surgeons who safely undertake significant proportions of such cases albeit using conventional surgery.

Some care processes that relate to the perioperative period have been shown to influence outcome. Approximately twenty perioperative processes including early nutrition (Fearon et al., 2005), epidural analgesia (Gendall et al., 2007) and goal-directed fluid administration (Noblett et al., 2006) have been incorporated into Enhanced Recovery Programmes (ERP's). A recent consensus document has been produced by the Enhanced Recovery After Surgery (ERAS) group which provides recommendations for each evidence based item for inclusion in an ERP (Lassen et al., 2009). Benefit differences have not been shown between either laparoscopic and open procedures within ERPs due to the paucity of high quality trials addressing this issue (Vlug et al., 2009). However, ERP's have demonstrated consistent efficacy at accelerating recovery after major elective colorectal surgery (King et al., 2006b, Wind et al., 2006). The overall benefits of ERP in colorectal surgery has been demonstrated in a meta-analyses by Gouvas and colleagues which showed reduction in lengths of stay, reduction in hospital morbidities and complication rates in elective patients undergoing colorectal resections (Gouvas et al., 2009a). It is uncertain whether this 'package' of evidence-based components that comprise ERP's represents the determinant behind improved outcomes, or, whether these are achieved through standardization of clinical care. It is known that team communication and coordination are qualities that are linked to low-morbidity hospitals (Young et al., 1998). Perhaps greater emphasis on team-work and clinical care processes underlie, at least in part, the benefits observed with fast-track surgery. Moreover, the latter underlying factors may only be effective if the correct processes are followed and furthermore if these processes are faithfully reproduced to a high standard.

4.6.3 Outcome measures in general

Outcome markers are obvious attractive quality metrics that commonly represent objective markers of clinical performance. Patient interpretation of clinical services through Patient Reported Outcome Measures (PROM's), as well as evaluation of the patient experience, can also be used to denote service quality and satisfaction.

Perioperative mortality has been used as a performance marker across various surgical specialties. Caution regarding its use as an index of quality in colorectal surgery is perhaps warranted as it is an outcome not necessarily determined solely by quality of surgical care. In fact, Iezzoni (Iezonni, 2003) described determinants of outcomes as the sum of patient factors, effectiveness of care as well as random variation that occurs between individual patients and providers. Thus over reliance upon mortality monitoring, or even outcome measurement in general, for benchmarking performance within colorectal surgery may mislead. Despite the above cautions mortality following surgery may represent an important marker of a hospital's ability successfully to manage patients who develop postoperative complications. In a recent report Ghaferi and co-workers demonstrated from eighty five thousand patients undergoing general or vascular surgery in American hospitals that patients suffering major postoperative complications were more likely to die in hospitals with high overall mortality rates (Ghaferi et al., 2009b). In contrast a patient's chance of survival following complications was greater when managed at low mortality centers. This study suggests that the management of complications, rather than the ability to avoid them, is perhaps an emerging, important marker of strong clinical care.

Four methodological factors provide potential impediments to the use of clinical outcomes as reliable discriminators of surgical performance. These include difficulties associated with accurate case-mix adjustment, the integrity of the dataset used for performance measurement and the magnitude of absolute numbers required for outcomes to identify poor performers in an appropriate time frame. Regarding risk adjustment, well-known scoring systems have been applied to colorectal surgery. These include systems such as the American Society of Anaesthesiologists Classification (ASA) (Davenport et al., 2006), the Association of Coloproctology of Great Britain and Ireland Colorectal Cancer Model , the NSQIP risk-adjustment method (Khuri et al., 1997, Young et al., 1998), APACHE II (Acute Physiology and Chronic Health Evaluation II)(Khwannimit and Bhurayanontachai, 2009b) , SAPS II (Simplified Acute Physiology Score-II) (Khwannimit and Bhurayanontachai, 2009a) and the Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM) score. The latter is a British scoring system that is used widely to predict mortality (Copeland et al., 1991) as well as benchmark surgeon performance (Sagar et al., 1996) in patients undergoing colorectal surgery (Midwinter et al., 1999, Sagar et al., 1994, Al-Homoud et al., 2004). Variants of the POSSUM score, such as the Portsmouth POSSUM (P-POSSUM) and the Colorectal POSSUM (CR-POSSUM) have also been developed. Comparison of the latter POSSUM scoring systems for the purposes of mortality prediction in colorectal surgery has failed to reach consensus regarding the superiority of one system over another (Senagore et al., 2004, Bromage and Cunliffe, 2007, Vather et al., 2006).

When considering statistical power calculations approximately 250 cases are required to identify a meaningful difference between a national average 30-day mortality rate of 3% and an institution with a mortality rate of 6% - i.e. a doubling of mortality risk. Two hundred and fifty cases may represent an appropriate caseload for institutional benchmarking as it reflects the cumulative caseload for an NHS institution over a 2-3 year period. It might however demand a decade of data to identify an outlying individual surgeon. Similar arguments may extend to other quality metrics when applied to the individual surgeon, such as local recurrence rates or use of restorative and non-restorative procedures for rectal cancer surgery. Amalgamation of outcomes into meaningful composite metrics may in the future address this problem.

4.6.4 The patient perspective

Patient experiences may not always directly coincide with performance outcomes as evaluated by surgeons. Health Related Quality of Life (HRQoL) questionnaires seek to quantify the effect of a disease, and its treatment, on patient wellbeing. Generic tools such as the EuroQol 5D (EUROQOL, 2009) and SF-36 (Medical Outcomes Trust, 2009) estimate overall quality of life irrespective of underlying disease. Specific tools used in colorectal cancer are the European Organisation for Research and Treatment of Cancer (EORTC QLQ CR38) (EORTC, 2009) and Functional Assessment of Cancer Therapy – Colorectal Scale (FACT-C) questionnaires (Ristvedt and Trinkaus, 2009). Quality of Life (QoL) scores are however time-dependent. In a study that examined QoL between two and five years following rectal cancer treatment the factors that were associated with lower scores on the FACT-C HRQoL instrument included severe faecal incontinence and male gender. Being further out from treatment was associated with more favourable scores. Interestingly, patients who had a colostomy did not have lower FACT-C scores or post-traumatic stress disorder symptoms (Ristvedt and Trinkaus, 2009) whilst having an anxiety trait predicted for poor QoL scores at 2-5 years following treatment.

	Generic	**Specific**
Structure	Hospital size, volume of cases, staffing levels, equipment in use, no. theatres.	No. of colorectal procedures performed by surgeon and Trust, No. colorectal surgeons, No. specialist colorectal nurses.
Process	Operative time.	Blood loss in CRC procedures, Lymph node retrieval.
Outcome	Diagnostic lead time, length of stay post op, mortality rates, re-intervention rates.	Case mix adjusted LOS, mortality rates, morbidity, functional outcome.

Table 7- Generic and specific markers of performance in surgery and colorectal surgery

Included within a patient's experience are both evaluations of the treatment received as well as appraisal of the service delivered. The latter is often termed satisfaction. Recently in England assessment of patient satisfaction has become compulsory amongst NHS hospitals. The Care Quality Commission (CQC), an independent healthcare regulator, surveys patients each year on their opinions regarding the care that they have received (Care Quality Commission, 2009b). These surveys cover broad aspects of care such as the admission process, ward stay and attitudes of medical and nursing staff. The relationship between patient satisfaction, HRQoL and actual clinical outcomes remains understudied at present.

4.7 SUMMARY

Considerable effort has gone into defining quality in colorectal surgery yet there is still no one metric or marker than can encapsulate this. Perhaps this stands to reason; with so many variables the likelihood that one metric serves this purpose is too optimistic. Having highlighted the importance of a multitude of different metrics and measures it is clear both quantative and qualitative measures are important.

What is apparent from the introduction is that choosing the correct dataset for performance benchmarking is vital. In the following study an appraisal of the two datasets available is undertaken. The HES administrative dataset will be compared with the colorectal specific NBOCAP registry to assess the strength of going forwards with either dataset for national perioperative colorectal resectional benchmarking.

5.0 COMPARISONS OF A VOLUNTARILY SUBMITTED CLINICAL REGISTRY WITH MANDATORY COLLECTED NATIONAL DATA FOR ASSESSING PERI-OPERATIVE MORTALITY IN COLORECTAL CANCER RESECTIONS.

5.1 Chapter overview

In this chapter comparisons are made between the reported outcome from two sources to appraise their potential use for national surgical benchmarking in colorectal surgery. Firstly the National Bowel Cancer Audit Programme (NBOCAP) is considered. This at the time of analysis is a voluntarily submitted clinical registry for colorectal cancer. This is compared to the HES data submission where submission is compulsory. To keep the comparison as reliable and free from interpretation bias as possible mortality, length of stay and readmission within 28 days, between the two data sources, for matched time periods have been chosen to compare. The overall aim is to assess which source may be more reliable given the differing reporting natures of each. This will then identify a data source going forwards for further analysis for the rest of the thesis.

5.2 Introduction

Quality and outcome measurement is becoming increasingly important within the National Health Service. Following the introduction of statutory regulation from the Department of Health in 1999, continual review of clinical performance, both from within one's own hospital and from outside, has become mandatory (Health, 1999). Compulsory self-audit became part of the General Medical Council (GMC) revalidation process for doctors from late 2009 (GMC, 2009). Given the importance of outcome measurement and the high prevalence of colorectal cancer (CRC) in the United Kingdom (UK) ensuring accurate and reliable outcome statistics for procedures undertaken is important for continual individual and service improvement (UK, 2006). The validity of outcome measurement may, however, depend upon the nature of how information is submitted. Voluntary reporting to clinical registries perhaps permits appraisal of only the best performing institutions whilst poor performing units that choose not to submit data remain undetected. This calls in to question the very nature of whether voluntarily reported outcome datasets are appropriate to use for performance measurement.

National patient outcome measures can be derived from datasets such as clinical registry data or routinely collected datasets. In the United Kingdom, the National Bowel Cancer Audit Programme (NBOCAP) is the principal clinical registry for colorectal cancer outcome. The Hospital Episode Statistics (HES) database comprises information relating to all patients admitted to NHS hospitals in England (all specialities). HES data are primarily collected for administrative purposes but have also been utilised extensively for clinical and health service research purposes and has been described in the introductory chapters in more detail (Dawson, 2005).

At the time of this study, NBOCAP represented a self-reporting dataset. Its data are derived from individual units submitting their own annual data. HES data are derived directly from Trusts' administrative data reporting systems (Faiz et al., 2008b) and it is compulsory to submit records to this dataset. National colorectal caseload and postoperative mortality have previously been compared between these two respective data sources (Garout et al., 2008). It is not, however, currently known whether voluntary data submission is associated with improved, or differing, clinical outcomes.

The aim was specifically to assess whether institutions that voluntarily self-report to NBOCAP demonstrate differing peri-operative clinical outcomes, as recorded by HES, than non-reporting institutions. This will help ascertain which data source of the two nationally available ones to take forward, for further use throughout the thesis.

5.3 Methods

Using the NBOCAP dataset, released in October 2009 (this was the most recently published document at the time of the analysis), Trusts was assigned to either 'submitting' or 'non-submitting' status depending on whether they had submitted data to NBOCAP from 1st August 2007 to 31st July 2008. Trusts that submitted <10% of their colorectal cancer resection workload were also termed non-submitters. Amongst Trusts that submitted <10% of their total colorectal cancer resection workload the range of submission was 1.3%-7.1%. Outcome for all patients operated on in the two groups using information from the HES database were compared. All NHS Trusts in England that performed the selected procedures were included.

Procedures were chosen that appeared on both datasets and were classified according to Office of Population, Census and Surveys codes, version 4 (OPCS-4). During the study period, HES records were retrieved for patients with a primary diagnosis of colorectal cancer. This included colon, recto-sigmoid and rectal cancer. Patients that had undergone the following operations were included: right hemicolectomy (OPCS-4 codes, H06-7), transverse colectomy (OPCS-4 code, H08), left hemicolectomy (OPCS-4 code, H09), sigmoid colectomy (OPCS-4 code, H10), abdomino-perineal excision of rectum (APER; OPCS-4 code, H33.1) , Hartmann's procedure (OPCS-4 code, H33.5), anterior resection (OPCS-4 codes: H33.2-4, H33.6) in keeping with the NBOCAP dataset. Site of tumour was re-coded into three categories- colon, recto-sigmoid or rectal according to the NBOCAP registry.

Patients undergoing excision of benign lesions were excluded as were patients that underwent procedures in two Trusts on HES that did not appear in the NBOCAP dataset were excluded. Only three consultant episodes were excluded through removal of the latter two Trusts. Furthermore, one Trust was excluded that appeared in the NBOCAP report, as it is an oncological centre not performing any surgery. Laparoscopic and open, elective and emergency operations were included in the initial analysis.

5.3.1 HES database

The HES database is a nationally collated data warehouse containing details of admissions of NHS patients in England since 1986. The data are taken from each hospital's Patient Administration System (PAS) for every inpatient and (since 2003/4) outpatient Finished Consultant Episode (FCE). Demographic and any procedural/diagnostic information within that admission period is captured against a unique patient identifier. The patient identifier allows identification of previous or subsequent admissions/procedural data pertaining to that patient. Each record also contains up to 13 International Classification of Disease 10th revision (ICD-10) secondary diagnoses and up to 12 [at the time of the analysis] procedural interventions recorded using the OPCS-4 codes (Faiz et al., 2008a). Deriving clinical outcomes from HES has been previously reported (Faiz et al., 2010c).

5.3.2 Charlson co-morbidity Index

The Charlson comorbidity scoring system is validated for use to predict patient outcomes based on concurrent co-morbidities (Newschaffer et al., 1997). It is a commonly used comorbidity index developed for administrative datasets and has been previously used in CRC patients (Rieker et al., 2002, Almoudaris and Omar Faiz, 2010). Comorbidities conferring worsening outcomes are given greater values. The secondary diagnosis fields were used to create the Charlson comorbidity index.

The Carstairs index is a composite deprivation score calculated at the output-area level (average population 1500) converted into population-weighted quintiles. The score has been shown to be a good reflection of material deprivation factors. The index is used by health economists and government agencies for this purpose (Dolan et al., 1995, Morgan and Baker, 2006). This was included for assessment as a social deprivation co-variate in to the model.

5.3.3 Emergency Status

Admission status on the HES database is recorded as 'elective' or 'non-elective'. All non-elective admissions were considered as emergencies.

5.3.4 Outcomes

30-day in-hospital mortality, length of stay (the mean natural logarithm with back-exponentiation) and 28-day readmission rates were primary end points from the HES database. 30-day in-hospital mortality is defined as death from all causes occurring in hospital within 30 days of admission.

5.3.5 Statistical Analysis

Mortality, length of stay and readmission within 28 days were analysed using HES data and compare to outcomes between submitting and non-submitting units using the NBOCAP dataset. Chi-square testing was used to assess categorical variables including age that had been re-coded into age bands. Owing to its non-normal distribution, length of stay was analysed by log-normal transformation and independent t-testing with back exponentiation. Multiple logistic regression analysis of 30-day mortality was performed. Covariates with significance $p \leq 0.1$ on univariate analyses were included in multifactorial regression analyses.

Statistical Package for Social Sciences, version 17 for Windows was used for the statistical analysis (SPSS, Chicago, IL, USA).

5.4 Results

Between 1st August 2007 and 31st July 2008, 17,722 patients were identified from HES as having been diagnosed with colorectal cancer and who underwent a major resection for colorectal cancer. Over the same time period NBOCAP reported 14,780 cases (83.4%). The same 152 Trusts in England were compared between the two databases. There were 20 Trusts in the non-submitter group (15 Trusts that submitted no data and a further 5 Trusts that submitted less than 10% of their total caseload). 3 missing emergency/elective admission status records were found in the submitter arm and were thus excluded from further analysis.

Submitters and non-submitters were statistically similar in terms of age and gender and type of procedure performed (Table 8).

Table 8-Table of demographics of submitter versus non-submitter groups

	Submitter (%) n=15815	Non-Submitter (%) n=1907	p-value
Sex			
Male	8829 (55.8)	1039 (54.5)	0.265
Female	6986(44.2)	868 (45.5)	
Age group			0.717
20-40	390(2.5)	40 (2.1)	
41-60	4227(26.6)	505(26.5)	
61-80	9925(62.8)	1198(62.8)	
>80	1273(8.0)	164(8.6)	
Charlson Index			
0	229 (1.4)	40 (2.1)	
1	60 (0.4)	14 (0.7)	
2	8984 (56.8)	1074 (56.3)	
3	2629 (16.6)	289 (15.2)	0.013
4	524 (3.3)	48 (2.5)	
5	2537 (16.0)	332 (17.4)	
6	843 (5.3)	110 (5.8)	
Carstairs			
Less deprived-1	3196(20.2)	377(19.8)	
2	3695(23.4)	481(25.2)	
3	3534(22.3)	497(26.1)	<0.001
4	2950(18.7)	328(17.2)	
Most deprived-5	2423(15.3)	224(11.7)	
Unclassified	17(0.1)	0(0.0)	

Co-morbidity scoring differed statistically between submitters and non-submitters (p=0.013). Fewer patients with comorbid disease were present in the submitting group. Submitters performed relatively more operations utilising the laparoscopic approach than non-submitters 18% (2847/15815) vs. 15% (287/1907) (p=0.001) (Table 9). Submitting Trusts performed fewer operations as emergencies (i.e. non-electives) than non-submitting Trusts: 17.5% (2761/15815) vs. 19.3% (369/1907) (p=0.041) Table 9.

Table 9- Table of type of resection and mode of admission by submitter status

Table 9	Submitter (%)	Non-Submitter (%)	p-value
Diagnosis			
Colon	10213 (64.6)	1237 (64.9)	
Rectosigmoid	1322 (8.4)	156 (8.2)	0.954
Rectum	4280 (27.1)	514 (27.0)	
Laparoscopy			
Yes	2847 (18.0)	287 (15.0)	0.001
No	12968 (82.0)	1620 (85.0)	
Admission type			
Elective	13051 (82.5)	1538 (80.7)	
Emergency	2761 (17.5)	369 (19.3)	0.041

A statistically significant increased risk of in-hospital death within 30 days of admission was observed in the non-submitting institution group [5.2% (100/1907) vs. 4.0% (628/15815), p=0.005]. Following case-mix adjustment for age, gender, comorbidity, social deprivation, use of laparoscopic technique, emergency/elective admission status and bowel location of tumour, submission of data to NBOCAP was associated with an independent reduction in 30-day mortality risk (OR=0.76, CI=0.61 0.96, p=0.021) (Table 10). A greater mean length of stay, following natural logarithmic transformation and back-exponentiation, in the non-submitter group was also observed (Table 11).

Table 10- Univariate and multivariate analysis of variables influencing 30 day mortality

	Univarate analysis				Multiple regression analysis			
		95% CI				95% CI		
	OR	**Lower**	**Upper**	p	**OR**	**Lower**	**Upper**	p
Submitter vs non-submitter	0.77	0.61	0.96	0.022	0.76	0.61	0.96	0.021
Elective vs Emergency	4.26	3.57	5.07	<0.001	4.04	3.43	4.77	<0.001
Age	1.08	1.07	1.09	<0.001	1.08	1.07	1.09	<0.001
Laparoscopy vs open	0.59	0.44	0.78	<0.001	0.58	0.43	0.77	<0.001
Carstairs				<0.001				<0.001
Less deprived-1	1.00				1.00			
2	0.96	0.74	1.25	0.777	0.96	0.74	1.25	0.782
3	1.18	0.92	1.51	0.200	1.18	0.92	1.52	0.196
4	1.39	1.08	1.79	0.011	1.39	1.08	1.79	0.010
Most deprived-5	1.56	1.20	2.02	0.001	1.56	1.20	2.02	0.001
Gender female vs male	0.82	0.70	0.96	0.014	0.81	0.69	0.95	0.010
Charlson Score				<0.001				<0.001
1	1.00				1.00			
2	2.93	1.45	5.92	0.003	2.94	1.45	5.93	0.003
3	1.14	0.70	1.84	0.609	1.14	0.70	1.84	0.607
4	2.16	1.32	3.55	0.002	2.15	1.31	3.54	0.002
5	3.29	1.89	5.73	<0.001	3.27	1.88	5.69	<0.001
6	1.99	1.21	3.28	0.007	1.97	1.20	3.24	0.008
7	3.52	2.10	5.89	<0.001	3.48	2.08	5.83	<0.001
Diagnosis								
Colon	1.00			0.131				
Rectosigmoid	0.97	0.71	1.33	0.855				
Rectal	1.24	0.99	1.53	0.051				

Table 11- Study endpoints combined for patients undergoing elective and emergency surgery in submitting and non-submitting Trusts

Study End points	Submitter (n=15815)(%)	Non-Submitter (n=1907)(%)	p-value
30-day mortality	628 (4.0)	100 (5.2)	0.005
Length of stay			
Mean	10.9	11.5	0.013
+/- SD	+/-2.0 (8.9,12.9)	+/-1.9 (9.6,13.4)	
Readmission			
No	14023 (88.7)	1703(89.3)	0.408
Yes	1792 (11.3)	204 (10.7)	

5.5 Conclusion

A minor but significant difference in mortality rates between reporting and non-reporting Trusts to a clinical registry have been demonstrated. The clinical significance of the difference observed in real terms appears small however it may belie wider implications. If the findings potentially reflect genuine outcome differences that arise between Trusts that do and do not voluntarily submit data to clinical registries, then mandating reporting to such registries may be a future consideration if they are to be used nationally to benchmark performance and quality amongst surgical providers. However at the time of this analysis, it appears that the HES database is more reliable in reporting of important outcome measures in colorectal cancer surgery during the peri-operative period. For this reason and given the fact that this thesis is concerned with peri-operative outcome it has been decided to progress using HES as the primary data source.

However, the limitation of HES that makes the use of clinical registries attractive, with respect to peri-operative outcome, is the recording of complications and patients subsequent outcome. At the time of writing HES had not been used to report on complications in colorectal cancer surgery. If some inference of complications and their management could be made, this would greatly support the future use of HES for national quality performance appraisal.

6.0 FAILURE TO RESCUE-SURGICAL (FTR-S): A NOVEL MARKER OF SURGICAL COMPLICATION MANAGEMENT IN COLORECTAL CANCER PATIENTS DERIVED FROM HOSPITAL EPISODES STATISTICS (HES)

6.1 Chapter overview

If the HES database is to be used going forwards, as identified from the quality metrics and the systematic review, as well as the wider literature, complication management is a vital determinant of patients' subsequent outcome. This is especially true in the peri-operative period. What the HES database lacks in its native form is any reflection of such complications. In this chapter attempt is made to derive a novel metric from the HES database and see if it can be used to report on aspects of surgical complication management that would strengthen the use of HES for national benchmarking and the identification of high performing units.

6.2 Introduction

The management and prevention of complications has been the focus of increasing research interest in contemporary surgical literature (Berenguer et al., 2010). This is especially true in patients undergoing major colorectal surgery, where the presence of serious postoperative morbidity significantly increases the likelihood of death (Longo et al., 2000). Similar findings have been observed in other specialities also (Lebeau et al., 1990). Silber and colleagues termed the phrase

‘Failure to Rescue’ (FTR) describing patients who died after an acquired complication in hospital (Silber et al., 1992). The metric they described represents the proportion of deaths that occur in those patients who experience complications. In a study investigating patients undergoing cardiac surgery they ranked hospitals according to their case-relevant mortality. They observed significant differences in FTR rates, between the best and worst ranked units despite equivalent complication rates (Silber et al., 1995). This suggests that the institutional management of complications is an important factor determining whether patients survive following the occurrence of morbidity. As such, FTR potentially represents a useful metric of complication management.

The Agency for Healthcare Research and Quality (AHRQ) is the health services research arm of the United States Department of Health and Human Services (HHS) (Services, 2010). It produces freely available indicators of hospital quality, termed ‘Patient Safety Indicators’ (PSIs) that make use of inpatient hospital administrative data. Currently, one such indicator is termed ‘Death among surgical in-patients with serious treatable complications’(AHRQ, 2010) and is synonymous with FTR. The AHRQ defines this as the number of deaths in patients suffering one or more complications, from a select group of medical or surgical complications, per 1000 discharges. Furthermore, potentially avoidable complications represent a significant economic burden also. Medicare and Medicaid, two US governmental programmes that provide medical and health services to non-privately insured US citizens, have introduced measures withholding remuneration for certain perceived avoidable complications (Services, 2008).

While FTR is undoubtedly an important concept, current analyses using administrative datasets may not effectively distinguish between pre-existing conditions (Talsma et al., 2010). Conditions Present On Admission (POA) have been shown to impact upon FTR rates when recalculated using refined algorithms (Moriarty et al., 2010) or upon case record review (Horwitz et al., 2007, Moriarty et al., 2010, Talsma et al., 2010). This questions the fitness for purpose of FTR in such a form. Under these circumstances prior studies that have included total morbidity, i.e. complications of all severity, yield complication rates of up to 36.4% (Ghaferi et al., 2009b). As a result, it may not be immediately apparent to what degree complications occurring in surgical patients are attributable to the surgical processes/decision making and what degree they are attributable to conditions present on admission that manifest after surgical interventions.

Refinement of the FTR paradigm in the surgical context should consider the fact that patients experiencing some of the worst preventable complications are those that necessitate a return to theatre in the postoperative period. Evidence also suggests that patients that are re-operated are more likely to die (Ricciardi et al., 2012). It is suggested that this group of patients with the most serious surgical complications i.e. requiring a reoperation represent a well-defined group that could be a useful target for quality improvement. The term '**Failure To Rescue - Surgical** (FTR-S) is introduced as- the proportion of patients with surgical complications that die during their index admission following an unplanned re-operation.

The aim is to firstly see if such a metric can be derived from the HES. If possible to derive, the applicability of using FTR-S rates, as a reflection of complication management will be explored. If derivable from routinely-collected data, this may represent a potential novel marker of surgical quality assessment. Secondly, the aim is to assess if any variability in FTR-S rates exists amongst units in England, in patients who have undergone a primary resection for colorectal cancer. This may help to validate this novel metrics for use.

6.3 Methods

6.3.1 Patient identification

The methods follow those described in Chapter 5. For this analysis, patients were grouped into four age cohorts: 17-54, 55-69 and 70-79 and>79 years for purposes of the analysis.

6.3.2 Risk adjustment models

Similarly, the models were created in-line with the methodology described in Chapter 5. In this analysis, the Charlson score was considered in three categories 0, 1-4 and ≥5.

6.3.3 Re-operation

A patient was classed as undergoing a re-operation if, on their index admission they were returned to theatre for a procedure as shown in Table 12. Elaine Burns is acknowledged for coding the initial re-operation dataset. From this, discrete re-operative procedures were chosen. The choice of procedures aimed to reflect those that were most likely to represent unplanned returns to theatre, due to surgical complications. Time to re-operation was calculated as the number of days from the index operation to re-operation. Returns to theatre up until midnight on the same day as the index resection are not discernible from the HES database for any procedure except for re-opening of abdomen- which is considered in the washout of abdomen category. Returns to theatre for examination under anaesthetic alone were excluded as these procedures were rarely solely coded for as the only reason for a return to theatre. More often they were coded alongside an additional reason for return to theatre; therefore, these patients were considered under the more 'major' codes.

Table 12-Types of reoperation and association with high and low mortality quintiles

	Lowest Mortality Quintile		Highest Mortality Quintile		p- value	
Primary reason for reoperation	No. of reoperations (n=1144)	Deaths (n=127)	No. of reoperations (n=1386)	Deaths (n=233)	Reoperations	Deaths
Washout of abdomen	177 (15.5)	22 (17.3)	221 (15.9)	45 (19.3)	0.681	0.281
Small bowel resection	75 (6.6)	14 (11.0)	80 (5.8)	19 (8.2)	0.413	0.351
Further colorectal resection	192 (16.8)	26 (20.5)	238 (17.2)	67 (28.8)	0.933	0.073
Drainage of intra-abdominal abscess	45 (3.9)	5 (3.9)	54 (3.9)	5 (2.1)	0.873	0.247
Division of adhesions	86 (7.5)	7 (5.5)	78 (5.6)	14 (6.0)	0.037	0.881
Stoma formation / operation on stoma	351 (30.7)	55 (43.3)	430 (31.0)	105 (45.1)	0.681	0.690
Wound complications	218 (19.1)	18 (14.2)	285 (20.6)	20 (8.6)	0.380	0.094

6.3.4 Organizational structural information

Information on hospital structural variables were derived from the Department of Health Hospital Activity Statistics website (Department of Health, 2010). Yearly data were collated and averaged over the study period for each Trust. Consideration has been made for the merging of units by aggregating the average values of merged units over time. To account for differences in unit size, structural factors were calculated per bed number and subsequently compared. For ease of description, average Intensive Care Unit (ICU) and High Dependency Unit (HDU) bed numbers and the number of theatres per in-patient bed were each multiplied by a factor of 100. Inclusion of structural data is intended to be reflective of the wider hospital within which the colorectal units function.

6.3.5 Statistical analysis

Statistical analyses were carried out using SPSS Version 18.0 (Statistical Package for Social Sciences, SPSS, Chicago, Illinois, USA). Categorical variables were investigated using the Chi-squared test. The Mann–Whitney *U* test was used for non-parametric tests of structural data. Logistic regression analysis was used to investigate predictors of postoperative re-operation and for risk adjustment. Factors with a significance level of ≤ 0.1 on univariate analysis were included in the following step of the regression analyses. For tests of significance, p values of <0.05 were considered significant.

The methodology of ranking and comparing units was based on a previously published study (Ghaferi et al., 2009a). The overall risk-adjusted mortality was calculated for each unit. The risk-adjustment model included the type of resection, patient age, use of laparoscopy, gender, admission status (elective or non-elective), Charlsons comorbidity score and Carstairs index as co-variates. Logistic regression was used to predict the probability of death for each patient; these were summed for patients at each unit to estimate expected mortality rates. Next, the ratio of observed to expected mortality was multiplied by the overall mortality rate for each operation type to obtain the risk-adjusted mortality rate for each unit. Units were then stratified into quintiles (n=30 units per quintile) according to their risk-adjusted overall mortality rates.

Units in the lowest mortality quintile are abbreviated to LMQ and those in the highest mortality quintile are abbreviated to HMQ. LMQ and HMQ units were compared based on overall risk-adjusted mortality rate. Intermediate quintiles were also analyzed and to demonstrate any stepwise same direction effect in outcome- otherwise known as a montopic effect.

6.3.6 Funnel plot

Funnel plots were constructed using the tools available at http://www.erpho.org.uk/topics/tools/funnel.aspx. Funnel plots are validated methods of graphically representing performance data (Mayer et al., 2009a). The control limits use a normal approximation to the Poisson distribution.

6.4 Results

6.4.1 Demographic clinical characteristics

144,542 patients underwent primary colorectal resections for colorectal cancer between the study dates in 150 English NHS units. 110,587 (76.5%) patients underwent elective procedures with the remaining 33,955 (23.5%) undergoing non-elective procedures. There were 80,400 (55.6%) men and 64,142 (44.4%) women in the study. The number and type of resections for the whole cohort is described in Table 13.

6.4.2 Demographics according to quintile

There were 25,082 (17.35%) patients in 30 units within the LMQ compared with 27,630 (19.12%) patients in 30 units within the HMQ. Table 13 offers a description of characteristics for patients operated on between the extreme quintiles. Overall older patients were operated on in the LMQ units (p<0.001). More men were operated on in HMQ units (55.6% vs. 54.7%, p=0.026). There were more socially deprived patients in the LMQ units compared with more co-morbid patients in the HMQ units (p<0.0005). There were no differences in open / laparoscopic approach between the quintiles (p=0.927), but more rectal and subtotal/total procedures (p<0.008) were performed in the HMQ units (Table 13).

Table 13 Demographics between the different quintile groups

	Lowest mortality quintile n=25,082 (%)	Highest mortality quintile n=27,630 (%)	p-value
Age bands			
17-54	2310 (9.2)	2630 (9.5)	<0.001
55-69	8246 (32.9)	9405 (34.1)	
70-79	8863 (35.3)	9845 (35.6)	
>79	5663 (22.6)	5750 (20.8)	
Sex			
Male	13,709 (54.7)	15,369 (55.6)	0.026
Female	11,373 (45.3)	12,261 (44.4)	
Admission method			
Elective	19,393 (77.3)	21,177 (76.6)	0.067
Emergency	5689 (22.7)	6453 (23.4)	
Charlson			<0.001
0	13,536 (54.0)	16,686 (60.4)	
1-4	2105 (8.4)	2379 (8.6)	
>5	9441 (37.6)	8565 (31.0)	
Carstairs			<0.001
Less deprived 1	5558 (22.2)	5103 (18.5)	
2	6699 (26.7)	6100 (22.1)	
3	5737 (22.9)	6064 (21.9)	
4	4340 (17.3)	5468 (19.8)	
5	2704 (10.8)	4890 (17.7)	
Unclassified	44 (0.2)	5 (0.0)	

Colectomy / resection type			
Left sided *	5790 (23.1)	6297 (22.8)	0.008
Right sided	8978 (35.8)	9588 (34.7)	
Subtotal / total	1007 (4.0)	1192 (4.3)	
Rectal	9307 (37.1)	10,553 (38.2)	
Surgical approach			
Open	23,580 (94.0)	25,970 (94.0)	0.927
Laparoscopic	1502 (6.0)	1660 (6.0)	

**included Hartmann's resection*

6.4.3 Organisational structural factors

No significant differences were demonstrated between the quintiles in terms of the average number of in-patient beds (p=0.196), use of radiological imaging (Computerised Tomography [p=0.174], non-gynaecological ultrasound [p=0.515] or fluoroscopy [p=0.069]), the number of available theatres [p=0.233] or Intensive Care Unit beds [p=0.425]). A difference was however observed in the number of High Dependency Unit (HDU) beds available between the LMQ and HMQ units favouring more beds in the lower mortality quintile units (p=0.011) Table 14.

Table 14- Median values of the number of scans performed per patient-bed per year, number beds per inpatient beds multiplied by a factor of 100 and number of operating theatres per inpatient bed multiplied by a factor of 100, Mann-Whitney U test

Structural factor compared	Lowest mortality quintile (median)	Highest mortality quintile (median)	p value
Size-			
Average no. beds per unit	683.82	791.00	0.196
Imaging-			
CT scanning	17.33	14.66	0.174
Ultrasound scanning (non gynaecological)	27.54	28.90	0.515
Fluoroscopy	9.82	7.40	0.069
Level I+II beds-			
ITU beds	1.32	1.05	0.425
HDU beds	1.04	0.78	0.011*
Theatres-	2.30	2.20	0.233

6.4.4 Overall Mortality by quintile

A significant difference was observed between unadjusted mortality rates at the LMQ units (4.1% [1016/ 25,082]) and in the HMQ units (7.6% [2106/27,630], p<0.001). When this is adjusted for the previously described co-variates the overall risk-adjusted mortality for the best and worst quintile units was 5.42% and 9.31% respectively (p=0.029).

6.4.5 Re-operation rates

Table 15- Mortality, reoperation and FTR-S rates per quintile with figures in parenthesis representing adjusted values

Quintile	Mortality rate (adjusted) %	Re-operation Rate (adjusted) %	FTR-S rate (%)
1st (lowest mortality)	4.1 (5.4)	4.6 (4.8)	11.1
2nd	5.3 (5.2)	4.7 (4.7)	13.0
3rd	6.2 (5.5)	6.0 (5.0)	14.4
4th	7.2 (6.7)	5.1 (4.8)	15.4
5th	7.6 (9.3)	5.0 (4.8)	16.8

When including all quintiles the overall re-operation rate was 4.8%. (6911/144,542). When units were stratified into quintiles for overall risk-adjusted mortality, re-operation rates were similar at the LMQ and HMQ units respectively [4.6% (1144/25,082) and 5.0% (1386/27,630), p=0.125]. When the re-operation rates were risk adjusted, the re-operation rates were comparable at 4.8% and 4.8% respectively (p=0.211) see Table 15.

6.4.6 Time to re-operation

Figure 7 shows the distribution of days from index procedure to re-operation between quintiles. This demonstrates that patients were returned to theatre at equivalent times after their index procedure between LMQ and HMQ units (6.5 days vs. 7.0 days, p=0.858) and peaked in a bimodal distribution for both types of units at 24-48hours and then at 7-9 days.

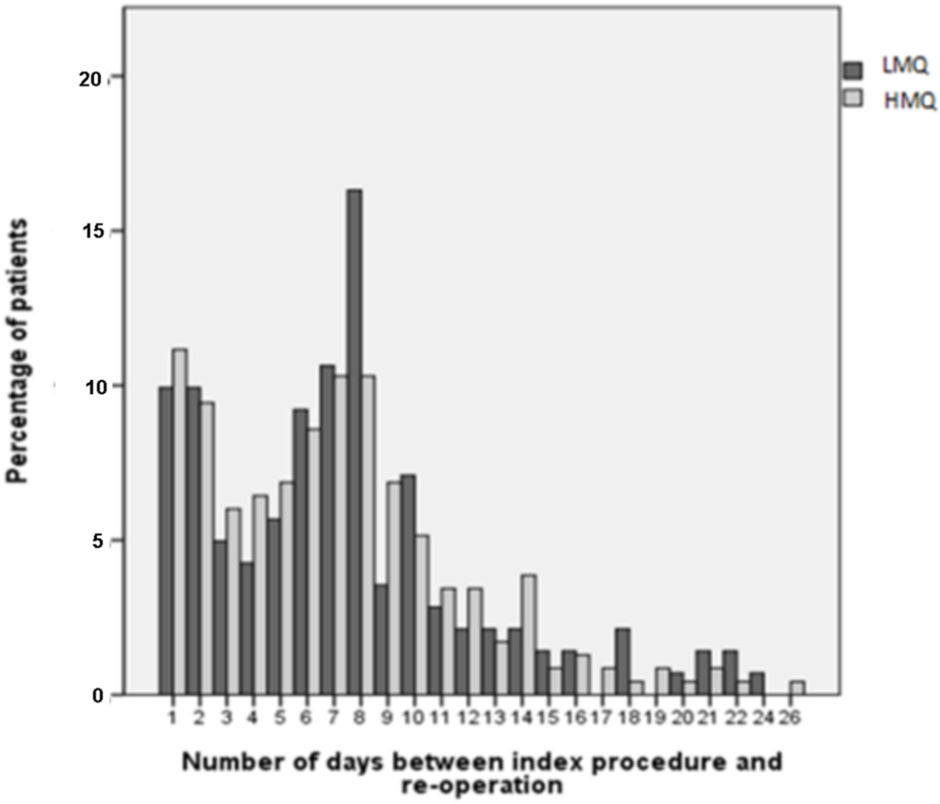


Figure 7- A paired bar chart demonstrating time to re-operation for high and low mortality quintile groups

6.4.7 Reasons for return to the operating room

In patients that required a re-operation, besides division of adhesions where there was a small statistical difference in frequency between the quintiles, there was no statistically significant difference in the reasons for return to theatre between the two quintiles for any other cause of returns to theatre see Table 12.

6.4.8 Failure to Rescue rate after serious surgical complications (FTR-S)

Despite differences in overall risk-adjusted mortality, little significant differences have been demonstrated in the patient demographics, availability of structural factors, re-operation rate, types and lead time to re-operation between the extreme quintiles. However, when patients underwent a re-operation, 11.1% (127/1144) died in LMQ units versus 16.8% (233/1386) of patients in HMQ units based on risk-adjusted mortality (p=0.002). Overall patients requiring a re-operation in the HMQ units were 1.7 times more likely to die than had they undergone a re-operation in a LMQ unit.

6.4.9 Re-operation type before death

In those patients that were not rescued after a re-operation (i.e. FTR-S), Table 12 shows which re-operations were undertaken before death, by quintile. Relatively, operations for wound complications were more likely to be performed before death in LMQ units when compared with HMQ units (14.08% vs. 8.58%, p=0.094) as were drainage of intra-abdominal abscesses (4.23% vs. 2.15%, p=0.247); however there were no statistically significant differences observed in re-operation types undertaken before death between the quintiles.

6.4.10 FTR-S and Volume

Figure 8 is a funnel plot showing the FTR-S as observed over expected (O/E) FTR-S following re-operations for both quintiles. This demonstrates the O/E mortality of FTR-S is not clearly related to overall re-operation volume. FTR-S at 14 units lay above the 2SD control limit signifying greater than merely random variation in this outcome measure.

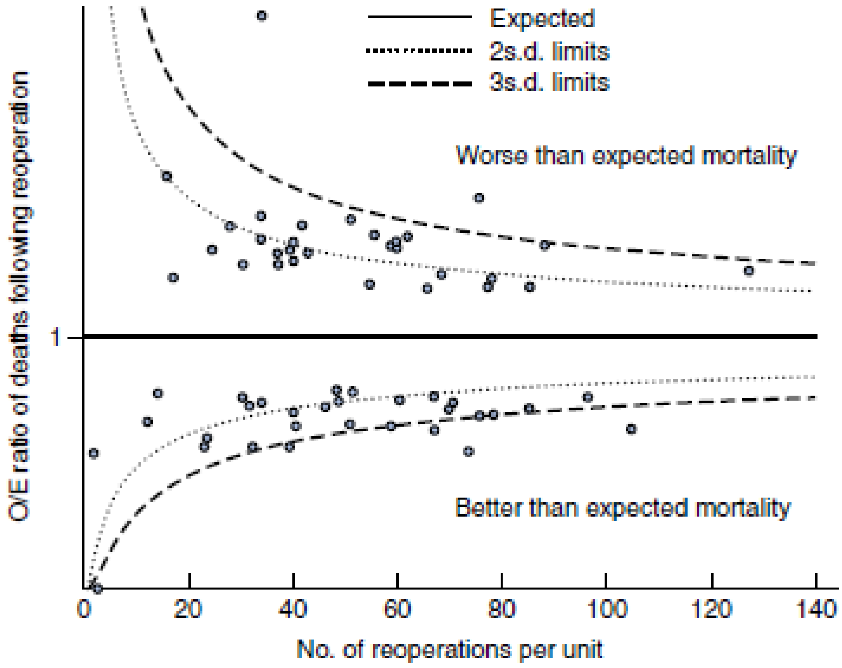


Figure 8- Funnel plot demonstrating the Observed over expected number of deaths following reoperations by unit volume with those units above one being in the high mortality quintile

7.0 Notable limitations

Whilst operative re-interventions are important the importance of non-operative interventions such as the use of image guided drains in the management of serious surgical complications is acknowledged. On analyzing the dataset, such interventions were poorly coded. As a result they were excluded from the analysis to minimize the influence of reporting bias in the final results.

6.5 Conclusion

A new surgically relevant marker termed FTR-S has been defined. This marker has been derived from routinely collected HES data. The results demonstrate that high mortality hospitals (despite equivalent reasons and frequency) of returns to theatre have worse outcome from complication management. It has also been shown to discriminate between high and low mortality units giving it construct validity. This study demonstrates the ability to identify serious surgical complications and importantly also appraise their outcome. This is likely to be a very important metric for identifying high performing units from HES.

Due to the limitations of the data from HES a notable exception to this study is the influence of re-interventions that were non-operative- specifically the use of image-guided drains. These are important and should be considered in any evaluation of a unit's complication management. As previously stated, at present coding in colorectal surgery is not robust enough to include these in a formal analysis. Given the importance of such interventions when considering outcome after complications it will be attempted to assess whether it is possible to discern their use in allied specialities (e.g. upper gastro-intestinal cancer surgery) where they occur relatively more commonly, as a proof of concept.

7.0 INFLUENCE OF OPERATIVE AND NON-OPERATIVE RE-INTERVENTIONS IN OESOPHAGO-GASTRIC CANCER RESECTIONS IN ENGLAND FOLLOWING SURGICAL COMPLICATIONS

7.1 Chapter overview

It remains to be seen whether non-operative re-interventions e.g. image guided drains can be derived from HES. Any complete consideration of complication management in major surgery must be able to appraise this type of intervention. Due to data limitations in colorectal surgery an attempt has been made in this chapter to use oesophago-gastric cancer surgery where non-operative re-interventions occur more frequently and are more robustly coded for.

7.2 Introduction

There were 462,000 new cases of oesophageal cancer (UK, 2011a) and 988,000 gastric cancer diagnosed annually worldwide (UK, 2011b). Surgery is the mainstay for cure. Oesophageal and gastric cancer resections are, however, associated with significant postoperative mortality with notable variability between centres. The volume-outcome relationship has been demonstrated in upper gastro-intestinal cancer surgery (Skipworth et al., 2010, Pal et al., 2008, Markar et al., 2011, Anderson et al., 2011), which has led the drive to centralization of services in England.

Another major determinant of outcome is the quality of postoperative complication management. The relationship between a hospital's ranking for complication rate and rank for mortality is not a linear one. It has been shown that hospitals that rank worst for mortality are not necessarily those that rank worst for the number of complications accrued, suggesting that complication rate is not the only determinant of outcome following postoperative complications (Silber et al., 1995). In colorectal surgery it has been shown that despite similar surgical complication rates, variability in outcome following the management of these complications occurs between high and low mortality hospitals- this is termed failure to rescue-surgical (FTR-S) (Almoudaris et al., 2011b). This metric is thought to represent a marker of how well post-operative surgical complications are managed. It may explain to a degree why such mortality differences and variability occur between colorectal cancer surgery units. In oesophago-gastric cancer surgery, such a relationship between re-interventions and postoperative outcomes has not been examined.

The hypothesis is that units with lower overall postoperative mortality re-intervene more often and are subsequently more successful than units with high mortality. This study aims to observe for differences in patient survival following re-operations for complications that may occur, between high and low mortality hospitals undertaking oesophago-gastric cancer resections in England. A further aim of the study is to assess the usage and subsequent outcome of non-operative interventions such as endoscopic and interventional radiological therapies.

7.3 Patients and Methods

7.3.1 Study Design

This study was a retrospective national cohort study of all patients diagnosed with a gastric or oesophageal cancer that underwent an elective primary major surgical resection between April 2000 and March 2010 inclusive in England. Data were obtained from the Hospital Episodes Statistics (HES) database as described in previous chapters.

7.3.2 Patient selection

All patients that underwent a primary resection for oesophageal or gastric cancer were included in the analysis. International Classification of Disease version 10 (ICD-10) codes were used to identify these patients. OPCS-4 procedure codes used to identify oesophagectomy and gastrectomy were G01, G02, G03, G27 and G28. Ravi Mamidanna is acknowledged for undertaking the initial coding for procedural identification.

7.3.3 High and low Mortality group selection

Units were stratified according to overall risk-adjusted mortality (see risk-adjustment section below). A cut-off of 5% was chosen as an arbitrary threshold for combined mortality of a unit's oesophageal and gastric 30-day mortality rate. This figure was chosen as high volume centers worldwide reportedly reproduce such rates for patients undergoing elective resections for cancer (Allum et al., 2002, Siewert et al., 1998, Al-Sarira et al., 2007, Sano et al., 2004). Units with an adjusted mortality of 5% and lower were termed low mortality units (LMU) and those units with a higher adjusted mortality (>5%) were termed high mortality units (HMU).

7.3.4 Complications / Re-operations

Serious surgical complications that necessitated re-operations on the index admission were termed re-operations. Thus re-operations in this context can be thought of as a surrogate for serious surgical complications.

Patients that emergently underwent one or more of a panel of operative procedures subsequent to the index procedure during the primary admission were deemed to have undergone a re-operation. The re-operative procedures were selected as those that most likely reflect serious surgical complications rather than planned "re-looks". The re-operations broadly fall under three categories - thoracotomy, laparotomy and laparoscopy. Examples of indications for re-operation include bleeding, organ space infections and bowel obstruction. Common procedures included under the heading 'laparotomy' were drainage of abscess/collection, bowel resection +/- stoma formation and washout of abdominal cavity.

7.3.5 Non-operative re- interventions (radiological / endoscopic)

Patients that did not undergo a re-operation that underwent a radiologically guided percutaneous drainage procedure or an upper gastrointestinal endoscopic procedure post-operatively were identified and termed a non-operative re-interventions. Procedures before midnight on the day of the index operation were excluded as these may have been planned procedures that were undertaken as a part of the primary resection. Reasons for undertaking these re-interventions are not discernible from the database. However radiological guided drains and endoscopies performed after the day of surgery are likely to reflect a surgeon's desire/threshold to investigate any deviations from a 'normal' recovery.

7.3.6 Outcome measures

The outcome measures were-

1. Mortality following re-operations (a surrogate for surgical complications)- and,
2. Mortality following non-operative re-intervention (endoscopic and radiological),

7.3.7 Outcome measures and the modified Clavien-Dindo surgical complication classification

The validated modified Clavien-Dindo classification of surgical complications stratifies surgical complications according to seven grades, I to V, with two sub grades for grades III and IV (IIIa, IIIb and IVa, IVb) (Dindo et al., 2004). For this study grade IIIa and IIIb complications were of specific interest. The primary outcome measure were Failure to Rescue after non-operative re-intervention and would represent outcome following Clavien-Dindo grade IIIa complications and FTR-S would represent outcome following Clavien-Dindo grade IIIb complications. Of note the original failure to rescue definition would have considered deaths following Clavien-Dindo grades I-IVb complications together see Table 16.

Table 16- Modified Clavien-Dindo grading system of surgical complications and how they relate to the study outcome measures.

Clavien-Dindo Grade	Definition	Study outcome measures
I	Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic and radiological interventions.	
II	Requiring pharmacological treatment with drugs other than such allowed for grade I complications. Blood transfusions and total parenteral nutrition are also included.	
IIIa	Surgical, endoscopic or radiological intervention not under general anaesthetic	Failure to Rescue following non-operative intervention
IIIb	intervention under general anaesthetic	Failure to Rescue- Surgical (FTR-S)
IVa	Single organ dysfunction (including dialysis)	
IVb	Multi organ dysfunction	
V	Death of a patient	

7.3.8 Statistical Analysis

Risk adjustment

Risk adjustment was carried out by creating a multiple regression model to predict the likelihood of binary outcomes with covariates gender, patient age (considered in bands <60, 60-70, 71-80, >80 years), Carstairs Index (Taylor et al., 2003), Charlson comorbidity score (grouped as those with scores of <2 and those with scores of >=2 where latter group indicates more co-morbid patients) and the type of procedures performed (oesophagectomy or gastrectomies). Factors with a significance level of ≤0·1 on bivariate analysis were included in the multiple regression analyses. Unit-level adjusted death rates were obtained for each hospital by dividing the hospital's observed deaths by its model-predicted deaths and multiplying by the national crude death rate. Statistical analyses were carried out with SPSS version 18·0 (SPSS, Chicago, Illinois, USA). All p-values stated refer to two-sided values.

7.4 Results

7.4.1 Whole cohort patient demographics

A total of 25,626 patients were electively admitted to 141 National Health Service (NHS) units in England over the 10 year period and included in the study. The demographics of the patients are shown in Table 17 as are the odds ratios of 30-day mortality as independently predicted by the considered covariates. Patients aged >80 years, more socially deprived patients (higher Carstairs index) and patients with more comorbidities (Charlson score >2) were more likely to die in hospital within 30 days (respective Odds ratios - 5·00, 1·36, 1·84, p<0·05 for all- see Table 17Table 17.

Table 17- Final multiple regression model of independent predictors and the relative odds risk of death in hospital at 30 days for the whole study cohort

Covariate	Number of patients n=25,626 (%)	Odd ratio of 30-day mortality (95% C.I)	Significance (p-value)
Resection Type			
Oesophageal	14955 (58·4)	1·00	<0·001
Gastric	10671 (41·6)	0·78 (0·70, 0·88)	
Age			
<60	6450 (25·2)	1.00	
60-70	9402 (36·7)	1.75 (1·45, 2·10)	<0·001
71-80	8165 (31·9)	3.28 (2·74, 3·93)	
>80	1609 (6·3)	5.00 (3·96, 6·33)	

Gender Male Female	 18350 (71·6) 7276 (28·4)	 1·00 0·90 (0·79, 1·02)	 0·097
Carstairs Index (1=less deprived) 1 2 3 4 5 6	 4487 (17·5) 5610 (21·9) 5633 (22·0) 5157 (20·1) 4644 (18·1) 95 (0·4)	 1.00 1.02 (0·84, 1·23) 1.03 (0·86, 1·25) 1.23 (1·02, 1·48) 1.36 (1·13, 1·65) 0.86 (0·31, 2·39)	 0·002
Charlson Score (≤2=less comorbid) ≤2 >2	 15141 (59·1) 10485 (40·9)	 1·00 1·84 (1·65, 2·06)	 <0·001

7.4.2 Mortality rates

There were 1348 deaths in the whole cohort giving an overall crude death rate of 5·3% for oesophageal and gastric resections combined. The crude death rates for patients undergoing oesophagectomy were 5·1% (758/14955) and 5·5% (590/10671) for those undergoing gastrectomies respectively.

7.4.3 By mortality group (LMU versus HMU)

When the units are stratified using a risk adjusted mortality threshold of 5%, 65 units with 11,803 (46·1%) patients appear in the LMU group (≤5% adjusted mortality) with 13,823 (53·9%) patients in the HMU group (>5% adjusted mortality).

7.4.4 Patient demographics

There were no statistical differences between the two groups in terms of the type of resection undertaken (oesophageal or gastric, p=0·198), patients' presenting age (p=0·765) or patients' gender (p=0·341). There were relatively more socially deprived patients (indicated by a higher Carstairs index) in the HMU than the LMU (p<0·001). There were also more co-morbid (indicated by higher Charlson score) in the LMU (p<0·001). More patients underwent minimally invasive approach (MIA) procedures in the LMU (985 versus 792, p<0·001) see Table 18.

Table 18- Demographics between the mortality groups

	LMU n=11803	HMU n=13823	p value
Resection Type Oesophageal Gastric	 6922 (27·0) 4881 (19·0)	 8033 (31·3) 5790 (22·6)	 0·198
Age <60 60-70 71-80 >80	 2999 (11·7) 4306 (16·8) 3769 (14·7) 729 (2·8)	 3451 (13·5) 5096 (19·9) 4396 (17·2) 880 (3·4)	0·765
Gender Male Female	 8467 (71·7) 3336 (28·3)	 9883 (71·5) 3940 (28·5)	 0·341
Carstairs Index (1=less deprived) 1 2 3 4 5 (most deprived) 6 (unknown)	 2039 (8·0) 2510 (9·8) 2567 (10·0) 2357 (9·2) 2237 (8·7) 93 (0·4)	 2448 (9.6) 3100 (12·1) 3066 (12·0) 2800 (10·9) 2407 (9·4) 2 (0·0)	<0·001*
Charlson Score (≤2=less comorbid) ≤2 >2	 6673 (56·5) 5130 (43·5)	 8468 (61·3) 5355 (38·7)	 <0·001*
Use of MIA	985 (8·3)	792 (5·7)	<0·001*
Teaching hospital status	10/65 (15.4)	14/76 (18.4)	0.402

7.4.5 Re-intervention and mortality rates

LMU and HMU had equivalent re-operation rates for surgical complications (thoracotomy and laparotomy combined) rates (5·4% versus 4·9%, p=0.105). LMU and HMU had significantly different non-operative re-intervention rates with the LMU performing more non-operative re-interventions (6·7% versus 4·7%, p<0·001).

Patients were significantly more likely to die after re-operations in HMU units than in LMU units - what is termed failure to rescue-surgical: FTR-S (24·1% versus 15·3%, p<0·001). Patients are more likely to die after a non-operative re-intervention in the HMU (12·5% versus 7·0% p<0·001).

7.4.6 Limitations

The study is based upon an administrative data source. Any coding errors may potentially influence the results. Systemic under-reporting of re-interventions by disparate organisations within the same mortality grouping over the study periods could in theory lead to reporting bias. This is however unlikely. More general coding error could influence the results. However, this is unlikely given the proven accuracy of this dataset in recording diagnostic fields (Campbell et al., 2001, Burns et al., 2011c). Unfortunately, HES does not capture cancer stage (though the Charlson index includes metastases which was included in the adjustment model) and this may well have an influence on perioperative outcome, as will case-mix not fully adjusted for using the available data. However, it is unlikely that patients with stage IV disease would be undergoing major surgery, unlike in colorectal cancer where some patients with obstructing colon cancer are stented as a bridge to elective (albeit expedited) surgery.

The strengths of the study are that it is not subject to reporting bias. HES has been shown to record more deaths when compared with a voluntarily recorded clinical registry over similar time periods (Almoudaris et al., 2011a). Given the sample size and number of years considered, this study truly reflects the national outcome from oesophago-gastric cancer surgery.

7.5 Conclusion

It has been shown, using English national administrative data when units undertaking oesophageal and gastric cancer resections are grouped according to adjusted mortality, index admission re-operation rates for complications are equivalent between low and high mortality units. Low mortality units are, however, more likely non-operatively to intervene and are more likely to rescue patients from subsequent death after both re-operative and non-reoperative re-interventions.

What this study reiterates is that variability in outcome following management of serious complications does also occur in patients undergoing upper gastrointestinal resection. This is in keeping with the current literature in other specialties (Ghaferi et al., 2009a). Our study also demonstrates that Failure to Rescue-Surgical rates differ by mortality grouping as in lower gastrointestinal surgery (Almoudaris et al., 2011b).

There is variability in outcome following serious surgical complications requiring re-operations between units undertaking oesophago-gastric cancer surgery in England. Units with lower overall mortality re-intervene more often and are subsequently more successful. Future work should focus on why such variability occurs and identify methods for mitigating this variability. This study has shown that non-operative re-interventions are feasible to derive from the HES dataset.

8.0 BENCHMARKING COLORECTAL CANCER RESECTIONAL UNITS IN ENGLAND ON A PANEL OF METRICS USING HOSPITAL EPISODES STATISTICS

8.1 Chapter Overview

In this chapter an attempt is made to appraise the performance of all units in England undertaking colorectal cancer resections. This will be performed using a panel of metrics. These metrics will include those directly calculated from HES (e.g LOS, readmission within 28 days) as well as those sourced using external references (e.g death within 30 days- via linkage with ONS) as well as those that have been derived during the course of this thesis (e.g. FTR-S).

8.2 Introduction

In the United Kingdom comprehensive and mandatory healthcare data collection is performed routinely (The Information Centre (England), 2011, National Services Scotland, 2011, Department of Health Social Services and Public Safety (Northern Ireland), 2011). Measures of surgical quality are readily available from such administrative data sources. The utility of these measures depends upon their use to define quality and influence decision-making at a clinical, managerial or policy level. Several metrics relevant to colorectal surgery are derivable from routinely held data. These could potentially be used to benchmark performance in colorectal surgery. If this process were reliable it could inform broadly on surgeon-specific and institutional colorectal surgical performance. Moreover, if reliance upon existing National Health Service (NHS) data sources (that lie within the public domain) were maintained for these purposes, transparent reporting of outcome to the public would follow. For such a system to be fair and robust two conditions are pre-requisites. Firstly, the accuracy of data used for benchmarking must be consistent at an institutional and surgeon level. In addition, an understanding of how individual metrics inter-relate to reflect high and poor surgical performance is needed. High achievement across all measured domains almost certainly reflects a proactive and competent provider. Secondly, can one however meaningfully comment upon provider performance from measurement of one domain alone? If so, what limits are meaningful? Alternatively, do these metrics reflect unique aspects of performance and demand individual appraisal (and remedial intervention)?

Validated metrics may be used to benchmark performance between surgical providers and to underpin quality improvement initiatives. Surgical measures that are easily obtained from routinely collected data include: 30-day mortality (Dimick et al., 2010b), in-patient length of stay (McPherson, 1984), and re-admission rates (Chambers and Clarke, 1990). Additional markers of performance that are of relevance to colorectal surgery can also be derived from such data sources including short-term re-intervention (Merkow et al., 2009b) and Abdomino-Perineal Excision (APE) rate (Morris et al., 2008). Other metrics, such as lymph node yield, R0 resection rate and quality of mesorectal excision, may be obtained from clinical registries (Association of Coloproctology of Great Britain and Ireland, 2011). The latter may also be used to evaluate service quality between providers (Jeyarajah et al., 2007). The aforementioned measures potentially form the basis for future quality improvement programmes (Lindenauer et al., 2007).

A publication by Morris and colleagues used administrative data linked to cancer registry information to report on the variation that currently exists in 30-day mortality rates between English NHS institutions undertaking colorectal surgery (Morris et al., 2011). Performance concerns, with regards to short-term survival outcome, may be justifiable in a limited number of outlying institutions. Clarification that poor peri-operative mortality rates denote poor global standards of colorectal surgical practice (e.g. associated high re-intervention rates, poor oncological outcomes etc.) clearly justifies public reporting of single performance measures. As such, it is necessary to elucidate what it means to be an institutional outlier for 30-day mortality?. Lastly, an understanding of the limitations of specific outcome measures is becoming increasingly significant to individual surgical practitioners in the United Kingdom due to the implementation of compulsory revalidation of doctors (General Medical Council, 2011). Furthermore, as previously mentioned, there has been a drive publicly to report surgeons' outcome data from nationally collected data. At present the Association of Coloproctology of Great Britain and Ireland (ASGBI) is openly reporting unit and individual surgeons' outcome data following colorectal cancer resections. The outcome measure appraised is 90 day mortality following resection for 2012-13 (http://www.acpgbi.org.uk/surgeon-outcomes/search) from data submitted by individual surgeons via their hospitals' own registration to a national clinical audit program (https://clinicalaudit.hscic.gov.uk/nboca).

The primary aim of this study was to explore from an English national administrative database the relationships between commonly collected and derivable metrics. A secondary aim is to identify whether units can be appraised on a panel of metrics that may be more informative than just the reporting of one. Specifically, national data was used to correlate institutional 30-day mortality rate with other outcome metrics. In addition, the performance of statistical 'outliers' for 30-day mortality across other quality domains at an institutional level was analysed.

8.3 Methods

Hospital Episodes Statistics (HES) database and patient selection

The HES database is an administrative dataset to which all National Health Service (NHS) hospitals compulsorily submit patient level information and has been described previously.

All patients that underwent a primary major colorectal procedure with a diagnosis of colorectal cancer between April 2000 and March 2008 in English NHS Trusts were included. Patients were identified using diagnostic and procedural codes from the relevant International Classification of Disease 10th revision (ICD-10) and Office for Population Census and Surveys Classification of Surgical Operations and Procedures 4th Revision (OPCS-4) codes on the HES database. A detailed methodology of this process has been described previously (Faiz et al., 2010c). The following resections were analysed - right and extended right hemicolectomies, transverse colectomy, left hemicolectomy, sigmoid colectomy, Hartmann's procedure, subtotal colectomy, panproctocolectomy, total colectomy, anterior resection (AR) and abdominoperineal resection (APER). The corresponding OPCS-4 procedural codes are described previously.

The Charlson co-morbidity scoring system was developed for administrative datasets (Charlson et al., 1987). Co-morbidities that are associated with worse outcomes are given greater scores. Secondary diagnosis fields on HES were used to create the Charlson co-morbidity index. Charlson score was re-classified into three categories 0, 1-4 and ≥5. The Carstairs index (Morgan and Baker, 2006) is a composite socio-economic deprivation score calculated at the output-area level and converted into population-weighted quintiles (Morgan and Baker, 2006).

8.3.1 Outcome metrics

i) *Length of stay, re-admission within 28 days and mortality within 30 days*

Institutional lengths of admission stay for the above procedures (taken as a basket) were described as mean values following logarithmic conversion as the percentages of patients that had lengths of stay greater than the population 75th centile. 28-day re-admission and 30-day mortality rates were expressed as proportions (in percentages) of the total caseload.

ii) *Reoperation rates*

Reoperation rates are computed from HES data employing a methodology that has been described previously. Reoperation describes any patient returned to theatre after their index procedure for a select group of interventions within 28-days. The codes for reoperations include those denoting: washout of abdomen, small bowel resection, further colorectal resection, drainage of intra-abdominal abscess, division of adhesions, stoma formation or operation on a stoma and wound complications requiring return to theatre. Reoperation rates are calculated as a proportion of the total volume of index procedures undertaken.

iii) *Abdomino-Perineal Excision (APE) rate*

Where the primary diagnosis was rectal cancer, the APE rate was calculated as the number of non-emergency APE resections performed over the total volume of other procedures performed for excision of a rectal cancer. Other procedures included Anterior Resections, Hartmann's resections, excision of rectum unspecified/other, panproctocolectomy, total colectomy, sigmoid colectomy and excision of left hemicolon. These were converted into a percentage. OPCS-4 and ICD-10 codes used are previously described.

iv) *Failure to Rescue – Surgical (FTR-S) rate*

FTR-S rate is defined as the proportion of patients that die on their index admission after being returned to theatre (for the procedures listed above). The methodology used to calculate FTR-S has been described in Chapter 6 (Almoudaris, 2011).

8.3.2 Statistics and funnel plots

Outcome rates were calculated per institution for correlation and comparison purposes. Correlation between linearly distributed outcome variables was investigated using Pearson's statistic. Length of stay required logarithmic transformation before application of linear statistical methods.

Adjustment was carried out using multiple regression analyses. Models incorporated covariates including patient gender, age, Charlson co-morbidity score, Carstairs deprivation index, type of resection and method of admission - elective/emergency. These were aggregated on a per institution level for each metric considered and used to create case-mix adjusted funnel plots for each dependent variable. Funnel plots (Spiegelhalter, 2005) were created using the tools available at http://www.erpho.org.uk/topics/tools/funnel.aspx. The control limits displayed are the exact Poisson control limits.

8.4 Results

8.4.1 Patient Population

144,542 patients were analysed that had undergone a primary major colorectal cancer resection between April 2000 and March 2008 in 149 NHS units. Patient demographics are described in Table 19.

Table 19- Demographics of the patients included in the study

	Primary major colorectal resections 2000-2008 n=144,542 (%)
Age bands	
17-54	13,705 (9.5)
55-69	48,600 (33.6)
70-79	51,598 (35.7)
>79	30,639 (21.2)
Sex	
Male	80,400 (55.6)
Female	64,142 (44.4)
Year of Procedure	
2000	18004 (12.46)
2001	17290 (11.95)
2002	17692 (12.24)
2003	17488 (12.10)
2004	17751 (12.28)
2005	18837 (13.03)
2006	18511 (12.81)
2007/8	18969 (13.13)
Admission method	
Elective	111,037 (76.8)
Emergency	33,505 (23.2)
Charlson	
0-1	82,675 (57.2)
2-4	12,598 (8.7)
>5	49,269 (34.1)

Carstairs	
1 (least deprived)	28,141 (19.5)
2	33,338 (23.1)
3	31,917 (22.1)
4	28,135 (19.5)
5 (most deprived)	22,911 (15.8)
Unclassified	100 (0.1)
Colectomy / resection type	
Left sided *	33,354 (23.1)
Right sided	50,804 (35.1)
Subtotal / total	6,174 (4.3)
Rectal	54,210 (37.5)
Method	
Laparoscopic	6,380 (4.4)
Open	138,162 (95.6)

A funnel plot describing risk adjusted in-hospital 30-day mortality rate for all 149 NHS institutions was charted (Figure 9). The funnel plot depicts the adjusted upper and lower 2nd and 3rd standard deviation (SD) control limits for varying caseload. Units were described as outliers if they lay above or below the respective 3SD control (99.8%) limits. Units were described as lying within acceptable limits if they lay below the upper 2SD control (95.0%) limit.

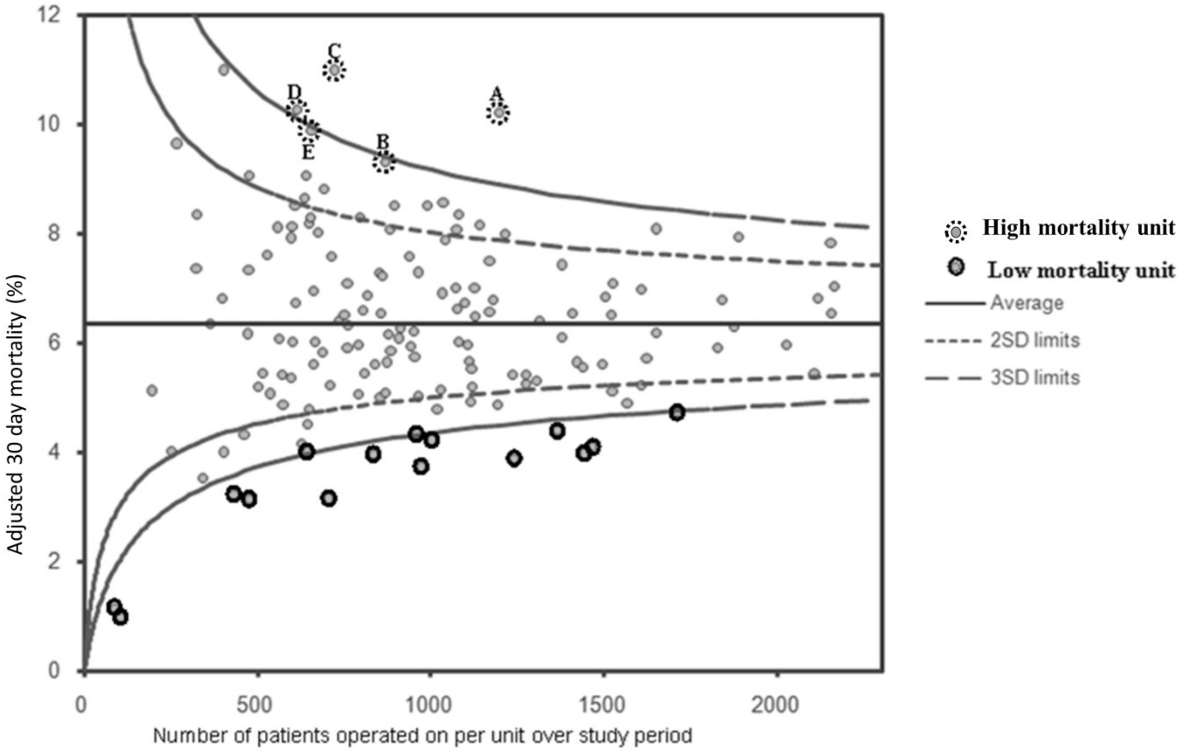


Figure 9- Funnel plot demonstrating the risk adjusted 30 day mortality of all units undertaking colorectal cancer resections in England

The funnel plot highlighted five institutions whose mortality rates lay on, or above, the upper 3rd SD control limit (i.e. have significantly higher than expected mortality rates at the 99.8% confidence level). These institutions were termed High Mortality Outlier (HMO) units. Fifteen units were identified below or on the lower 3rd SD control limit (i.e. significantly lower mortality than expected at the 99.8% confidence interval). These were described as Low Mortality Outlier (LMO) units (Figure 9).

8.4.2 HMO units

All five HMO units lay within acceptable limits for re-admissions and APE rates (Figure 10 and Figure 11) when identified on case-mix adjusted funnel plots. All but one unit (institution D) lay between the control limits for length of stay (Figure 12).

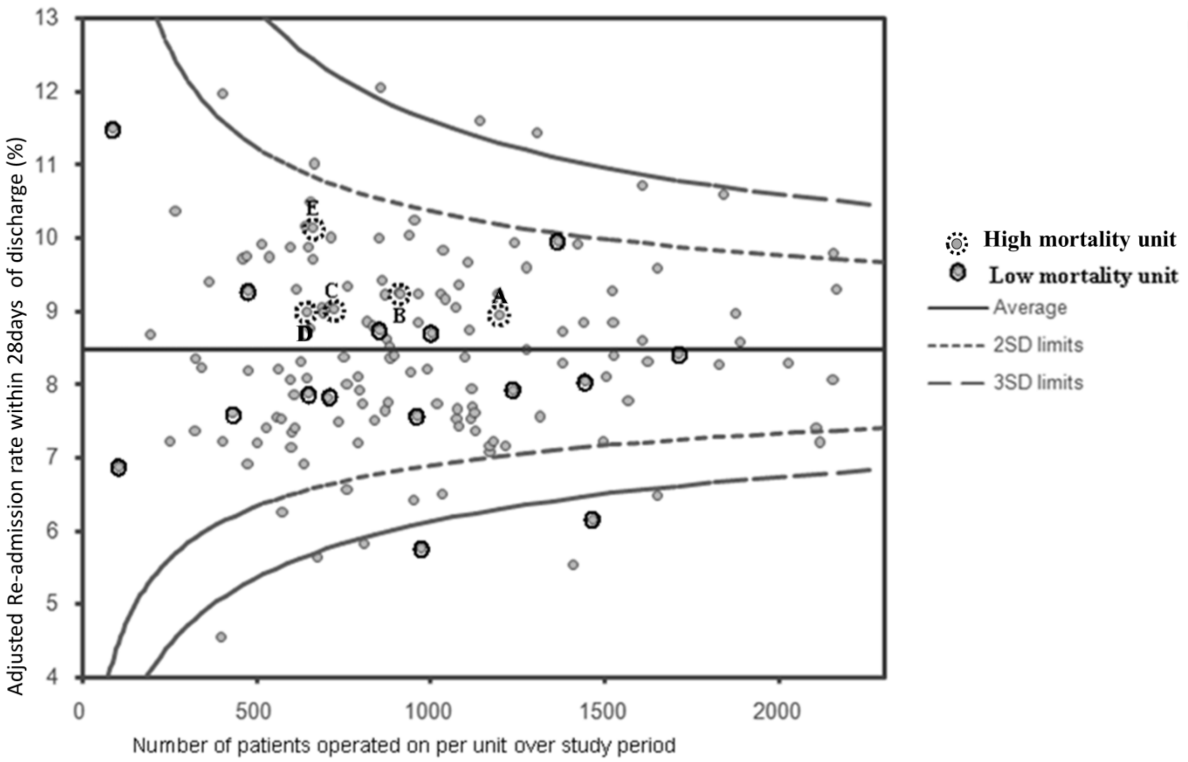


Figure 10- Funnel plot of unit's re-admission rates within 28 days

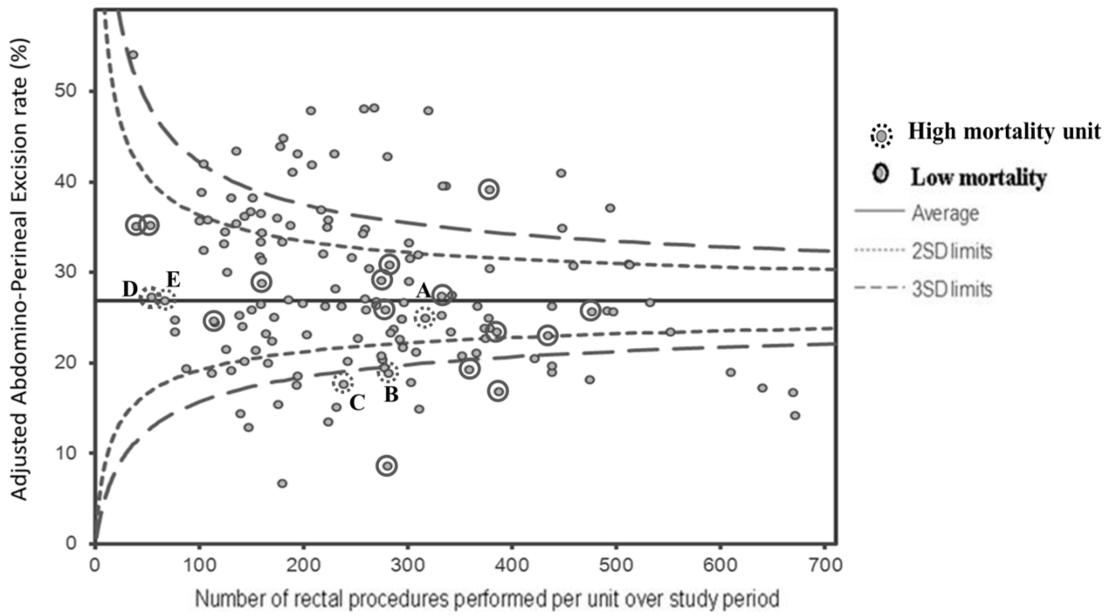


Figure 11- Funnel plot of adjusted APE resection rates

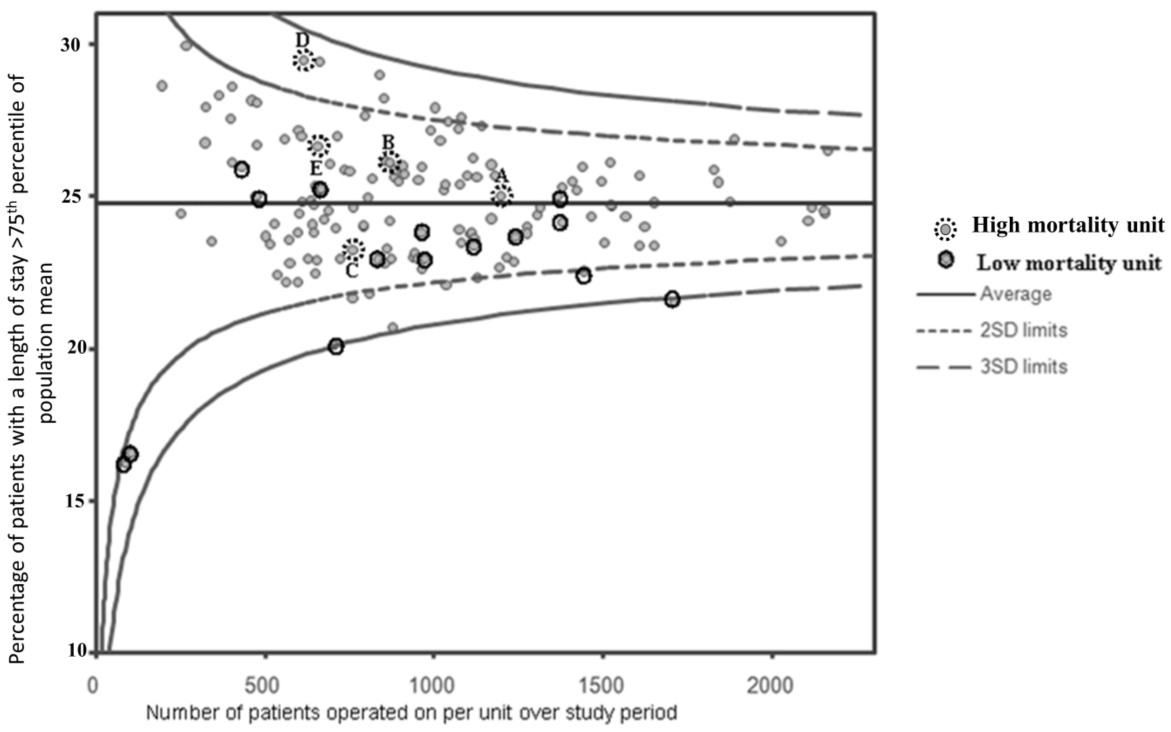


Figure 12- Funnel plot of percentage of patients with lengths of stay greater than the 75th percentile

When the HMO units A-E were charted on a case-mix adjusted funnel plot for reoperation rates, unit D lay within acceptable limits whereas unit C lay below the lower 3rd SD control limit (i.e. lower than expected reoperation rates). In contrast, units A, B and E all demonstrated higher than expected reoperation rates- Figure 13.

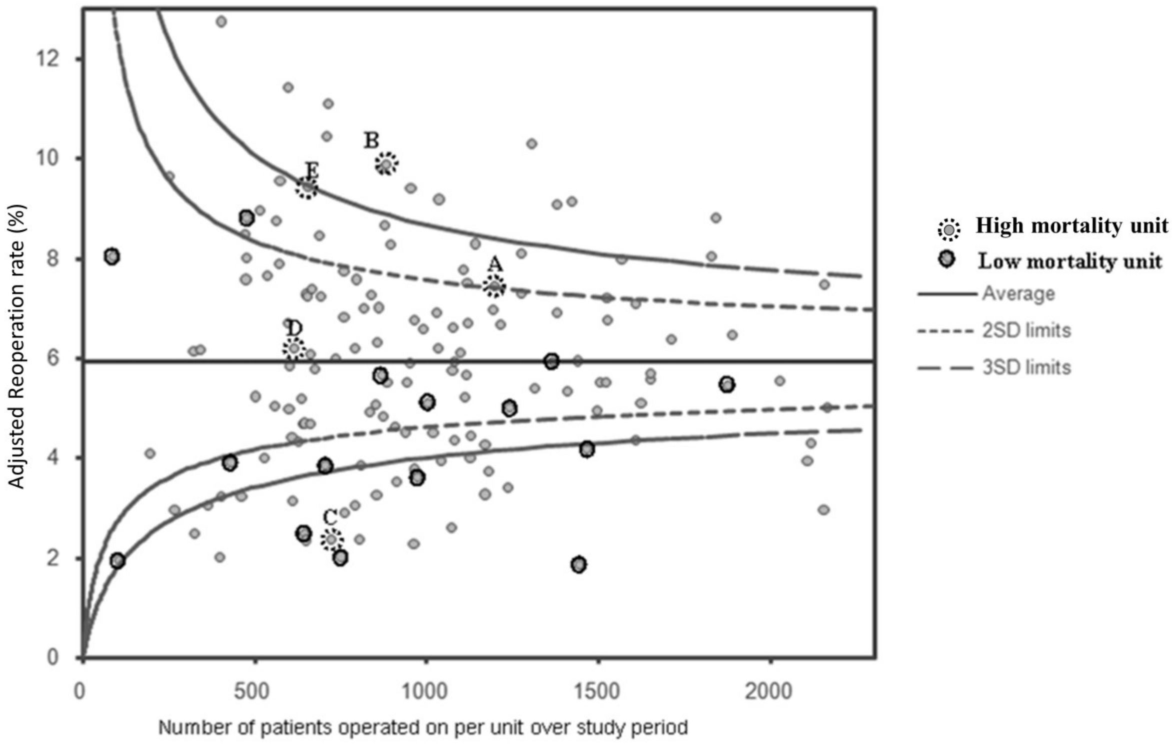


Figure 13- Funnel plot of units adjusted re-operation rates

A funnel plot describing case-mix adjusted institutional FTR-S rates demonstrated that units A, D and E lay within the acceptable limits (Figure 14). Unit C demonstrated a significantly lower FTR-S rate than expected. In contrast unit B lay above the upper 2nd SD control limit indicating that a greater number of patients at the institution failed to be rescued.

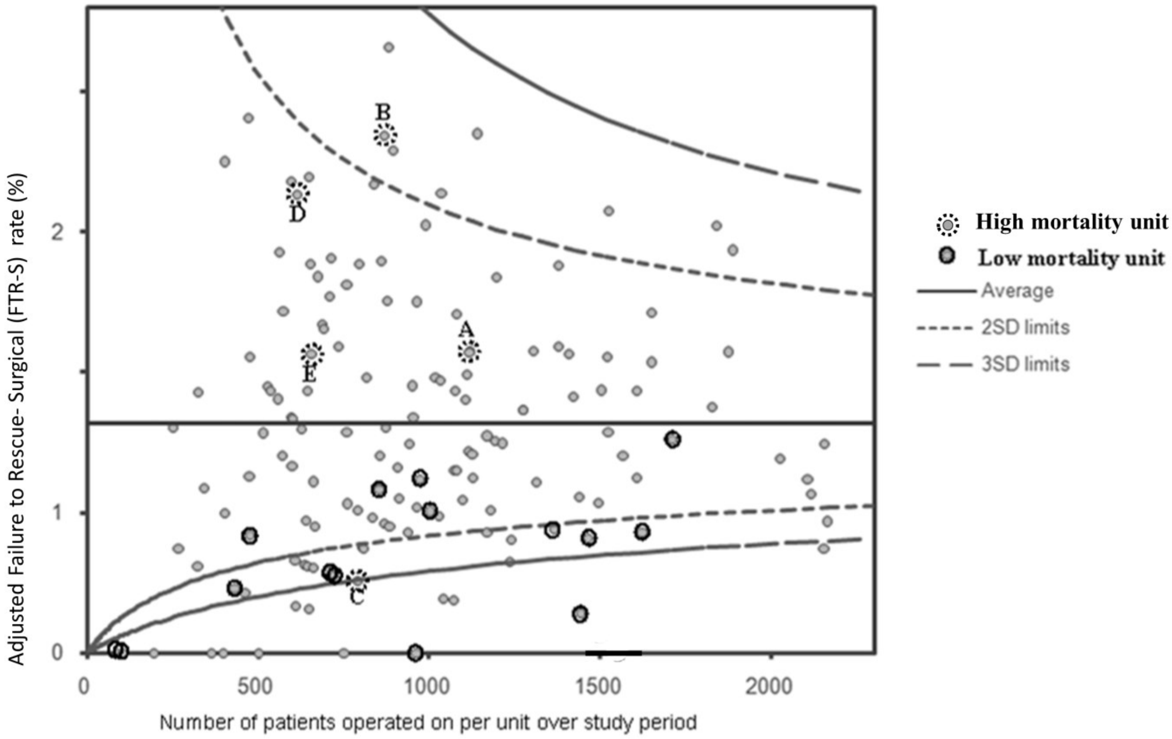


Figure 14- Funnel plot of adjusted Failure to Rescue- Surgical rates

Table 20 summarises the performance of the HMO units and how they performed on the other considered metrics in a simple matrix.

Table 20- Summary table of how the High mortality outlier units performed on other measures

High Mortality Outlier Unit	Reoperation	FTR-S	Re-admission rates	Length of stay	APER rate
A	HE	-	-	-	-
B	HO	HE	-	-	-
C	LO	LO	-	-	-
D	-	-	-	HE	-
E	HO	-	-	-	-

HO=High Outlier (lying above upper 3rd s.d control limit),
LO= Low Outlier (lying below lower 3rd s.d control limit),
HE=Higher than expected (lying above upper 2nd s.d. control limit),
LE=Lower than expected (lying below lower 2nd s.d control limit),
'-' indicates performance within the expected range.

8.4.3 LMO units

13/15 LMO units performed within acceptable limits with several performing better than expected when charted on case-mix adjusted funnel plots for re-admission rate, length of stay and FTR-S (Figure 10, Figure 12, Figure 14) . When reoperation rate was charted one LMO unit was observed to perform less well than expected lying above the 2nd s.d. control limit. The remaining 14 units performed as well as, or better than, expected (Figure 13). One LMO unit lay above the upper 3rd SD control limit for APER rates (Figure 11).

8.4.4 Institutional outcome metric correlation

Correlations between institutional 30-day postoperative mortality rate and other outcome measures are described in Table 21. At institutional level, when 30 day mortality is correlated against FTR-S rates a significant correlation is observed (R=0.445, p=<0.001). When institutional 30-day postoperative mortality rate is correlated against reoperation rate, a weak statistical correlation is observed (R=0.191, p=0.020). Mortality did not correlate statistically with: 28 day re-admission rates (R=0.143, p=0.082), APER (R=0.119, p=0.147) or length of stay (R=0.148, p=0.072).

Table 21- Table demonstrating individual Pearson's correlation statistics and significance values per metric against 30d mortality (adjusted)

		Return to theatre rate	FTR-S rate	28d readmission	Length of stay	APER rate
30-day Mortality	R	**0.191**	**0.445**	0.143	0.148	0.119
	p-value	**0.020**	**<0.001**	0.082	0.072	0.147

Significant correlation at p<0.05 have been highlighted in bold
R= Pearsons correlation R value, sig.=two tailed significance value.

8.5 Conclusion

The current study suggests that high institutional postoperative mortality rate following colorectal surgery does not necessarily predict how such units perform on other measures of service quality. Benchmarking institutional colorectal surgical performance is complex and not generalisable from a single measure of outcome, but rather demands global service appraisal across a range of outcome measures including those that are novel and derivable from HES.

However whilst appraisal using the currently available data has been shown to be feasible, Lilford's intervening variables of more subtle markers of institutional performance are recalled. Such analysis has not taken such factors into account. What follows is an attempt to construct a novel methodology and to pilot it, in an attempt to discern whether a truly global perspective of a surgical unit's performance is possible.

9.0 DEVELOPING A TOOL TO INVESTIGATE DRIVERS OF INSTITUTIONAL QUALITY IN COLORECTAL SURGERY

9.1 Chapter Overview

This study attempts to ascertain whether it's possible to identify factors that are not discernible from appraisal of databases for the assessment of high performance in surgery using a novel methodology. This study has been termed the High Performance (HiPer) methodology study. In essence this chapter attempts to identify "what lies behind the numbers". Can a tool/methodology be devised that appraises aspects of care such as team-work and institutional culture, which can then be related to hard-outcome data?

9.2 Introduction

High performing organisations exert an understandable fascination for everyone concerned with the safety and quality of healthcare and the often inconsistent and unreliable nature of much of the care actually delivered. In healthcare the original studies by the Berkeley group of nuclear power, naval aviation and air traffic control have been very influential and inspired much commentary and interpretation (La Porte, 1996). These High Reliability Organisations (HROs) are those which, in the face of considerable hazards and operational complexity, manage to achieve high levels of both safety and performance (Hofmann et al., 1995). HROs are frequently referenced as models to which healthcare should aspire, particularly because their environment and challenges seem to have much in common with the dynamic and uncertain healthcare environment (La Porte, 1996, Roberts et al., 2005).

Studies of high reliability organisations, whether in healthcare or other industries, have tended to be largely descriptive in nature and present a number of difficulties for those wanting guidance on the best route to high performance (Vincent et al., 2010). The original studies and later interpretations drew attention to a very wide range of characteristics said to be important to reliable performance and the range of alleged high reliability concepts is now enormous. More recent studies of high performing systems have greatly clarified the key elements underlying high performance and introduced some welcome conceptual clarity (Baker, 2008). Even so, the field has remained resolutely descriptive with few attempts to measure high reliability characteristics or relate them to substantive clinical outcomes. The challenge now is to take this rather diffuse set of ideas, refine them, and test them in a healthcare context. This requires finding a robust and efficient method of assessing relevant clinical and organizational characteristics.

Both researchers and regulators have struggled to find an effective way of assessing a hospital or unit's clinical and organisational processes. Self-assessment, as used by many regulatory agencies, while useful in that context, is obviously subject to numerous potential biases. Survey data suffer from low response rates and are unlikely to enable us to discern the nuances of patient care. Site visits, while attractive in the depth of understanding that might be gained, are expensive, difficult to arrange and potentially disruptive influences in already stretched organisations (Vincent, 2013). However studies in the commercial sector suggest that structured telephone interviewing can provide a reliable means of assessing organisational characteristics. This method of assessing organisational performance has been found to be remarkably effective, with high reliability obtained between interviewers, robust scoring and validity demonstrated by the association of findings with objective data (Bloom and Van Reenen, 2007). This management measure has been found to correlate with general financial and operational outcomes in both healthcare and industry.

The aim is to describe the development of the High Performance (HiPer) methodology, a structured method of assessing clinical and organizational characteristics of clinical systems producing quantitative data which subsequently, may, be linked to clinical outcomes.

9.3 Methods

9.3.1 The evolution of the HiPer Methodology

The methodology was a planned stepwise progression. There were 4 main steps to the creation of the methodology.

9.3.2 Step 1 -Systematic review

A comprehensive review entitled "Establishing Quality in Colorectal Surgery" was initially undertaken (Almoudaris et al., 2011c). The purpose was to identify salient structural, process and outcome factors that pertain to high performance and quality in colorectal surgery from contemporary literature. The review was used as the basis of identifying important factors and themes that needed to be considered in appraising and understanding how high performance and clinical quality are achieved in a real-world setting.

9.3.3 Step 2- Expert consensus

The purpose of this stage was to develop consensus as to which themes, from those identified in the systematic review, should be used to populate the final tool. The specific brief the experts were given was "What aspects of the surgical care pathway do you think should be further investigated in a study of high performing colorectal surgical units?". The 5 experts were offered all of the themes from the review and were allowed to suggest others. The experts were also encouraged to explain why they thought selection of the specific facet of care was important. This subsequently aided and guided development of the individual questions. Experts were independently practicing colorectal clinicians of Consultant grade. Each had published over 50 peer reviewed articles. The experts were chosen from around the Western world and reflected the continents of the final participating units. The process was undertaken via email with all responses sent to back to me. I co-ordinated the process and collated the final responses.

9.3.4 Step 3- Development of semi-structured interview protocol

Once a list of 9 themes had been reached with consensus agreement the development of the semi-structured interview began. These were preceded by one very broad, open ended theme/question to give a total of 10 questions. Questions were estimated each to take 4 minutes to discuss. This resulted in a standard interview length of 40 minutes. For each of the identified themes from the expert consensus exercise, a question was derived with multiple stems. The questions were open ended and purposefully non-specific at times to encourage participants to discuss factors that they felt were important rather than be guided by the question. Alongside the interview protocol a marking schedule was also developed. For each of the 9 marked questions, a 7 point Likert scale was created. For each point an indicative response to attain that mark was given. The intention was that a score of 1 would represent detrimental care and a score of 7 would represent exceptional care. The questionnaires would be delivered by trained interviewers.

9.3.5 Step 4- Interviewer training

A bank of interviewers was selected and trained in interview techniques. The interviewees were also given the brief of the project. All were senior colorectal surgical trainees undertaking higher research degrees. Interviewers were given the interview schedule and the mark scheme one month before their first scheduled interview to assimilate. The interviewers were each assigned units and arranged the telephone interviews via email. The interviewers would score the response in real-time for each question. The interviewers were also encouraged to make free text notes as well.

9.3.6 Choice of participating units and subjects

9.3.6.1 Participating units- the Global Comparators Project

Collaboration with the Global Comparators initiative run by Dr Foster Intelligence (http://globalcomparators.com/) for piloting this methodology was chosen. Dr Foster Intelligence has brought leading global academic hospitals together which have individually submitted patient level data. The aim of the project was to share best practice, collective learning to improve cost-effectiveness, efficiency and ultimately patient outcome. The group consists of large academic teaching hospitals from mainland Europe, England and the United States. At the time of the study there were over 30 leading academic hospitals involved (more have joined since). On joining the project each hospital submits its patient level outcome data on a range of pathologies including colorectal surgery to Dr Foster Intelligence. Units' data are then compared by pathology and risk-adjusted. The overview of the process and exact methodology has been published by the Global Comparators group (Bottle et al., 2013).

All participating units undertook the necessary ethical and board clearances including Institutional Review Board (IRB) for the participating units in the United States.

Specifically, the main reason for choosing the Global Comparators for the source of data comparisons rather than continuing to use HES data for institutions in England alone pertains to data governance issues of using HES data. The data governance that bound access to HES data meant that it was not possible to identify units specifically by name. The inclusion of English units in HiPer was made possible as the unit's individually and independently submitted data outside the HES submission pathway and directly to Dr Foster Intelligence hence allowing for their inclusion. Without using this collaboration it would not have been possible to compare any results from the HiPer study with subsequent outcome - one of the final aims of the thesis. All participating units in the Global Comparators group were invited to take part in piloting of the HiPer tool. The final units that completed the necessary ethics clearance and replied to the request to participate in the pilot study are given in Table 22.

Unit Name	Continent
AMC Amsterdam	Europe
Chelsea and Westminster Hospital	
Royal United Hospital Bath NHS Trust	
University Hospitals of Leicester NHS Trust	
University Hospital of South Manchester	
Uzleuven Belgium	
Hospital of the University of Pennsylvania	USA
UC San Diego Medical Center	
UT Southwestern Medical Center	

Table 22 Table of the final units included in the HiPer study

9.3.6.2 Selection of interviewees within each participating unit

It was decided to interview one colorectal surgeon, a ward nurse that cares for colorectal patients, an anaesthetist/intensivist and a manager with direct responsibility for colorectal surgical patients/department. The aim was to give a balanced view of the day to day workings of the department. By gaining an insight from different respondents it allowed for triangulation of responses where relevant. The term manager in this context related to any person with responsibilities for managing the day to day operation of the unit as well as financial and staffing responsibilities. The managers could be clinical or non-clinical. Interviewees were selected using a judgement sampling methodology (Perla and Provost, 2012).

9.3.7 Reliability testing

The interviewers each underwent a mock interview with me to ensure consistency and clarity of the interview schedule. All interviews were recorded to allow for reliability testing subsequently. A sample of 10 interviews were listened to again and the questions were re-scored for inter-rate reliability. I was blinded to the original scores during this process. The pooled Kappa value for the sample analysed was 0.83 with a $p<0.001$.

9.3.8 Co-ordination of the process

To ensure the process was delivered on schedule and that all participating units were interviewed in a timely fashion I co-ordinated the process. Contact was made with one representative member of the Global Comparators group either in person or via email that had expressed an interest in participating using a bespoke standardised correspondence (Appendix 2). I undertook training and received a National Institute of Health Human Subjects Protection training certificate before commencement of the study. A database of participants was created at each institution with their institutional contact details that was encrypted. Once all interviewers were trained and tested, the process went live. Initial contact details for a single institution were given to one interviewer at a time. The interviewers then made initial contact via e-mail and sent the four participants introductory emails and the overall goals of the project. I was carbon copied into all email correspondences. The interviewers were then free to arrange mutually agreeable times with the participants to undertake the recorded interviews. Track of all interviews performed and progress was noted centrally. An overview of the whole process is given in Figure 15 .

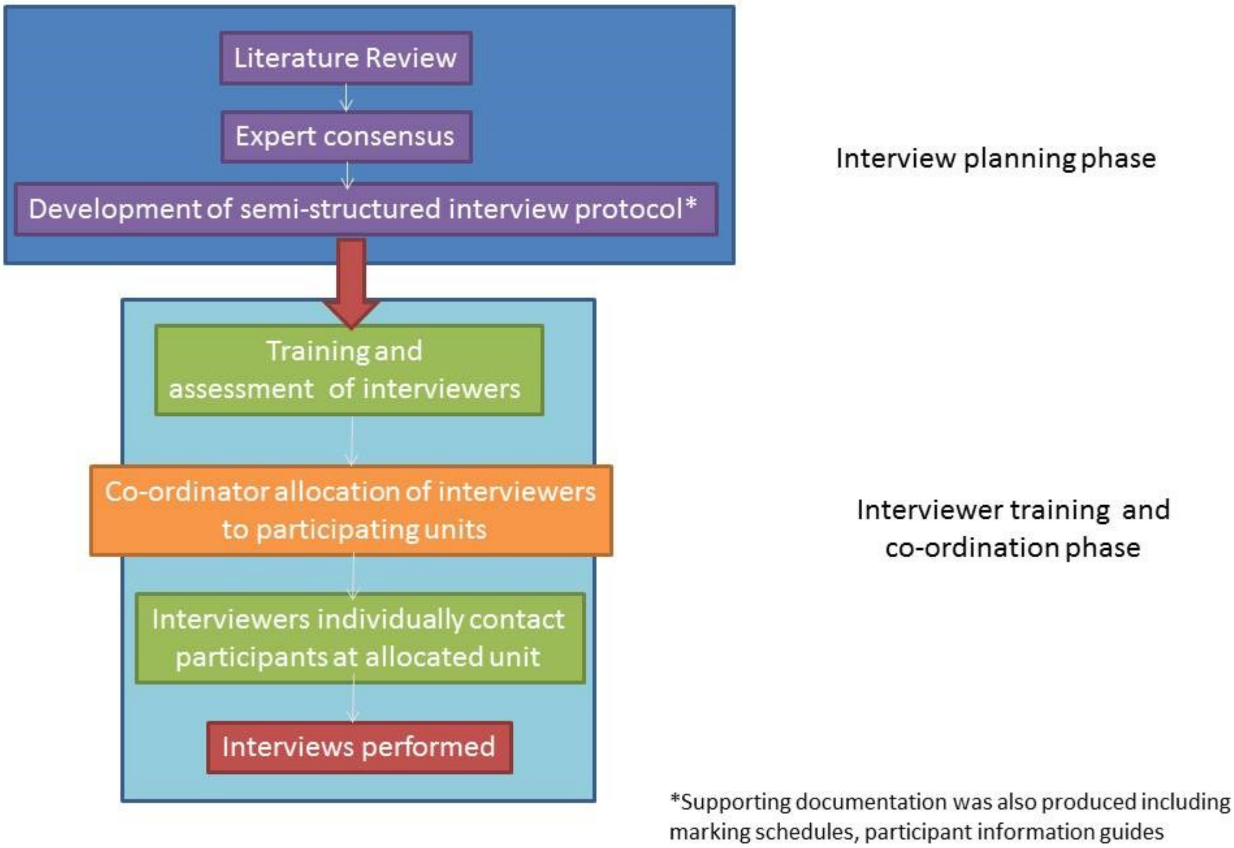


Figure 15- Scheme of how HiPer was planned and undertaken

9.4 Results

What are reported here are the final results from the interview planning phase and the development of the questionnaire. The included themes are given in Table 23. Furthermore, preliminary findings are also given to demonstrate the practical feasibility of undertaking such a study.

9.4.1 The included themes for the interview protocol

An outline of the final themes selected for inclusion into the interview schedule with example questions and reasoning are given in the table below-

Table 23- Table of the final themes selected by the expert consensus for inclusion into the questionnaire with example questions

Facet of care analysed	Example question	Reason
Opening question	"Can you talk me through the process a patient with colorectal cancer that needs an operation at your unit undergoes?"	Gives the participant the opportunity to describe what they think is important in the pathway at their unit
Structural facilities available *(2 questions)*	"Tell me about provision of CT scanners, interventional radiology and access to the operating room in your unit" "What about out of hours?"	To discern whether there were any wide variations in the availability of structural factors
Decision to operate	"Who decides when to operate?, is there a multi-disciplinary meeting / tumour board meeting?"	This question aims to understand the inputs and multidisciplinary nature of decision making.
Pre-operative care	"How are patients pre-assessed and risk stratified for surgery?"…."Who does this?"	To identify what routine processes patients undergo to be optimized for surgery.
Post-operative care	"Who reviews patients after their operation"…"How many years post-graduation are they?"	Quantifying how often and how senior a review is undertaken in the post-operative period
Identification of complications	"What methods does your unit employ to detect complications in the peri-operative period?"	To discern whether any early warning scoring systems are used or other ward based interventions

Teamworking	"How are patients handed over / handed back between nursing staff and doctors and clinical teams?"	To understand the process patient information is shared in the hospital
Informatics	"Are any outcome measures discussed by the team" "If so tell me more"	Aims to identify whether any goals or standards are worked towards. Also attempts to identify what and who drives this.
Quality improvement	"What drives quality improvement in your unit?" "What role do managers have in the unit"	The role of people with managerial responsibilities is probed.

9.4.2 Illustrative findings of feasibility

In all 9, units participated in the initial feasibility study. Included is an excerpt from one question to illustrate the methodology in practice. Described are the findings from the question entitled "Post-operative care". This question attempts to identify certain key aspects of post-operative care-

- How often and by whom are patients seen post-operatively (surgical and anaesthetic)?
- What seniority are these people performing the reviews in post graduate years?

In these examples a Consultant represents the most senior clinician which translates as an attending in the United States. For brevity just the term Consultant will be used.

Table 24 is an excerpt from the scoring schedule for this question. The score with the typical indicative response is given. A higher score denotes higher quality of care.

Table 24- A table demonstrating the indicative scores and responses as per the marking schedule

Score	1	3	5	7
Indicative response	consultants do not review patients before discharge juniors only review patients if 'required'	consultants review patients only if required only most junior member of team reviews patients daily*	consultants review patients daily on weekends patients reviewed by an on-call team (not necessarily own team)	Operating consultant reviews own patients daily including weekends and national holidays

*seniority of clinician was standardised by determining how many years post full qualification if there were any doubts due to regional variation.

The table below gives illustrative results as recorded by the interviewers real-time for surgeons and nurses for the question in relation to post-operative care in one institution.

Table 25- A table demonstrating the scores as recorded by the interviewers between two respondents for one question in the same institution

Unit	Respondent Score	
	Surgeon	**Nurse**
A	**5**	**5**
B	**7**	**6**
C	**4**	**4**
D	**4**	**4**
E	**7**	**6**
F	**6**	**6**
G	**6**	**6**
H	**5**	**4**
I	**7**	**6**

9.4.3 Free text and on re-analysis of audio-recordings

Below are excerpts from the audio recordings that illustrate the findings as tabulated above. All recordings were re-listened to. The purpose of doing so was to extract more in-depth findings. The findings of the free text analysis is illustrated using the table below-

Table 26- Indicative quotations from free-text analysis for two separate units on the same question

Unit	Indicative quotations	
	Surgeon	**Nurse**
F	***"They get seen every day by me, well definitely during the week…mine might not get to see me on a Sunday but a registrar will see them, but we** (the consultants) **generally come in on weekends even if we are not on-call"***	***"Patients are seen daily by the surgical consultant and again by the juniors daily"*** ***"On weekends Consultants do see them but it won't necessarily be their own consultant"***
C	***"Well day to day management is done by the registrar… and most consultants do a round twice a week, but we** (consultants) **do see them on day one post-op always"*** ***"On weekends they are seen by the on-call team"***	**"Well they** (the patients) **would be seen twice a day by a registrar…and the consultants seeing them, well not every day…well, probably, I would say, not every day, but again depends on the consultant"** **"Usually a consultant does a ward round on the weekend, if it's not a colorectal consultant on-call then they** (colorectal consultants) **sometimes come in on the weekends**

9.5 Conclusion

Feasibility has been demonstrated in this pilot study. It has been shown that a telephone based interview study using the described methodology for attempting to discern more subtle factors of surgical performance is feasible and on initial analysis that the methodology described, appears to reveal interesting and valuable data. It remains to be seen whether this method will translate to clinical outcome, however the depth of the information may be valuable to feedback to the participating units alone. In the next and final chapter the HiPer study will be continued and results of the full study will be reported as well as an attempt to correlate the study findings with site specific outcome data.

10.0 THE COLORECTAL HIPER PILOT PROGRAMME-ALL UNIT RESULTS

10.1 Chapter overview

In the previous chapter the development and piloting of the HiPer study protocol in colorectal cancer units was described. In this final empirical chapter the findings of the full study are correlated the study findings with outcome data from each unit. This is a two part chapter. In the first part, the full findings of the HiPer study are illustrated with the results of the telephone questionnaire study. In the second part of the chapter, these findings are then correlated to outcome data from each individual participating unit. A final appraisal of the process is then given.

10.2 Introduction

Having developed and piloted the HiPer telephone questionnaire, interview process and assessment of feasibility, this final study will bring together the completed findings of the study. This study attempts to bridge the gap between hard outcome data, such as those reported in routinely collected data using a variety of methods as described in earlier chapters (e.g outcome data from databases), with the questionnaire data from the HiPer study. The aim is to elucidate if any concordance can be demonstrated between the hard outcome data collected, and the novel methodology or whether HiPer may offer another facet to understanding what lies beneath routinely collected data.

10.3 Methods

Whilst a detailed methodology of the HiPer feasibility study and process is given in chapter 9, the focus of this methodology will describe how the findings of the HiPer study were correlated with the outcome data held for each unit.

For analysis of the questionnaire data, this was undertaken at multiple levels. Firstly the overall scores between interviewees at each institution will be compared against one another as well as against those of other units. A more detailed analysis on a per unit basis with more in-depth analysis will also be performed. The audio recordings have been re-analysed individually and in full length to create a one page pro-forma, per unit, that summarises the findings. Specific mention is made where scores have been less than (<4 ex 7) or greater than (>5 ex 7). Furthermore any aspect of unique care is highlighted. This section will be listed under 'Results-Questionnaire Study'.

For linkage to the outcome data, a different methodology will be employed. For each participating unit, outcome data is held on a secure database by Dr Foster Intelligence. These data were specifically uploaded to the Dr Foster Intelligence server using a standard data submission form by individual units. These data then underwent a standardised 'cleaning' procedure by Dr. Foster Intelligence and risk-adjustment (Bottle et al., 2013). The final available outcomes included - length of stay, risk-adjusted mortality and re-admission rates for all colorectal cancer operations from 2010-2012 (inclusive). This section will be listed under 'Results-HiPer linkage to outcome data'.

10.4 Results

10.4.1 Demographics of participating units

Initially, 18 units were invited to participate in the HiPer study. 9 units completed the full study, ethical and research board clearance. Data collection concluded in November 2012. These were geographically disparate units with 4 from England, 3 from the United States of America and 2 from mainland Europe.

10.4.2 Demographics of participating interviewees

In all 29 interviews were undertaken. It was planned to contact 4 people from each unit giving a total of 36 interviews potentially possible. One unit had a clinician manager (who answered for both surgeon and manager) hence reducing the total number to 35. This gave an interviewee response rate of 83% (29/35) for the whole study.

There were 9 surgeons interviewed (n=100%), 8 nurses (n=89%), 5 separate managers plus one clinician manager (n=67%) and 6 intensivists (n=67%) in total were interviewed. Parenthesis represents the response percentage from those available.

10.4.3 Length of interviews

In total, 966 minutes or just over 16 hours of interviews were conducted. On average, surgeons interviews lasted 40 minutes (range=19-52), nurses lasted 33 minutes (range=20-46), intensivists lasted 41 minutes (range=28-51) and managers lasted 17 minutes (range=13-21). Note - managers were only required to answer the final two questions, hence the shorter interview times.

Table 27- Table of all of the individual scores per unit

	STRUCTURAL	OOH	DECISION TO OPERATE	RISK STRATIFICATION	POST OP CARE	COMPLICATION Rx	TEAMWORK	INFORMATICS	QUALITY IMPROVEMENT	
Interviewee Position	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q9**	**Q10**	**TOTAL**
Colorectal surgeon	5	5	6	6	5	4	6	4	2	43
Head nurse	5	4	4	5	5	4	4	4	2	37
Head of Intensive Care	6	5	5	4	6	6	6	7	4	49
Colorectal surgeon	7	5	5	4	7	6	6	5	5	50
Head nurse	3	3	4	4	5	5	5	5	5	39
Head of Intensive Care	4	4	5	3	5	5	6	6	5	43
Manager								3	5	8
Colorectal surgeon	6	4	5	5	4	4	4	6	4	42
Head nurse	5	5	4	4	4	4	4	3	3	36
Manager (non-Surgeon)								6	4	10
Colorectal surgeon	4	5	4	5	4	4	5	4	3	38
Head nurse	5	4	5	4	4	4	4	4	2	36
Head of Intensive Care	6	5	5	4	7	6	6	4	3	46
Colorectal surgeon	7	5	5	4	7	6	6	5	5	50
Manager (Surgeon)								5	6	11
Colorectal surgeon	4	3	4	4	6	5	5	6	5	42
Head nurse	5	6	5	4	6	5	5	6	5	47
Head of Intensive Care	5	3	4	4	3	4	4	5	3	35
Manager (Surgeon)								6	7	13
Colorectal surgeon	3	6	6	6	6	6	6	6	5	50
Head nurse	3	5	2	5	6	5	5	3	5	39
Head of Intensive Care	3	5	2	4	5	6	6	5	4	40
Head of Surgery	4	4	4	5	5	3	4	4	5	38
Colorectal Nurse Speciali	6	6	6	4	4	4	4	4	5	43
General Surgery Manager								6	6	12
Colorectal Surgeon	6	5	3	6	7	5	4	4	3	43
Nurse	6	5	4	4	6	4	4	4	6	43
Manager								6	7	13
Intensivist	6	5	4	5	6	5	4	5	5	45

Results- Questionnaire Study

Interviewee scores within units

When the overall scores are compared, there is good correlation of average scores for the whole interview between the interviewees. This is demonstrated by a median score for surgeons of 4.81 (ex 7) and a median score for nurses of 4.44. The median scores for the intensivists was 4.81 and for the managers 5.58. The scores between the surgeons and nurses (in an individual unit) were statistically significant with a p-value of 0.023 - in other words, nurse and surgeons scores differed statistically. There were however no other statistically significant scores (i.e demonstrating significant differences between the scores) between the other interviewees (Table 28).

Table 28- Table demonstrating the statistical significance of scores between the different interviewee types for the whole study

Interviewee	Surgeon	Nurse	Intensivist	Manager
Surgeon	-	p=0.023*	p=0.528	p=0.101
Nurse	p=0.023*	-	p=0.550	p=0.750
Intensivist	p=0.528	p=0.550	-	p=0.343
Manager	p=0.101	p=0.750	p=0.343	-

*p-values between the interviewee groups, *denotes statistical significance*

The clustered bar chart (Figure 16) below demonstrates the score for each question on the y-axis. The blue lines represent the surgeons' scores and the green bars represent the corresponding nursing scores adjacent to the surgeons scores (Figure 16). The numbers on the x-axis represent the unit number. Unit 5 only shows blue bars as this unit did not have a nurse respondent. This chart demonstrates how closely the nurse's responses were to the surgeons for each question. In general the chart demonstrates how closely the nurses and surgeons scored on each individual question with very few scoring +/- 2 away from the other respondent for any individual question. Another way to demonstrate this is by formally assessing the correlation using a scatter plot.

Figure 16- Clustered bar chart of the individual scores per question for all units (blue represents the surgeon's scores and green represents the corresponding nurse's score for the same question).

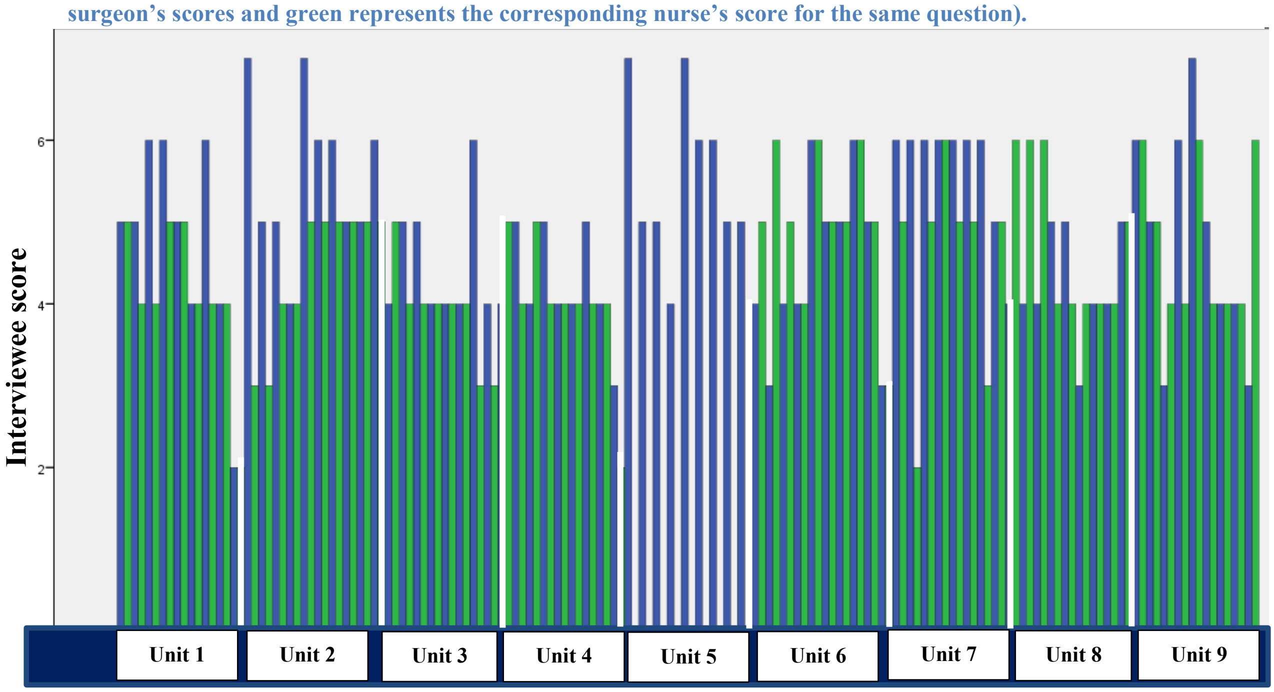
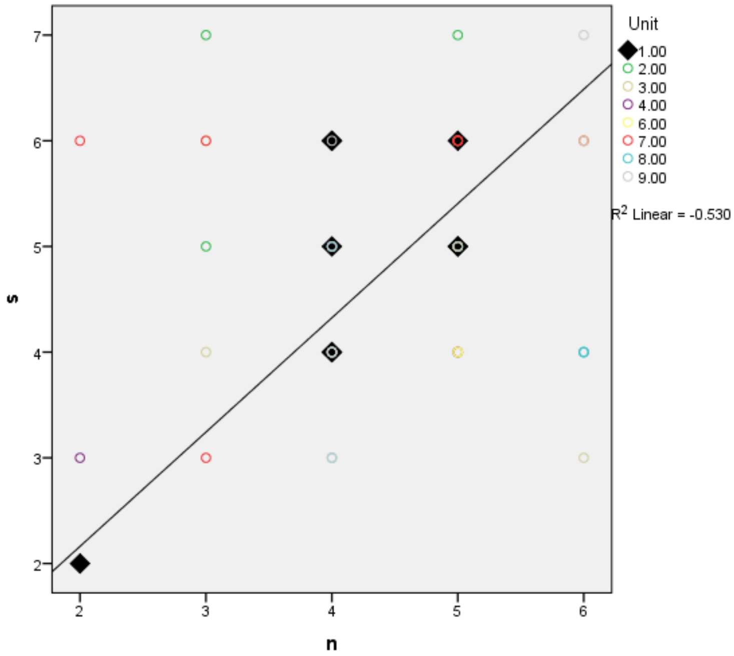


Figure 17- Scatterplot of surgeons (s) scores versus nurses (n) scores for all questions



When the correlation of the scores for each question is plotted on a scatter plot there is a good correlation between scores between the interviewees. In this case between the surgeons and the nurses. The scatterplot below shows all the scores of all the units for both surgeons and nurses. Initially there appears to be little correlation however when one unit is exaggerated by black diamond's six major markings are seen. As there were 10 questions it would be expected to see 20 diamonds however due to the close correlation of the scores many of the scores overlap. Furthermore when a line of best fit is plotted, it is seen that there is a linear correlation. The strength of this correlation is good as demonstrated by an R^2 value of 0.530 (Figure 17).

A similar finding is seen when the correlation is performed with surgeons and intensivists questionnaire data. A stronger correlation is seen with an R^2 value of 0.608. To reiterate, what is displayed here are the individual scores for surgeons plotted against the scores of the intensivists. Again in the scatter plot below one unit (unit 9) has been selected and emphasised using black triangles. Here again the close correlation is seen for this individual unit on the back-drop of all the units. In other words this scatter plot shows that the best fit line demonstrates less variability than the previous plot. Hence loosely it can be said that surgeons and intensivists results are more closely predictable to one another than surgeons and nursing scores. Only 6 units are displayed as these where the ones were there were both respondents from surgeons and intensivists (Figure 18).

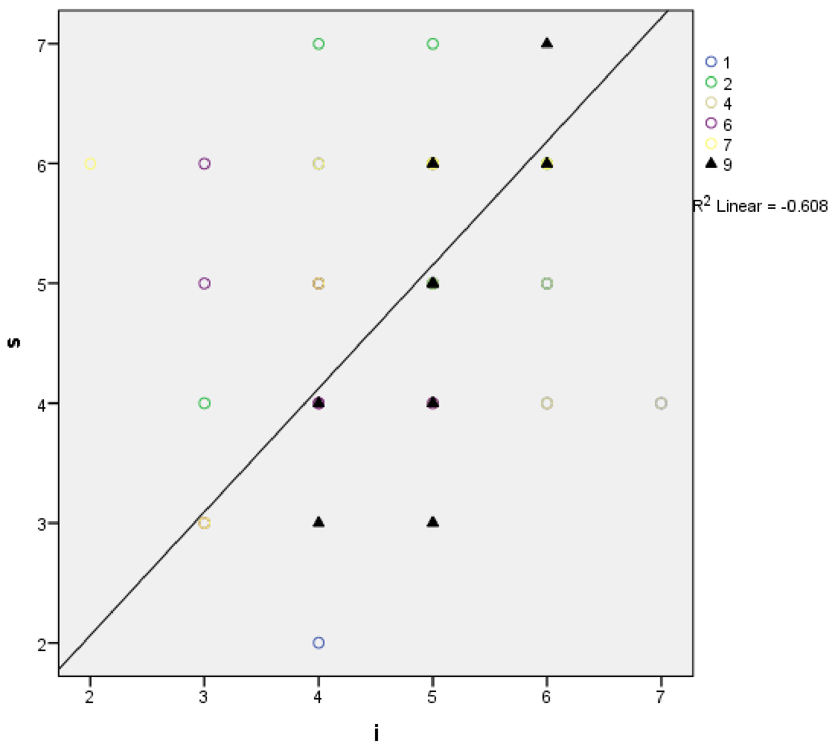


Figure 18- Scatterplot of surgeons (s) scores versus intensivists (i) scores for all HiPer questions highlighting unit 9

The table below outlines and summarises the individual R^2 values between the groups. The table shows that there is at least a moderate correlation between the respondents (Table 29Table 29).

Table 29- Table of the correlation statistics between the interviewees

Interviewee	Surgeon	Nurse	Intensivist	Manager
Surgeon	-	R^2=0.530	R^2=0.608	R^2=0.566
Nurse	R^2=0.530	-	R^2=0.395	R^2=0.397
Intensivist	R^2=0.608	R^2=0.395	-	R^2=0.588
Manager	R^2=0.566	R^2=0.397	R^2=0.588	-

Per unit assessment

To investigate any relationships further, assessment of the scores on a per unit level will be undertaken. Despite good correlation of scores between nurses and surgeons responses the scores were statistically significantly different, indicating a difference in the scores to the responses between the nurses and the surgeons. To explore this further the table below describes in which units the responses were divergent (Table 30).

Table 30- Table demonstrating whether there were statistical differences between the surgeon and nursing scores on a per unit level

Unit	Surgeon-Nurse (p-value)
1	0.072
2	0.390
3	0.387
4	0.029*
5	-
6	0.950
7	0.592
8	0.283
9	0.453

From the above table statistical differing responses are seen in unit 4, with the strongest statistical significance in unit 6. All the other groups showed no statistical differences in the responses.

Analysis of the results

Formal thematic analysis is not appropriate as the interview was designed on the basis of several themes as identified from the systematic review. Thus the analysis of the responses will be performed using a novel approach. The questionnaire was designed with the standard of expected care would score 4. Given the median scores were within this region for further analysis where care was rated to be specifically good i.e scores of >5 will be investigated further. Furthermore, for quality improvement and for less well performing units to identify areas of weakness analysis will also be performed where a response scored <4. This will be performed on a per unit basis. This has been undertaken by re-analysing all audio recordings and identifying the salient points that led to the scores being given. Notes in italics represent corroboratory views from other interviewees.

Unit 1

UNIT 1	High score (6 or 7)	Low score (1-3)
Surgeon	Electronic patient records available from clinicians homes include MDT data Nuclear medicine and radiotherapists attend MDT All patients seen by cardiologist if >45yrs old. Pre-op. assessment week of surgery >70% laparoscopic rates all enrolled ERAS Surgery usually <2/52 from MDT date	Managers viewed as money makers and savers Infrequent meetings with managers and clinicians
Nurse		Little known about managerial input into day-day workings Managers helpful if goal doesn't cost money
Intensivist	Full time senior intensivist cover Dedicated radiology sessions for ITU Strong focus on patient centered outcome measures Key indicators (LOS/readmissions) audited monthly Very academic unit focusing on integrating research with clinical practice	*(Managers determine the resources available otherwise don't really impact on day-day care)*

Strengths- heavy research base has been demonstrably used to improve patient care

Improvements- interim step-down unit from ITU

Unit 2

UNIT 2	High score (6 or 7)	Low score (1-3)
Surgeon	Patients can be pre-operatively assessed on day decision made/discussed to operate with patient Patient can usually be seen by adjuvant therapists on same visit Stoma counseling (as well as marking/teaching) is mandatory for all patient undergoing ostomies. Electronic patient records accessible from anywhere in world via VPN. Patients seen once daily by consultant-level and twice daily by rest of team including weekends. Dedicated discharge team review every patient Established dedicated provision for unplanned returns to theatre with agreed timescales	
Nurse	Rapid response team for ward patients Full electronic patient records	Operating rooms not ideally placed from wards need to go through 'a few buildings' Can take over ten minutes to get ward patients to imaging
Intensivist	Return to theatre protocol highlighted again No barriers to return to theatre/out of hours imaging	
Manager	*(Viewed roles as facilitating the service and physicians rather than goals based)* *(Gave example that by prospectively auditing results of patients that were developing pressure ulcers new specialist beds purchased have improved results)*	

Strengths- Intermediate care facility able to take patients medically fit for discharge but requiring recuperation. Nurses very familiar with 'monthly dashboards' with indicators- PE/DVT/patient satisfaction. Weekly 'quality' meeting with clinicians and managers.

Improvements- no formal enhanced recovery pathway. Weekend theatre staffing of 'other surgical teams' may affect general surgery 'emergency lists'. Intensivists not aware of ERP. Early warning scores on wards may be of benefit- not currently used.

Unit 3

UNIT 3	High score (6 or 7)	Low score (1-3)
Surgeon	Weekly meetings with managers and clinicians and nurses and surgeons to prioritise patients and theatre utilization to ensure optimal care and resource usage	
Nurse	*(specialist cancer nurses always contactable by patients)*	
Manager	Role viewed as facilitating clinicians to manage their individual workload Heavy emphasis on adhering to national treatment targets and making sure operations (theatre lists/clinics/capacity) allows for these to be met	
Intensivist	Dedicated care pathway adhered to for each patient incorporating ERP All major surgery and intensive care and wards within one block as well as radiology	

Strengths- Cohesive and regular meetings between clinical staff and managers with direct ability to enact change. Safety net of two separate non-clinical staff ensuring cancer patients receive treatment and do not 'slip through nets'. Strong sense of team-work from booking clerks to clinicians and managers.

Improvements- Look into collecting data on post-operative patient care/experience.

Unit 4

UNIT 4	High score (6 or 7)	Low score (1-3)
Surgeon	All colorectal cancer seen by colorectal cancer nurse specialist Well established ERP program that is adhered to as standard Audio visual diary day-day shown to patients that explains all expected interventions e.g drains and what to expect day-day Prospective database held of all colorectal cancer patients holding >10years of patients	Feel unable to engage trust to get individual surgeons outcome data "There are no specific drivers for quality from higher up" There is a feeling finances are main driver rather than quality of care
Nurse		

Strengths- Pre-operative assessments are undertaken by anaesthetic doctors, surgical doctors and a nurse. All structural factors well linked and within proximity.

Improvements- Colorectal unit specific induction may standardise care at personnel change-over. Dedicated enhanced recovery nurse may benefit. "A bit more support from management would help".

Unit 5

UNIT 5	High score (6 or 7)	Low score (1-3)
Surgeon	All resources including outpatients, wards and theatres are in same area of hospital Ease of undertaking CT scanning on same day if required e.g. from outpatients to decide whether to admit or not Seen by Consultant level once to twice per day and usually by the head of service once per day No reduction in care from senior surgical clinicians out of hours	*(Feeling anaesthetic work-up could be better as are not routinely seen pre-operatively)* *(No real meetings with managers regularly focusing on quality and outcome)*
Manager		

Strengths- Lead clinician takes responsibility for ensuring all pre-operative investigations and co-ordinates the process from decision to discharge. Close follow-up of patients in first two years post-operatively (3 monthly in first year).

Improvements- more formalised structure/process of discussing patients at MDT/tumour board meetings. Current presentation of patients to MDT meetings is ad hoc on physician desire to discuss. Consider introducing enhanced recovery program. Discharge planning could begin pre-operatively. More formalised induction for trainees and consideration of introduction of early warning scores on the wards. No discernible audits for quality measurement. The unit may benefit from surgeons being aided to understand their clinical outcome to guide patient care decisions in the future.

Unit 6

UNIT 6	High score (6 or 7)	Low score (1-3)
Surgeon	Main surgical wards adjacent to theatres, HDU/ITU and above emergency department 2 dedicated colorectal nurse specialists for cancer, dedicated- stoma nurse, enhanced recovery and IBD nurses. Anaesthetists informally routinely review patients post-operatively on day 1. Monthly feedback of mortality data- prospective Outsourcing of data to external company that gives individualized clinician reports.	
Nurse	Patients seen daily at least once per day by consultant There are protocols in place for whom to escalate care to in case of EWS triggers Electronic patient board with real-time information on how the discharge planning process is progressing-e.g. flags if patients waiting to see OT/PT etc. Managers do visit wards regularly and attempt to familiarize how improvement initiatives can be adopted/rolled out	
Manager	Close relationship with clinical staff and walk rounds to see ‘front line’	

<u>Strengths</u>- There was a genuine sense that a cohesive team effort is likely to improve care. There is a current flux with respect to post-operative analgesia and attempts to improve the service within the anaesthetics department. There was a good use of early warning scores on the wards and amongst clinical teams. MDTs occur on Fridays thus all patients are verbally discussed for weekend handovers. Very strong focus on ensuring data collection is good and that clinical audits are acted upon where necessary.

Improvements- More assistance with discharge planning pre-operatively (even if concerns highlighted before admission) patient discharges are frequently delayed. More input from radiologist at MDT meetings was needed. Interventional radiology service was deemed as being of limited availability out of hours.

Unit 7

UNIT 7	High score (6 or 7)	Low score (1-3)
Surgeon	Patients discussed in the MDT are seen on the same day as the meeting. Medical oncologists, radiotherapists, specialist nurses as well present at MDTs Large focus on good data for national audit Dedicated colorectal team and surgeons (not standard of care in this country)	
Nurse	Well-structured introduction of the EWS on the wards with agreed protocols of whom to call Very keen on accreditation with Joint Commission to verify quality of care provided	
Intensivists	There is daily interaction with management level colleagues with more formalized meetings every two weeks Surgeons closely review patients when on the intensive care unit.	*Unsure of MDT meetings and make-up*

Strengths- Strong evidence basis for practice incorporating laparoscopy and enhanced recovery since 2005. Nurses engage with doctors as well to improve care e.g. removing patients' catheters at certain times to regulate when patients should be reviewed for signs of retention as a result of nursing evidence. Many protocols standardise the post-operative care process.

Improvements- operating surgeons do not necessarily review patients within 48 hours due to commitments and workload. There is a departmental requirement that data is fed back to clinicians on their own outcome.

Unit 8

UNIT 8	High score (6 or 7)	Low score (1-3)
Surgeon	Electronic records of MDT outcome stored locally (not accessible remotely). Dedicated geriatrician assigned to colorectal service to pre-optimize patients including anaesthetic work-up Nurse led follow-up clinics and community stoma nurses review patients at home if necessary.	
Nurse	2 dedicated stenting lists in the week for colonic stents Close adherence to network guidelines and meeting cancer waiting time goals. GP informed of MDT decision within 24 hours of any decisions being made Medical liaison pre-optimisation re-iterated with CPEX testing if needed Daily phone calls post discharge for 1 week by specialist nurses	
Manager	Keen to benchmark care delivery with comparable units. Executive drive to improve patient experience and costs by reducing LOS and unplanned readmission where possible	

<u>Strengths</u>- Close collaboration with medical and anaesthetic teams in pre-operative optimisation. Emergency surgery commitments do not impact on routine elective lists due to dedicated resident emergency surgeon. Strong emphasis on phoning patients post-operatively. Collation of PROMS and national databases to review delivery of care all kept prospectively.

Improvements- Senior opinions for out of hour's radiology opinions required to support surgical clinicians. Organisational efforts required to assist patients discharge when medically fit due to social circumstances or for patients that were marginally coping at home pre-operatively. Introduction of early warning ward based systems. Emergency returns to theatre difficult due to other specialities requirements. Consideration of induction program for rotating clinical staff. Access to diagnostics (especially endoscopy) highlighted as an issue at present.

Unit 9

UNIT 9	High score (6 or 7)	Low score (1-3)
Surgeon	Very close integration facilities and wards facilitating patient care Important recommendations including MDT outcome recorded on an electronic patient record available to all clinicians All patients seen by anesthetist pre-operatively Use of ileus reduction pharmaceuticals reducing length of stay demonstrably Senior surgeon will review own patients over weekend Weekly mortality and morbidity meetings	Little informatics feedback to assist clinicians audit care delivery Ad-hoc managerial meetings
Intensivist	Good location of ITU/theatres and wards for patient transfer. Same senior clinician on the unit for whole week both surgical and intensive care Very focused on patient related outcomes and auditing results Weekly mortality and morbidity meetings re-iterated	
Nurse	Cohesive staff and regular assigned team members ensures smooth efficiencies during patient care Surgeons round with ITU for patients on the units	
Manager	Rigorous quality process dictated by national requirements Focus on patient satisfaction as well	

Strengths- Consistent information from all clinicians due to agreed care pathways leads to patients receiving standardised care. There is a strong focus on auditing own results.

Improvements- Consideration of formalised early warning ward based systems. More structured familiarisation of rotating clinical staff with care practices and escalation policies. Informal identification of complication rates and potential areas for clinical improvement- could be formalised with assistance of institution. Assisting clinicians to feedback relevant data to them they can use to improve service.

10.5 Summary of result findings of the HiPer questionnaire study

	England	United States	Continental Europe
Highlights	All cancer patients discussed at MDT meetings	Electronic patient records facilitate decision making including remote access facilities	Electronic patient records facilitate decision making including remote access facilities
	Specialist nurses dedicated to colorectal cancer patients	Daily senior patient reviews including out of hours	Strong sense of evidence based practice and keenness to take up novel processes
	Early warning scoring systems common place	Synergistic working between clinicians and managers to facilitate clinicians roles	
	No issues accessing out of hours radiology		
	No barriers to returning patients to theatre if required		
Potential improvement	More assistance with discharging medically fit patients	Formalize criteria for presenting patients at tumour board meetings	
	No intermediate care facilities	Consideration of implementing EWS/MEWS ward based systems	
		Disparate usage of Enhanced recovery programmes	
	Consideration of formalized induction program for rotating clinical staff		
	Requirement for senior radiology opinions out of hours		
	More support required for clinicians to collect and analyse their own individual data		

Specific highlights of the HiPer questionnaire study

- One unit described that all patients undergoing major resections are shown a DVD diary of what to expect on a daily basis. This included the meaning of drains and expected milestones on a daily basis.

- Other units described the facility to access electronic patient records remotely so they can keep track of patients even off site.

- One unit is using alvimopan (Entereg) for post-operative ileus prevention and demonstrated a reduction in length of stay of 1 day on average for patients given this.

- One unit reported contacting patients by telephone regularly after discharge for one week. This initiative was perceived to reduce length of stay. Patients that are deemed borderline to go home were sent home with the safety net of the regular contact.

Results-HiPer linkage to outcome data

In this section of the results the scores from the interview studies will be linked to the actual outcome data for the participating units. As described in the methods, this is an exploratory study to analyse the data on records from these units submitted and analysed by Dr Foster Intelligence. The primary outcome that will be looked at is whether on a per unit basis the interview scores correlate with the outcome data held. The outcome end points that have been considered are risk-adjusted mortality, length of stay and re-admission rates.

When surgeon's scores are correlated with the risk adjusted length of stay (Figure 19) an interesting result is demonstrated. There is a weak correlation with an R^2 value of 0.140. Furthermore this correlation follows that the higher the doctors questionnaire score, the lower the length of stay.

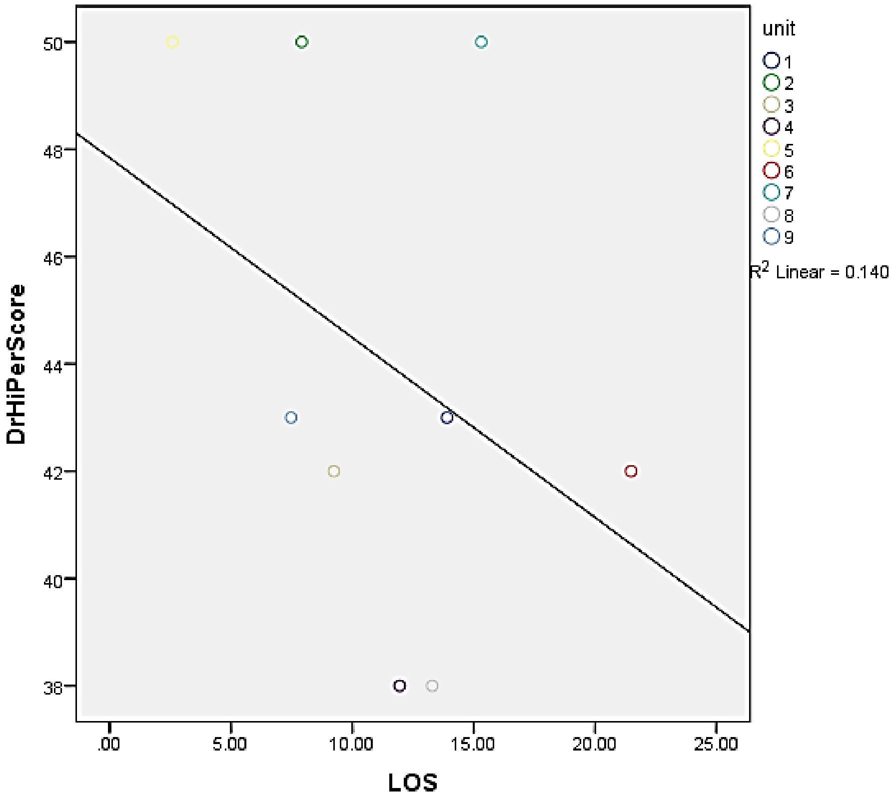


Figure 19- Scatterplot of doctors scores and length of stay

Specifically when the intensivists scores were analysed their scores have a moderate correlation with the institutional colorectal cancer length of stay with an R^2 value of 0.330. Furthermore, (as in with the surgeons scores and length of stay) the higher the scores of the intensivists the shorter the length of stay (Figure 20).

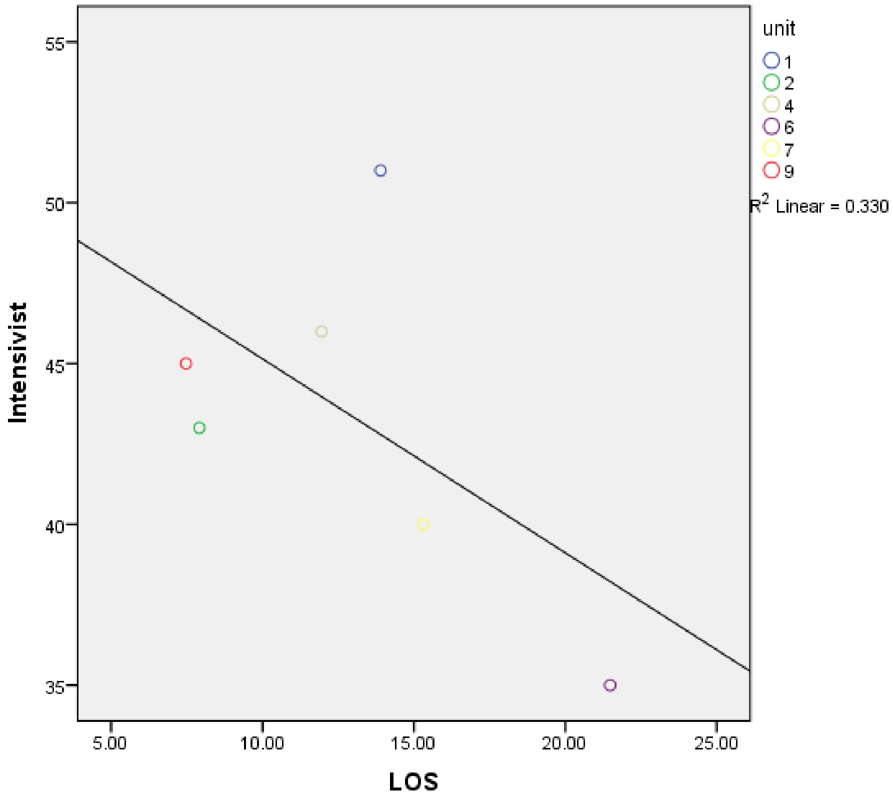


Figure 20- Scatterplot of Intensivists scores against length of stay

Again when the intensivists scores are analysed against the risk-adjusted mortality (Figure 21) rates for the individual units a stronger correlation with an R^2 value of 0.357 is seen. Again this correlation demonstrates that those scoring the highest scores on the questionnaire were in units associated with shorter mortality rates.

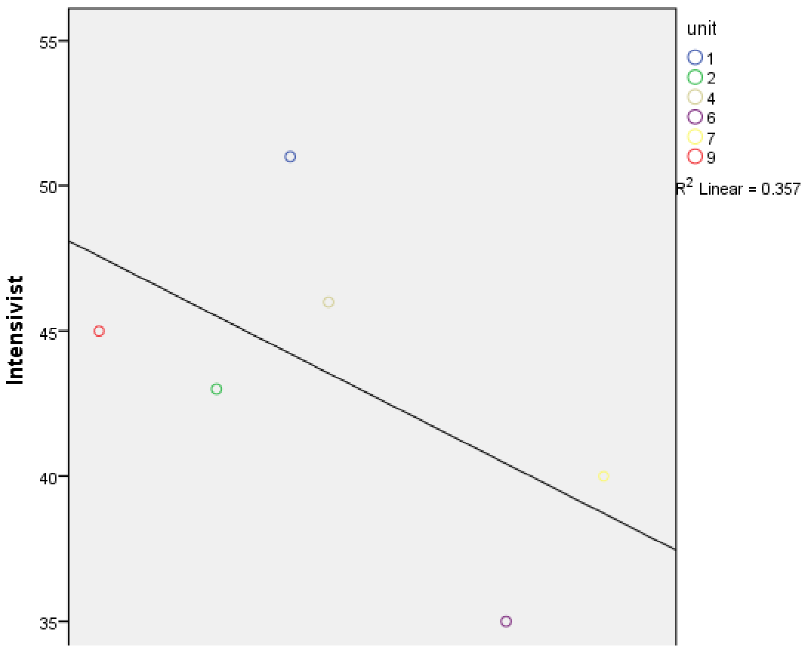


Figure 21- Scatterplot of intensivists scores with unit mortality

Nursing and managerial scores did not show any significant correlation with the outcome measures. Neither did surgeons or intensivists scores relate to any of the other outcome measures.

Specific comparison with the questionnaire study

Interestingly in the unit that described the use of ileus prevention pharmaceuticals, it was also one of the units with the shortest length of stay across the group.

In the unit that routinely telephoned patients on discharge, it had one of the lowest lengths of stay (in comparable units in the same country) with an average readmission rate.

10.6 CONCLUSION

This study has demonstrated that valuable information can be ascertained using the HiPer methodology. Specifically, correlation is seen between the different respondent groups in scores from the questionnaire study. When relating the questionnaire scores to the outcome data, surgeons and intensivists scores did correlate with some outcome measures. Whilst the numbers of data points are small the study has shown proof of concept and at some level correlation with some clinical outcome measures.

11.0 THESIS CONCLUSION

11.1 Overview of findings from thesis chapters

The primary aim of this thesis was to improve the understanding of what quality and high performance is in a surgical context and how it may be measured. This has increasing importance and relevance with the advent of open reporting of surgeon's data together with a growing national interest. Colorectal cancer surgery and the peri-operative period were chosen to contextualise this assessment for clarity and definition.

The introductory chapters reviewed the evolution of surgical quality appraisal originating from Codman to the contemporary open reporting of surgeons' results. In chapter 1 the historic and contemporary appraisal of variability in outcome and how and when variability is warranted was summarised. However, when variability breaches certain control thresholds or parameters further investigation is required. Unlike some industries, standardisation of care is very difficult to undertake in healthcare. The example of the Shouldice clinic in Canada is testament to the benefit of such a system. However, patient selection limits such care-process benefits across a wider healthcare system where all comers need treatment. Specifically, national healthcare systems need to provide care for all rather than selected patients. Thus absolute standardisation of all care pathways seems an unlikely answer to optimising care- although it has its place where feasible.

Chapter 2 studied paradigms of contemporary methods of performance measurement. This chapter appraised the Donabedian structure-process-outcome paradigm. Structural variables although easy to assess are the most difficult to influence. Process measures are useful if their utilisation can be directly shown to benefit care. Finally, outcome measures were considered. These are, in the most part, conceptually the easiest to understand. However, outcome measures can be ambiguous in nature. For example, length of stay is an outcome measure may be influenced by factors external to the quality of the care given in the hospital (e.g. social services provisions and the ability to discharge patients). Lilford's intervening variables introduce another aspect of quality appraisal that is typically very challenging to assess. Such intervening variables do however represent important factors, such as an institutions culture, cohesiveness and morale. Some appraisal of these factors is necessary when attempting to address the drivers of high quality care.

Chapter 3 analysed the contemporary methods of appraising national surgical performance. The Hospital Episodes Statistics (HES) database is introduced. This dataset may, if correctly analysed, not only provide 'standard' outcome measures such as length of stay and mortality rates but also generate novel ones. Examples of novel measures include return to theatre and re-operation rates. Such additional novel measures may well strengthen the use of such routinely collected data. Clinical registries are introduced while appreciating that the depth of clinical data makes their use for performance appraisal a very rich source of information. However, not all surgical registries at the time of writing were compulsory and thus omission of even a small percentage of patients may grossly misinform population averages. For example, if a unit reported data on 95% of patients operated on, this headline figure may be appealing. However, in theory at least, the unreported 5% cohort of patients may have all the mortality and morbidity within it. Patient perspective data were introduced by considering Patient reported outcome (PROM) data. The use of these has been shown to be more informative than reliable as a robust tool for quality appraisal due to the selective nature of the reporting and the subjective nature of the results from analysing PROM data.

Chapter 4 reviewed the literature to better understand what quality in colorectal surgery means. This review highlighted many potential markers of quality in colorectal surgery. It was demonstrated that the provision of a high quality colorectal surgical service demands consensus amongst many different groups. These include clinicians, patients, managers and societal stakeholders. Evidence-based structural and process metrics, as well as clinical and patient reported outcomes, may all be used for quality appraisal. With so many potential usable factors to appraise quality in colorectal surgery, some appreciation of what is feasible on a national level is required going forwards. The role of public and internal reporting of performance in colorectal surgery as drivers of quality improvement require further research and take us to the next chapter.

Chapter 5 assessed and compared a voluntary clinical colorectal registry- The National Bowel Cancer Audit Program (NBOCAP) outcomes with those derived from HES for the same patient cohorts. NBOCAP at the time was a voluntarily collected audit that at the time of analysis aimed to assessing peri-operative mortality and volumes following colorectal cancer resections. Surgeons were free to submit data to this registry or not. NBOCAP is a national clinical audit of bowel cancer run jointly by the Association of Coloproctology of Great Britain and Ireland and the National Clinical Audit Support Program (NCASP), part of the NHS Information Centre for health and social care. This dataset includes specific information on histological findings and pathology reports as well as adjuvant therapy administered. Trusts that submit data are able to compare their results with the 'national' outcomes (pooled data from submitting Trusts) as correlated by NBOCAP annually.

The aim of the study was to identify whether outcome of Trusts varied in submitting hospitals from non-submitting hospitals as compared to nationally held data (e.g. from HES) of the same units outcome for the same period. The aim being that if no outcome differences existed, clinical registries may be a valid appraisal of national outcome and performance and as discussed, *may*, be a richer source of information. The advantage of mining clinical registries lies in the fact that, for example, oncological measures represent important outcomes following cancer surgery. Specifically, oncological margins, lymph node yield and quality of the TME plane are predictors of successful surgical treatment (Stocchi et al., 2001, West et al., 2008b). Detailed clinical data and operative treatment intent are not available on large data sets such as HES, given its administrative origins. However, linkage of HES data with cancer registry data has been suggested as a means of overcoming the shortcomings associated with using HES alone (Garout et al., 2008). The crux of the study was to identify whether voluntary reporting yielded the same outcome (namely peri-operative mortality) as nationally held data from HES on a per unit basis.

In the comparison of Trusts that did and did not submit data voluntarily to the NBOCAP registry with outcome data from the HES database, postoperative mortality was higher in hospitals that did not submit data even after correction for case-mix differences. In the 2006–2007 NBOCAP report, only 44.3 per cent (70 of 158) of relevant Trusts in England submitted data. In the present study, using 2007–2008 data a significant increase in the number of Trusts submitting data was observed. Specifically, in the most recent NBOCAP report only 9.9 per cent of

Trusts (15 of 152) failed to submit any data. Despite the improvement, 23 per cent of Trusts still failed to submit more than 50 per cent of their colorectal cancer workload data. Garout and co-workers compared case volume and mortality for colorectal surgery between the NBOCAP and HES data sets and found that, at a national level, outcomes were comparable (Garout et al., 2008). However, when the data sets were compared at individual Trust level, significant inconsistencies were observed with respect to mortality, especially that following abdominoperineal excision (Garout et al., 2008). It must be noted that more recently submission to this audit has become almost mandatory with greater than 95% of units submitting data. Furthermore these data have for the first time become publicly available on a per-surgeon level for elective bowel cancer resections as of 2013 and reporting on 90 day mortality.

The present study was conceived on the basis that, if outcome differences exist between Trusts that report and those that do not, the status of voluntary reporting may need further evaluation. Following publication of the study submission has become mandatory and surgeons' data are now individually and openly reported.

Statistically significant differences between submitting groups were demonstrated for admission status and Charlson scores. Submitters operated on fewer emergencies and on patients with less co-morbidity. However, more socially deprived patients were operated on in submitting units. Although these could represent potential confounders to the crude observed differences in mortality rate, the adjusted analyses included the above factors as co-variables, and submitter status was identified as an independent predictor of increased relative risk of mortality amongst non-submitting Trusts.

Use of postoperative mortality as an outcome measure is appropriate only if it occurs frequently enough to discriminate statistically between high- and low-performing units (Daley et al., 2001). Arguably, outcomes such as mortality in patients undergoing elective surgery for colorectal cancer should be fairly homogeneous across Trusts. The present data yielded small, but significant, differences in absolute mortality outcome between submitting and non-submitting institutions. The clinical significance of such a finding is uncertain. Whether this mortality difference belies genuine broader differences in quality of care is unknown and cannot be deduced from this study.

Moreover, the present data relate to acute NHS hospital Trusts in England. Given that the average NHS Trust comprises more than two hospitals, Trust-level data potentially represent amalgamated outcomes. As such, outcome measurement from HES represents an oversimplification where opposing extreme outliers could theoretically negate one another within the same Trust. One might, however, expect this to occur indiscriminately within both submitting and non-submitting Trusts.

The observed difference in mortality between submitting and non-submitting Trusts may reflect either a true difference, whereby data submission is a marker of genuine, strong clinical performance, or reporter bias, whereby better units tend to self-report because of their better outcomes (Marshall et al., 2003). Another observation that may reflect differing quality of surgical care between institutions is the use of minimal-access surgery. The reduced number of laparoscopic procedures performed in the non-submitting group perhaps suggests a greater reluctance to adopt newer technologies. However, this difference was small (18.0 *versus* 15.0 per cent for submitters *versus* non-submitters) and may also be partly explained by case-mix differences such as the relative excess in the non-submitting group of patients having emergency surgery, who may be unsuitable for laparoscopic interventions.

In a study similar to the present investigation, Aylin *et al.* (Aylin et al., 2007b) compared outcomes between the HES data set and a voluntary reporting vascular clinical registry. In particular, they looked at caseload and outcome following vascular index operations including aortic abdominal aneurysm and infra-inguinal bypass. A principal study finding was that caseload was underrepresented in the vascular clinical registry in comparison with HES. The study has echoed such underreporting of cases in voluntary clinical registries.

Although this was not a longitudinal study, the analysis of 1 years' worth of data was chosen because the number of Trusts that submit data to NBOCAP differs year on year. Analysing a previous year's audit would identify 23 further units that would be termed non-submitters. These units, however, in the subsequent year's audit (the latest year available) would be termed submitters, and hence amalgamating 2 consecutive years may yield erroneous results. However, of the 15 Trusts that submitted no data in the 2009 report, 13 would have been termed non-submitters from the previous year's report, as well, demonstrating that these units were consistent non-submitters.

There has been much debate surrounding the interpretation of outcomes from voluntarily submitted data sets (Thompson et al., 2003). The Leapfrog Group is a collaboration of Fortune 500 companies and other large healthcare purchasers in the USA that, by publishing outcome data for individual hospitals, hopes to drive better outcomes from reporting (Leapfrog Group, 2009). Participation in the Leapfrog Group initiative and survey is voluntary. It has been suggested that the driver of submitting data to ‘hospital comparison’ websites favours hospitals that are seen to have better outcomes, as they are more likely to attract patients and funding (Leonardi et al., 2007). The argument is that, possibly, only the best institutions would report their data voluntarily and thus the Leapfrog Group may not be discriminating between good and poor units but merely reflecting variability within good units. Ghaferi and co-workers (Ghaferi et al., 2009c) have demonstrated that high-quality hospitals are not overrepresented and that voluntary reporting to the Leapfrog survey does not result in a significant bias in outcomes.

Evidence exists that reporting of clinical outcomes may lead to outcome improvement. In the 1990s, the risk-adjusted mortality rate for cardiac surgery decreased by 41 per cent in New York state following public reporting of surgeons’ outcomes (Hannan et al., 1994). Critics cite increased referrals of high-risk patients to out-of-state surgeons following public reporting as one reason potentially underlying this outcome improvement, although this assertion has been challenged (Omoigui et al., 1996, Chassin et al., 1996). In contrast, there is evidence from the American College of Surgeons’ National Surgical Quality Improvement Program that has shown reproducible clinical outcome improvement through a mechanism of internal reporting of clinical outcomes, where hospital performance is compared with national averages from participating units (American College Surgeons, 2010). This process has been shown to improve clinical outcomes and reduce complication-associated costs (Dimick et al., 2004).

Hibbard and colleagues demonstrated that open (public) reporting led not only to improvements in clinical outcomes, but also to institutions undertaking more quality improvement measures, and found that patients remembered the outcomes of such reports for up to 2 years after publication. Hospitals that had their results publicly reported demonstrated the greatest benefits when compared with hospitals that undertook only private or no reporting of their outcomes (Hibbard et al., 2005).

Submitter status alone may not be the only consideration of relevance. Even within the submitting group there was a wide range of total caseload submission, with units submitting between 10 and 100 per cent of their total caseload. If anything but 100 per cent of cases is reported, it is conceivable that reporter bias may be introduced. A unit may submit 90 per cent of its data, retaining all the morbidity and mortality in the remaining 10 per cent – hence misreporting their true outcomes despite appearing to be a submitting unit. This may further strengthen the argument for mandatory submission (Almoudaris and Omar Faiz, 2010). This concept would not affect the present results as only submitter status was inferred from the NBOCAP report. Subsequently, all patients in that Trust were analysed from the HES database on a per Trust basis. Further work would entail analysing clinical outcomes for Trusts before and after submitter status changes. However, this would rely upon submitter status being maintained year on year, which has not always been the case.

A minor but significant difference in mortality rates was observed between reporting and non-reporting Trusts to a clinical registry. The clinical significance of the difference observed in real terms appears small, but it may belie wider implications. If these findings reflect genuine outcome differences that arise between Trusts that do and do not voluntarily submit data to clinical registries, then, mandatory reporting to such registries may be a future consideration if the data are to be used nationally to benchmark performance and quality amongst surgical providers.

Given the result of the study, the use of clinical registries at present without full submission may not yield accurate inference of national performance and thus make identifying high performing units difficult. From a national perspective the HES dataset is left. Despite the potential shortcomings previously identified and reiterated in the subsequent section, it does capture national workload. One major limitation however is that to date no appraisal of complications has been possible from HES. In the review chapter it was identified that the management of complications is vital to establishing performance in colorectal surgery. Some appraisal of complication management needed to be considered. The following chapter attempted to ascertain whether any inference can be made from HES data of surgical complications and their subsequent management.

Chapter 6 presented a study of national outcomes from all English NHS institutions undertaking colorectal resections for cancer with the aim of commenting upon complications and their management. The aim as previously described was that the HES dataset was chosen to appraise units due to the outcome differences observed from voluntarily collected outcome data from clinical registries when compared to HES. However the main shortcoming of HES data was that no ability to comment upon complications and management was to date possible from HES. This study attempted to see if complications and subsequent outcome was derivable from HES.

In this chapter it was found that reoperation rates (for serious surgical complications) were similar among institutions classified within two groups- the high mortality quintile (HMQ) and the low mortality quintile (LMQ) groups- see Figure 22- Graph of overall mortality of the low mortality quintile (LMQ) units versus the high mortality quintile (HMQ) units with reoperation and subsequent failure to rescue-surgical (FTR-S) rates.

.

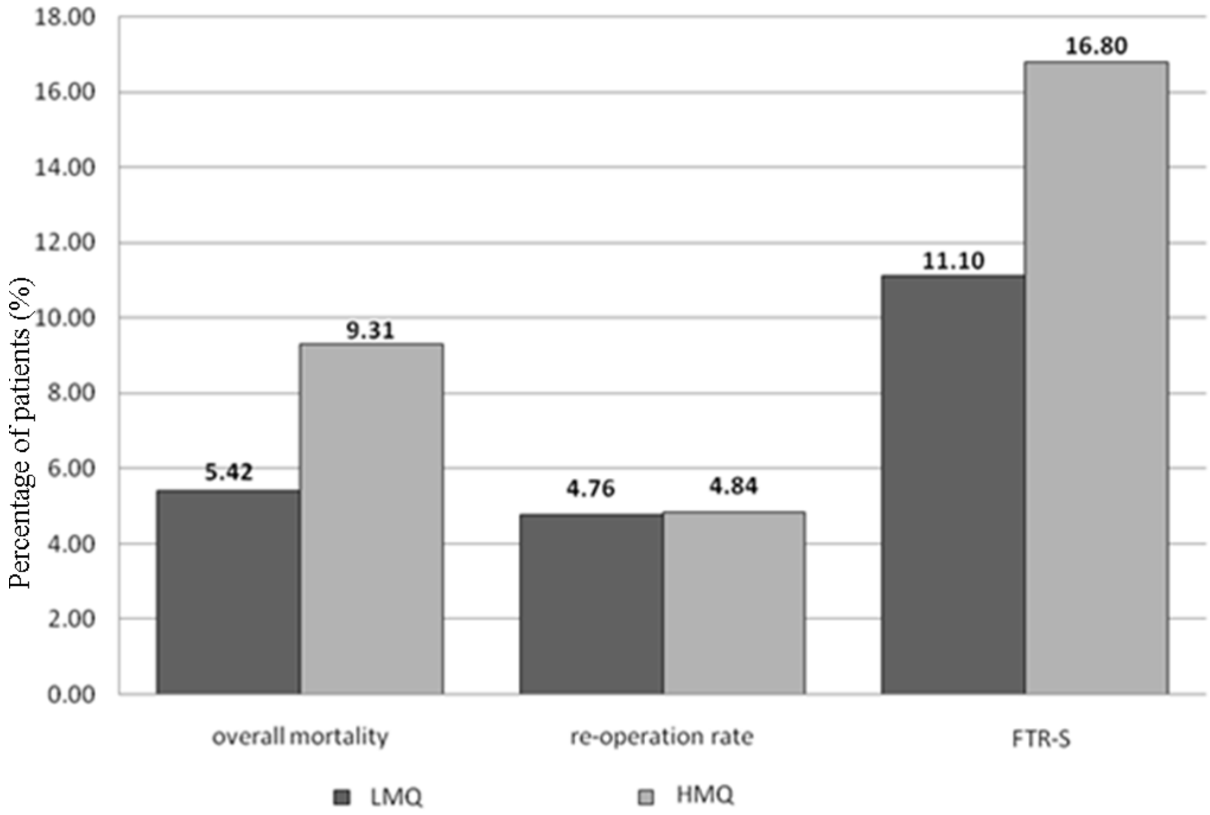


Figure 22- Graph of overall mortality of the low mortality quintile (LMQ) units versus the high mortality quintile (HMQ) units with reoperation and subsequent failure to rescue-surgical (FTR-S) rates

Moreover the types of complication that resulted in reoperation were similar among units in these extreme quintiles, as was the time lag between the index operation and re-intervention. However, high- and low-mortality units were distinguished by their ability to rescue patients following reoperation, with low-mortality units demonstrating an enhanced ability to prevent death in this context. This was termed Failure-to-rescue-surgical (FTR-S).

The management of emergency patients who develop complications may involve some different care processes (and teams) than those employed for elective patients. Nonetheless elective and emergency admissions were combined in the present study as this reflects actual practice. Elective colorectal resections are associated with low mortality but higher morbidity rates, whereas emergency resections have both high morbidity and mortality rates. Units that are inherently poor at managing complications are likely to be poor in dealing with both sets of patients, with the converse also being true. By combining these patients (elective and emergency) it was anticipated to demonstrate differences in FTR-S to a greater extent.

The results suggest that FTR-S is a more precise marker of surgical complication management than the more general FTR measure, which may include previously acquired medical complications as per Silber's initial description (Silber et al., 1992). Application of FTR to patients experiencing surgical complications that necessitate a return to theatre represents a meaningful measure of clinical and organizational ability to manage serious complications successfully. Moreover it is derivable from currently available data sources.

Most organizational structural factors did not appear to contribute to the observed differences in FTR-S rates between hospitals in the present study, although LMQ units had a greater number of HDU beds statistically-Table 31.

Table 31- Structural factors associated with low and high mortality quintile units

Structural factor compared	Lowest mortality quintile (median)	Highest mortality quintile (median)	p value
Size-			
Average no. beds per unit	683.82	791.00	0.196
Imaging^			
CT scanning	17.33	14.66	0.174
Ultrasound scanning (non gynaecological)	27.54	28.90	0.515
Fluoroscopy	9.82	7.40	0.069
Level I+II beds⁺			
ITU beds	1.32	1.05	0.425
HDU beds	1.04	0.78	0.011*
Theatres^	2.30	2.20	0.233

*indicates significant at the p<0.05 level
^average number of scans requested per patient bed per year
+average number of beds per in-patient bed multiplied by a factor of 100
^average number of theatres per in-patient bed multiplied by a factorial of 100

The significance of this finding is uncertain as greater resource availability could not be linked directly to greater HDU use among patients requiring reoperation. It is, however, likely that this compromised patient group would benefit from an HDU facility if it were available following re-intervention. Moreover, there is published evidence to suggest that managing patients in a surgical HDU leads to reduced morbidity, with a trend towards shorter length of hospital stay (Jones et al., 1999). In addition, low patient to nurse ratios are associated with a decreased risk of death within 30 days of admission among surgical patients (Aiken et al., 2002). One might expect that hospitals with low-quality perioperative care processes might fail to recognize patient deterioration promptly, thereby leading to delayed re-intervention. No difference was found, however, in the time to return to theatre between the two extreme mortality quintiles in this series. Nor was a relationship observed between ability to rescue patients following reoperation and number of reoperations. This suggests that institutional experience does not necessarily determine mortality risk in the event of a complication requiring reoperation.

FTR-S reflects processes of care other than traditional markers of surgical quality. For example anastomotic leak rates may reflect a surgeon's ability to perform a defunctioning procedure or anastomosis, case selection and technical factors. Such variables may also have an impact on mortality, but if a surgical complication arises further explanation is needed. An understanding of how well resourced units are, the intensity and seniority of ward care, access to radiology and other such factors is required. Mortality is not fully explained by standard metrics. For this reason FTR-S may explain why some units are able to prevent the conversion of serious morbidity to mortality. The rationale of a metric such as FTR-S is to understand why some institutions are better at preventing this conversion.

Overall, the reoperation rates reported in the study are consistent with contemporary literature (Merkow et al., 2009b). Patients who underwent reoperation were approximately 1.7 times more likely to die in HMQ than in LMQ units. No explanatory differences were found between the timing of re-intervention or operative caseload at HMQ and LMQ hospitals. As such, perhaps FTR-S is more a reflection of the quality of care after reoperation, than a marker of the recognition of complications and timely intervention.

In the study, the primary reason for return to theatre was generally in keeping with those reported in the literature, although direct comparisons are difficult owing to the discrepancies in definitions used between different studies. Morris and colleagues found that 21.1 per cent of patients who had surgery for colorectal cancer required re-intervention for wound complications, compared with 19.1 and 20.6 per cent for LMQ and HMQ units in the present study. Patients who died in this study were most likely to have required either formation of a stoma, further colorectal resections, abdominal washout or a combination of these procedures (Table 12). These are likely to represent procedures undertaken in the management of an anastomotic leak, although there are no data to support this assertion as there are no specific HES codes for anastomotic leak. In contrast, among patients who died, only 11.3 and 8.2 per cent (depending on quintile) had small bowel resection as a contributory re-operative procedure before death. Similarly, relatively small numbers underwent division of adhesions before eventual death.

These results raise some interesting questions regarding the quality of surgical decision-making and postoperative care of surgical patients. Further work is necessary better to understand the processes that underlie these findings which are not discernible from information recorded in administrative databases. This would probably involve more qualitative methods to elucidate salient factors such as patterns of on-call cover, availability of and seniority of specialist colorectal surgeons, and access to interventional radiology.

The reoperations selected in this study represent commonly occurring surgical morbidity following colorectal resection. This group of procedure codes defines a homogeneous group of patients who experience severe surgical complications demanding specialist expertise and intervention to enhance their chance of survival. A surgical provider's ability to have an impact on this group is therefore a potentially important measure of service quality. FTR-S has four criteria that strengthen its use as a quality metric. First, use of a discrete event such as reoperation is not open to interpretation – a patient is either returned to the operating theatre or is not. Attempted addition of less discrete medical complications from administrative data sets, such as wound infection not requiring intervention or chest infection, could perhaps render this metric a less reliable measure of quality due to the ambiguity of clinical definitions. Second, FTR-S is more likely to be based on discrete surgical complications derived from the index operation as opposed to medical conditions possibly present before surgery (although it is acknowledged that pre-existing conditions may make certain postoperative complications more likely). Third, given that reoperation is relatively uncommon, it should be feasible and practical to target these patients by using FTR-S primarily in quality improvement programmes. Finally, FTR-S may be derived from currently available routinely collected administrative data.

Whilst the study compared the influence of operative re-intervention, it is known the non-operative re-interventions are equally important in complication management. However, due to the rarity of non-operative interventions (e.g. radiologically guided drains/endoscopy) on the colorectal HES dataset it was not possible to appraise this in the study.

Given the importance of non-interventional procedures in complication management a further study was undertaken in upper gastrointestinal surgery to ascertain whether it is possible to derive the impact these non-interventional measures may have, in a setting where these interventions occur relatively more frequently.

Chapter 7 thus was concerned with appraising the impact that non-interventional measures such as endoscopy and interventional radiology have on outcome, and, if this was derivable from HES. This study showed that low mortality units rescue patients after re-operations more frequently than high mortality units. Through analysis of non-operative re-interventions, and subsequent outcome, it has been possible to reflect on those aspects of surgical care that are encountered when surgeons are faced with serious surgical complications in high and low mortality units. In those circumstances, surgeons have the option of watchful waiting +/- medical therapy, non-operative re-intervention (radiological drains / endoscopic therapy) or re-operation. What determines the treatment path taken is dependent upon many factors including the patient's co-morbidity and physiological reserve and the available resources and experience. Ultimately, surgeons bear the responsibility on the final decision taken.

Timely diagnosis of a major surgical complication and the performance and reliability of surgical teams are of vital importance. Supportive care such as intensive therapy units, physiotherapy and dietetic input may contribute towards the successful rescue of a patient. Similarly, structural factors such as out-of-hours radiology services and nurse-patient ratios may vary between LMU and HMU. Future studies are needed to examine the relationship between Failure to Rescue rates in different surgical units and the aforementioned factors that determine the treatment path chosen and the outcomes encountered.

The modified Calvien-Dindo group III types of complications were chosen for several reasons see Table 16. Firstly, certain post-operative complications may be present on admission such as pneumonia or deep vein thrombosis that are not necessarily discernible from administrative databases and are subsequently discovered in the post-operative period. This would influence failure to rescue rates and whilst important, if they cannot be adjusted for, this would not faithfully reflect the actual care given by a unit. The interventions chosen are not open to variability of definition. Certain post-operative complications such as respiratory tract infections and wound infections may be under-reported due to differences between unit's definitions. Our study has clearly defined complications in that patients were either returned to theatre or not, patients either underwent post-operative endoscopies or drains or they did not. Finally, Clavien-Dindo grade III complications are arguably those that most likely represent surgical technical quality and anastomotic complications (Lee et al., 2011). In oesophogastic cancer surgery these are factors that have important implications peri-operatively for survival (Yoo et al., 2011, van der Schaaf et al., 2012).

Re-operation following oesophago-gastric cancer surgery has been repeatedly associated with poor peri-operative outcomes in the literature. Such findings may influence contemporary decision making by surgeons faced with the difficulties of managing postoperative surgical complications. The implications of our study raise the question of whether more aggressive appropriate re-interventions can in fact confer better outcomes. Surgeons should interpret these findings within the context and limitations of their own units. Improved outcome from oesophago-gastric cancer surgery is more complex than just the volume-outcome relationship. Complication rates have repeatedly been shown to be equivalent between high and low mortality units in different specialities. Complication management is becoming more widely recognised as an important discriminator of surgical outcome. Surgeons should be supported with all the facilities and expertise necessary to ensure all facets of the care they deliver are optimal. Thus any appraisal of performance should also consider these elements. Thus this study has shown that where feasible, non-operative re-interventions are possible to derive from national datasets such as HES and can yield meaningful results. From this, the final data chapter attempts to bring together all the previous chapters. The chapter attempts to appraise all national colorectal cancer units on a panel of metrics for the first time.

Chapter 8 utilised HES in an attempt to appraise colorectal cancer units on a panel of metric and to examine the inter-relationship. It has been shown that serious surgical complication and re-operation metrics can be discerned from HES, however, what the following study highlights is the complexity associated with service quality appraisal in colorectal surgery. It questions the reliability of reporting individual metrics as universal markers of provider performance. The findings have important future implications regarding surgical benchmarking and quality improvement. The study suggests that units are not *necessarily* substandard performers across a range of outcome metrics despite being high mortality outliers for postoperative mortality. Although the study findings suggest that high 30-day mortality outlier status does not necessarily reflect poor overall institutional performance, low 30-day mortality outlier status does seem to convey at least 'standard' overall performance. When LMO units were considered across other outcome domains only two units performed worse than expected on two separate outcome measures.

Dangers potentially arise when 'good' and 'poor' performance labels are assigned to units on the basis of outlier status using only single metric evaluation. Certainly, factors such as case-mix could underlie outlier status and are potentially not fully accounted for on routinely collected datasets. The complexity of performance appraisal is appreciated when Figure 13 and Figure 14 (Chapter 8) are considered together. From the figures unit C has a significantly lower than expected rate of returning patients to theatre despite HMO status. Unit C has a lower than expected FTR-S rate also. The latter markers (reoperation and FTR-S rates) potentially represent high performance when taken in isolation; however the unit is a known HMO. The latter unit's high mortality rate is therefore not a consequence of surgical re-intervention. One explanation for this finding may however be that patients at this institution are perhaps not being returned to theatre when it is indicated. Alternatively, perhaps complex case-mix underlies this finding and patients are dying from non-surgically related causes postoperatively. When one however considers the outcome of units B and E, it can be observed that they both return patients to theatre more often than expected (Figure 13 in Chapter 8). Yet, these units are distinguished in Figure 14 where unit D lies within normal control limits whereas unit B lies above the upper 2^{nd} control limit for FTR-S. This suggests that the latter unit is not salvaging the patients it is returning to theatre. In contrast, in unit D the high mortality outlier status appears not to be due to failures in rescuing patients following reoperations.

Finally, it may be argued that 30-day mortality performance measures should be applied only to patients undergoing elective surgery. The role of the surgeon's ability, in terms of case selection and technical skills, must be influential in this context. In contrast, when both elective and non-elective admissions are considered together outcome is dictated also by surgical and intensive care teams' propensity to operate and support such high risk patients. Risk-adjusted models for both scenarios i.e. elective only (Figure 23) and elective & non-elective resections combined (Figure 9) were also considered.

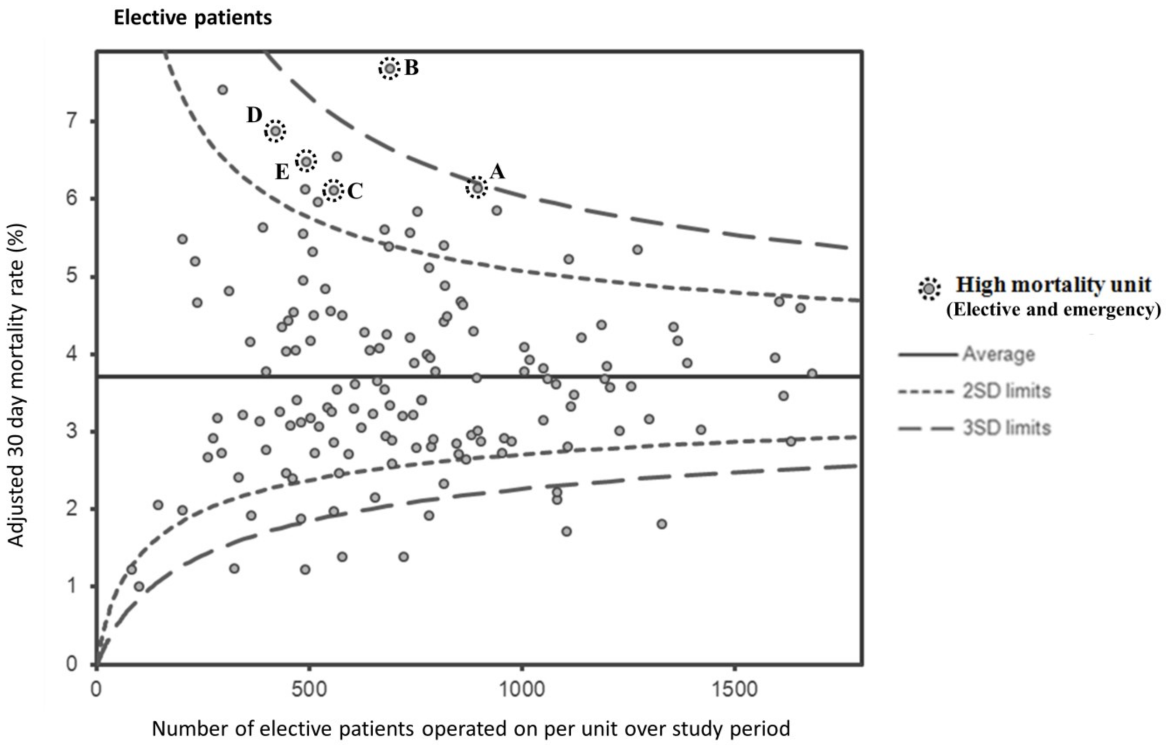


Figure 23- Funnel plot of elective mortality of the cohort with high mortality outliers highlighted for emergency-elective cases combined

Units identified as HMO (i.e. > 3rd s.d.) when considering both elective and emergency resections were also outliers (at the >2nd s.d threshold) when elective resections are considered alone (see Figure 23). This implies that factors such as mode of admission, are perhaps not overwhelming determinants of performance ranking. Moreover, 30-day mortality at all five institutions exceeded 6% within the elective setting despite a national adjusted mean of 3.8%. As such, it appears reasonable to consider elective and emergency patients together in this form of analysis. Furthermore, this method offers appraisal of how a colorectal cancer population is treated by a given institution rather than just how patients that present via discrete elective or emergency channels may do. Institutional performance in colorectal surgery is to some extent denoted by minimising patient exposure to non-elective presentation and consequent intervention. As such, efforts such as bowel stenting are practised in many centres to avoid emergency operations. Some clinicians consider this practice a marker of high quality service provision. Hospitals that successfully employ such procedures consequently operate on these patients electively but on an expedient basis. As such, examination of the elective workload in isolation may negatively bias their outcome despite arguably providing a better service than those that might just undertake an emergency operation. For this additional reason, inclusion of both elective and emergency colorectal cancer patient groups into perioperative mortality risk models appears warranted.

The low correlation between outcomes corroborates that defining quality in colorectal surgery is complex (Almoudaris et al., 2010). This further calls into question how quality in colorectal surgery can be quantified and meaningfully benchmarked. The fact that little correlation exists between postoperative mortality and other metrics suggests that achieving a definition of quality is potentially subjective and dependent upon what aspects of quality are prioritised. This perspective depends upon the viewpoints of the stakeholders concerned. Moreover, the lack of such correlation demands that overarching decision-makers (surgical professional bodies, health policy makers, hospital managers) decide upon the importance and relevance that should be placed upon individual performance targets. Furthermore, extrapolating high performance from such metrics may be complicated by the perceptions of good outcome dependent upon the agenda and goals of individual stakeholders (Solomon et al., 2003). For example, mortality and the formation of a permanent stoma are likely to be of greater concern to more patients than length of stay. By contrast, bed stay and its associated cost is likely to have greater implications for managers and service providers. As a result, there is a need to rationalise which measures should be targeted for benchmarking and quality improvement purposes.

In terms of groups of measures, moderate correlation was identified in the current study between the mortality related measures (i.e. 30-day mortality, reoperation and FTR-S). This indicates that, on some level, these measures reflect similar aspects of clinical decision-making and care received. This questions whether it would be possible to use a single composite metric to describe multiple mortality-related outcomes. Importantly, however, the funnel plots demonstrate very different performance levels amongst the five high mortality outlier institutions when FTR-S rates are considered, thereby suggesting that considerable institutional difference in prevention of death after re-operations occurs. 30-day mortality, despite risk adjustment, may offer little information regarding which deaths have arisen that could have potentially been avoided through re-intervention or better quality perioperative care. Appraisal of both metrics using funnel plots depicts differing apparent 'poor performers'. It is therefore a subjective decision to determine which measure (or indeed both) should be used to reflect the desired goal.

A clear understanding of the scope of each outcome metric used to reflect performance in colorectal surgery is required if these are to be openly reported (Thompson et al., 2010). Previous attempts to report variability in practice in colorectal surgery have been met with mixed response (Morris et al., 2008). This has mainly been due to the limited extrapolation that is possible when measures are heavily influenced by clinical factors (Faiz et al., 2009a) that are not represented fully in the datasets used for the analysis. It is therefore perhaps important that any audience is fully informed as to the complex relationships that surround these metrics should they be subject to open reporting. In addition, offering information on an institution's performance across multiple outcome measures might allow transparency with regards to overall performance.

Thus the difficulties associated with appraising units on single, combined and multiple metrics for the identification of high performance are acknowledged. As described earlier, some appraisal of other factors, such as those identified by Lilford, need to be considered. However, such analysis is not possible from the interrogation of databases alone. What was needed was a novel approach to assessing how units perform day-day and attempt to correlate these findings with the clinical outcomes described above. A more in-depth per unit analysis and novel methodology is required to try and uncover what 'lies beneath the numbers'.

Chapter 9 describes such a novel approach to assessing a complex healthcare process. This methodology takes from previous work demonstrating that assessment of complex managerial strategies can be performed using similar techniques. This is the first attempt at such a study in healthcare, in a specific focused clinical setting. Central to the success of this methodology is the fact that the units are invited to give information about how and what they are doing. There is no mention of linkage to outcome or assessment of performance. The understanding is that by collection of such data in this fashion, collectively there will be areas of excellence that can be shared within the group. There appears to be concordance (at least in the question chosen for this feasibility study) with the responses between differing healthcare professionals. This will allow for retrospective appraisal of those units that score the highest to be re-assessed for how they are achieving their results. As the interviews are recorded the opportunity is also given for deeper analysis of responses if particular units are found to be units performing well both from the interviewers' ratings and from potential future linkage to outcome.

This methodology is attractive for many reasons. Firstly, and perhaps of greatest importance, is the stepwise approach can be transposed onto any aspect/clinical healthcare setting. For example, using this methodology the same process can be undertaken to assess the performance of other healthcare settings, including those previously considered difficult to appraise. For example such a methodology would be transposable to assessing the performance of an acute medical service. Using the step-wise HiPer methodology and engagement with clinical units there is confidence that similar programmes could be undertaken. It is plausible that HiPer be used in any aspect of healthcare. For example, the interview schedule could practically be adapted for use in cardiac surgery or in the identification of high performing stroke units. This could be achieved by tailoring the interview schedule to the relevant speciality. With simple adaptation the methodology can be customised for both the institution and the interviewers as well. Given the relative flexibility this can be done with relative ease and minimal capital and personnel outlay. Bespoke interview schedules can be created with clinical input and these can then be used to train the bank of interviewers. Once familiar with the schedule, the interviews would take place as usual without the need for any alterations or changes.

A strength of the methodology is that at the heart of the process, all responses are blinded to the units performance. By de-coupling the HiPer methodology from the translation of any future analysis to outcome, the respondents are more likely to respond in more accurate/truthful ways. Specifically the interviewers are blinded to the unit’s performance on any measure. The units that are being assessed are also unaware of how they rate on any future linkage of measures, hence minimising reporting and reporter bias.

By allocating one rater to a unit this allows continuity of raters to engage with the participating units. Interviewers were allowed to disclose to interviewees the other member of the same unit being interviewed. This allowed both familiarity with the units and for the interviewers to state that they may not be as familiar with one aspect of care, as for example another of the future interviewers. This allowed the interviewees to respond from their own knowledge rather than guess what may happen in another domain. In contrast to selective reporting where the promotion of relationships and the creation of bonds between raters and the interviewees is thought to create a better environment for rating, this is a very subjective process and is not amenable to assessment or any form of reliability testing. However with the HiPer methodology, such familiarisation is not required and interviewees are considered equally by all raters.

Central allocation of interviewers to units removes any biases that may occur at this stage. By instructing the interviewers to arrange mutually agreeable times with the interviewers via email this allows a familiarisation and some degree of rapport. Feedback from the interviewers demonstrated that even with this informal email communication the interviewees were all primed to the process and were very willing to participate.

There was a minimal capital outlay cost involved in setting up HiPer. Interviewers and the interviewees gave up their time voluntarily. A small capital outlay was required to purchase telephone recording equipment which amounted to approximately £60 (90 US dollars). Another practical advantage was the flexibility of the programme. Once the interviewers were trained units were able to join the programme without additional resources needed. Interviewers were allocated more units once subsequent units joined the programme.

This pilot study has shown that the methodology is feasible to perform in practice. It has also shown that at some initial level the responses are rich and appear to show some consistency from the one question analysed. What remains to be seen is whether this methodology is able to correlate with the outcome measures from each unit.

The final results chapter 10 analyses all the responses from the whole project in the participating units. Overall, of the participating units there was a good response rate of 83% of the interviewees approached. The high uptake from the surgeons (100%) may reflect the fact that the surgeons were the primary investigators in each unit. It is therefore understandable that all took part. Secondly the nurses were the next highest responders (89%) which again may reflect the close working relationships that this group of respondents would have with the surgeons. In terms of the intensivists and the managers, the lower response rates may reflect the fact that these interviewees are less likely to associate with the study and its aims and thus the lower response rates (67% for both) may reflect this.

When the scores were analysed it was interesting to find that the only significant differences statistically were between the surgeons and the nurses ($p=0.023$). This may have more likely been a phenomenon of the small comparative samples sizes rather a true effect. On a per unit analysis that in fact there was only one unit (unit 4) where statistically significant differences were identified from the surgeon-nurses responses (Table 30) and this has had effect on the overall analysis.

Furthermore, this effect may have been due to the selection of the nurses. Whilst it was stipulated the nurse responder should be ward based and be involved in the day to day care of colorectal patients, one unit had a theatre nurse responder- this was unit 4. This may have affected the overall correlations as this nurse's responses were less detailed and resultantly scored lower. With such small sample sizes this may underlie why the statistical difference was seen. However, when the scores of the surgeons and the nurses were correlated (negating in part any individual differences) with one another, a good correlation with an R^2 value of 0.530 is seen. This is corroborated by the fact only one unit had significant differences statistically (unit 4) between the responses and this was the aforementioned unit. A stronger correlation is observed between the intensivists score and the surgeons. This is unsurprising as the questionnaire itself is strongly weighted to peri-operative management of patients. This may reflect the fact that the doctors would perhaps be more aware of the findings and reflect these in their answers as opposed to the nursing staff that may be less aware of the nuisances of peri-operative care.

When the audio recordings were re-analysed a wealth of information was discerned. Interestingly in the units where managers did not respond to the interviewers' invitations to participate, these were also the units that received the lowest scores for the managerial questions from those remaining clinical interviewees who did respond.

Despite the disparate units geographically and managerially there were however some generalizable findings, including the lack of senior radiological interpretation of out of hours scans, and the impact that social care and discharging patients out of hospital when medically fit pose as a common problem. Finally, most units expressed the desire to have greater assistance with recording and understanding their own results, mainly from managers that would be useful in identifying areas that needed closer inspection for improving the service offered. At present the overall feeling was that there were few resources available to surgeons to assist them in understanding/collecting their own longitudinal data for quality improvement purposes.

Interestingly, the units in the United States had the shortest length of stays and it is in these units that all units reported approximately 20% of their patients used 'interim care facilities'. These facilities allow patients who are medically fit for discharge to be transferred out to these units, thus reducing the mean length of stay for the base operating hospitals.

In summary, this final study has shown that HiPer and the methodology is feasible to undertake. There is good correlation of results between respondents, especially the medically trained ones. There is also close correlation between clinician's scores and nursing scores. The study has shown that a wealth of data not identifiable from routine analysis of databases is discernible from such an approach.

11.2 Methodological issues

There has been much discussion concerning the limitations of using HES data for clinical outcome measurement (Dixon et al., 1998). Specifically, the HES database relies upon accurate coding. In most Trusts, non-clinical staff interpret the clinical case notes and transcribe these into relevant diagnostic and procedure codes (Commission, 2009). In the studies where HES has been used, inaccurate HES data entry at source could obviously misinform regarding the differences in outcomes observed.

However, in a systematic review by Campbell and colleagues in 2001, median accuracy of HES coding varied between 91 and 69.5 per cent for diagnostic and operation codes respectively, with the overall conclusion that coding accuracy in HES data at the time was good (Campbell et al., 2001). A further limitation of the HES data set relates to its design rather than use. Tumour stage and curative intent cannot be derived from the data. The latter would have an expected impact on survival, but arguably not on 30-day in-hospital mortality and the peri-operative outcomes as measured in the studies where it has been used.

Further limitations with regards to the administrative nature of HES data are particularly important with regards to rectal cancer. Specifically, without clinical information relating to tumour height, APER rates are difficult to interpret. The validity of this metric has been questioned when it is derived from HES data (Faiz et al., 2009a). Inclusion in this context is for comparative purposes rather than judgement of the appropriateness of the procedure. Similarly with re-operation rate, clinical corroboration of re-operative need would facilitate performance appraisal. Data reliability is central to performance appraisal and benchmarking. High overall accuracy of the HES dataset has been demonstrated in a number of reviews (Burns et al., 2011a) (Campbell et al., 2001). Concerns still exist, however, that variation in coding accuracy between institutions renders performance benchmarking hazardous.

Case-mix complexity rather than differences in underlying performance is always a potential confounder of any epidemiological study. Although adjustment was made for all available parameters, there may still have been differences in the case-mix, including stage of presentation. Any differences in case mix that persisted over the studies could in part account for the observed differences in outcome between units. Although individual surgeon case mix may vary significantly within institutions, it is unlikely that the case mix presenting to Trusts varies widely across England above and beyond the co-variates considered- apart from specialist referral centres.

Surgeon experience cannot be derived from HES data. In the studies of complication management, complication rates were similar in low- and high-mortality units, suggesting that surgeon seniority is a less important factor. Seniority may impact on leak rates, oncological outcome and complications attributable to surgical technique. However, it is questionable whether the outcome from established complications is influenced by the seniority of the operating surgeon.

Finally, regarding the pilot qualitative telephone based questionnaire study, there are limitations to such an approach - where one interviewee in a unit does not participate this limits full analysis and comparisons due to the sample sizes involved. Although the process is very adaptable it is time consuming and requires flexibility by interviewers, especially when contacting non Greenwich Mean Time (GMT) based units.

11.3 Future Research

Future work would look to combine the richest sources of information available for performance appraisal beyond the peri-operative period. For example, it would be potentially possible to combine the appraisal of units using HES data and link this to national cancer registry data to provide more longitudinal outcome assessment. In other words, introducing other metrics such as disease free survival and cancer recurrence rates on a per patient basis would be highly informative.

Furthermore, creating a richer dataset by incorporating the data richness of clinical registries with the more comprehensive datasets such as HES would be potentially very useful. For example, if other co-variates such as neo-adjuvant therapy and exact cancer stage could be factored into subsequent analyses this would refine the outcome and performance results. This would also make inter-unit comparisons more meaningful.

Not only should clinical registries be considered, but also the linkage with primary care data, for example using general practitioner data on the use of anti-platelets and anti-hypertensive medication, could also give richer future results. This could help in some way to address the previously mentioned potential limitations of addressing case-mix adequately in such national studies.

With respect to the pilot study HiPer, future work could look at refining the methodology and focussing in more upon surgeon specific and nursing specific questions that are more likely to elicit focussed responses. In being limited to data governance factors that precluded performing the study in one country, such an approach may have given a more homogeneous data and responses. This in turn may have reflected in greater concordance with the hard outcome data.

11.4 Clinical Implications

The clinical implications of this work have in part already been shown. The Royal College of Surgeons Clinical Effectiveness Unit cited the work undertaken on comparing units that did and did not submit data to NBOCAP (Almoudaris et al., 2011a) as a driver for mandating submission at the national ACPGBI conference in 2012. Specifically, however, many metrics can be used to evaluate surgical performance in the peri-operative period. Aside from the 'standard' metrics of risk-adjusted mortality and length of stay that can be informative, more subtle markers need to be considered. Specifically when accounting for treatment intention (e.g. excluding those patients operated on for palliative purposes) the management of complications and the metric FTR-S seems to be an important one. Appraisal of *how* units manage their complications appears to be an important discerner of subsequent outcome and thus must be included in any appraisal of surgical performance. It has been shown that this metric is derivable from routinely collected administrative data.

In addition, performance measurement is too simplistic when it considers one or two metrics alone. Chapter 8- Benchmarking colorectal cancer resectional units in England on a panel of metrics using Hospital Episodes Statistics- demonstrated that appraising units using multiple metrics, using funnel plots, it is possible to elucidate the reasons behind the performance identified. Thus performance measurement should incorporate multiple metrics whilst adjusting for all possible co-variates and volume of cases.

Finally and most subjectively, appraisal of more subtle factors such as a unit's view of their management and how well all those that care for colorectal patients understand the roles and processes undertaken in caring for colorectal surgical patients may also be reflected in better performance. However, it is still open to debate as to what method is best for achieving this.

From what has been learnt, a high performing unit would be one with low risk-adjusted peri-operative mortality, comparable complication rates to national averages, and with the ability to rescue patients following complications. The unit would display cohesiveness, with all members of the team understanding the goals of treatment and standardisation of care where possible. These units should also be supported with the necessary high dependency care and nursing staff as well as senior-level radiological opinions out of hours.

OVERALL CONCLUSION

This thesis has explored whether high performing surgical units can be identified from available data sources. It has been found that the analysis of routinely collected and voluntarily submitted data can be enhanced with the mandating of submission and derivation of novel important metrics. Furthermore, the appraisal of other important factors needs to be complemented with a more qualitative approach as suggested by the HiPer pilot methodology. It has also been shown that such a methodology can in part relate to clinical outcome.

In future, no single metric or approach is likely to identify high performing units, given the complexity of what defines such units. What is clear, however, is that a multi-modal approach as described should be employed in any future work to identify high performance in surgery.

REFERENCES

AHRQ. 2010. *Patient Safety Indicators Overview* [Online]. Available: http://www.qualityindicators.ahrq.gov/psi_overview.htm [Accessed 18/10/2010.

AIKEN, L. H., CLARKE, S. P., SLOANE, D. M., SOCHALSKI, J. & SILBER, J. H. 2002. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA,* 288**,** 1987-93.

AL-HOMOUD, S., PURKAYASTHA, S., AZIZ, O., SMITH, J. J., THOMPSON, M. D., DARZI, A. W., STAMATAKIS, J. D. & TEKKIS, P. P. 2004. Evaluating operative risk in colorectal cancer surgery: ASA and POSSUM-based predictive models. *Surg Oncol,* 13**,** 83-92.

AL-SARIRA, A. A., DAVID, G., WILLMOTT, S., SLAVIN, J. P., DEAKIN, M. & CORLESS, D. J. 2007. Oesophagectomy practice and outcomes in England. *Br J Surg,* 94**,** 585-91.

ALLUM, W. H., GRIFFIN, S. M., WATSON, A. & COLIN-JONES, D. 2002. Guidelines for the management of oesophageal and gastric cancer. *Gut,* 50**,** v1-v23.

ALMOUDARIS, A., CLARK, S., VINCENT, C. & FAIZ, O. 2010. Establishing quality in colorectal surgery. *Colorectal Dis*.

ALMOUDARIS, A. & OMAR FAIZ 2010. Response to Thompson et al.: National Bowel Cancer Audit. *Colorectal Disease,* 9999.

ALMOUDARIS, A. B., EM. BOTTLE, A. AYLIN, P. VINCENT, C. FAIZ, O. 2011. Failure to Rescue-Surgical (FTR-S) a novel marker of surgical complication management in colorectal cancer patients. *British Journal of Surgery,* Accepted for publication.

ALMOUDARIS, A. M., BURNS, E. M., BOTTLE, A., AYLIN, P., DARZI, A. & FAIZ, O. 2011a. A colorectal perspective on voluntary submission of outcome data to clinical registries. *Br J Surg,* 98**,** 132-9.

ALMOUDARIS, A. M., BURNS, E. M., MAMIDANNA, R., BOTTLE, A., AYLIN, P., VINCENT, C. & FAIZ, O. 2011b. Value of failure to rescue as a marker of the standard of care following reoperation for complications after colorectal resection. *Br J Surg,* 98**,** 1775-83.

ALMOUDARIS, A. M., CLARK, S., VINCENT, C. & FAIZ, O. 2011c. Establishing quality in colorectal surgery. *Colorectal Dis,* 13**,** 961-73.

AMERICAN COLLEGE SURGEONS. 2010. *National Surgical Improvement Programme* [Online]. Available: https://acsnsqip.org/login/default.aspx 2010].

ANDERSON, O., NI, Z., MOLLER, H., COUPLAND, V. H., DAVIES, E. A., ALLUM, W. H. & HANNA, G. B. 2011. Hospital volume and survival in oesophagectomy and gastrectomy for cancer. *Eur J Cancer,* 47**,** 2408-14.

ANSARI, M. Z., ACKLAND, M. J., JOLLEY, D. J., CARSON, N. & MCDONALD, I. G. 1999. Inter-hospital comparison of mortality rates. *Int J Qual Health Care,* 11**,** 29-35.

ANSARI, M. Z. & COLLOPY, B. T. 1996. The risk of an unplanned return to the operating room in Australian hospitals. *Aust N Z J Surg,* 66**,** 10-3.

ARMSTRONG, B. G. 1995. Comparing standardized mortality ratios. *Ann Epidemiol,* 5**,** 60-4.

ARRIAGA, A. F., LANCASTER, R. T., BERRY, W. R., REGENBOGEN, S. E., LIPSITZ, S. R., KAAFARANI, H. M., ELBARDISSI, A. W., DESAI, P., FERZOCO, S. J., BLEDAY, R., BREEN, E., KASTRINAKIS, W. V., RUBIN, M. S. & GAWANDE, A. A. 2009. The Better Colectomy Project: Association of Evidence-Based Best-Practice Adherence Rates to Outcomes in Colorectal Surgery. *Ann Surg*.

ASSOCIATION OF COLOPROCTOLOGY OF GREAT BRITAIN AND IRELAND 2007. Guidelines for the management of colorectal cancer - 3rd edition 2007 (PDF - 3 Mb) </assets/documents/COLO_guides.pdf>

ASSOCIATION OF COLOPROCTOLOGY OF GREAT BRITAIN AND IRELAND. 2011. *National Bowel Cancer Audit Programme* [Online]. Available: www.nbocap.org.uk [Accessed 01/03/2011 2011].

ATHERLY, A., FINK, A. S., CAMPBELL, D. C., MENTZER, R. M., JR., HENDERSON, W., KHURI, S. & CULLER, S. D. 2004. Evaluating alternative risk-adjustment strategies for surgery. *Am J Surg,* 188, 566-70.

ATUL GAWANDE. 1998. *The New Yorker.com* [Online].

AYLIN, P., BOTTLE, A. & MAJEED, A. 2007a. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ,* 334, 1044.

AYLIN, P., LEES, T., BAKER, S., PRYTHERCH, D. & ASHLEY, S. 2007b. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's National Vascular Database. *Eur J Vasc Endovasc Surg,* 33, 461-5; discussion 466.

BAKER, R. 2008. *High Performing Healthcare Systems*, Longwoods Publishing Corp.

BAXTER, N. N., RICCIARDI, R., SIMUNOVIC, M., URBACH, D. R. & VIRNIG, B. A. 2010. An evaluation of the relationship between lymph node number and staging in pT3 colon cancer using population-based data. *Dis Colon Rectum,* 53, 65-70.

BEGG, C. B., CRAMER, L. D., HOSKINS, W. J. & BRENNAN, M. F. 1998. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA,* 280, 1747-51.

BENBASSAT, J. & TARAGIN, M. 2000. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Arch Intern Med,* 160, 1074-81.

BENTREM, D. J. & BRENNAN, M. F. 2005. Outcomes in oncologic surgery: does volume make a difference? *World J Surg,* 29, 1210-6.

BERENGUER, C. M., OCHSNER, M. G., JR., LORD, S. A. & SENKOWSKI, C. K. 2010. 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,* 210, 737-41, 741-3.

BERGMAN, S., FELDMAN, L. S. & BARKUN, J. S. 2006. Evaluating surgical outcomes. *Surg Clin North Am,* 86, 129-49, x.

BIRKMEYER, J. D., HAMBY, L. S., BIRKMEYER, C. M., DECKER, M. V., KARON, N. M. & DOW, R. W. 2001a. Is unplanned return to the operating room a useful quality indicator in general surgery? *Arch Surg,* 136, 405-11.

BIRKMEYER, J. D., HAMBY, L. S., BIRKMEYER, C. M., DECKER, M. V., KARON, N. M., DOW, R. W., WARSHAW, A. & ZINNER, M. 2001b. Is unplanned return to the operating room a useful quality indicator in general surgery? *Archives of Surgery,* 136, 405-410.

BIRKMEYER, J. D., SIEWERS, A. E., FINLAYSON, E. V., STUKEL, T. A., LUCAS, F. L., BATISTA, I., WELCH, H. G. & WENNBERG, D. E. 2002. Hospital volume and surgical mortality in the United States. *N Engl J Med,* 346, 1128-37.

BLOOM, N. & VAN REENEN, J. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics,* 122**,** 1351-1408.
BOROWSKI, D. W., KELLY, S. B., BRADBURN, D. M., WILSON, R. G., GUNN, A. & RATCLIFFE, A. A. 2007. Impact of surgeon volume and specialization on short-term outcomes in colorectal cancer surgery. *Br J Surg,* 94**,** 880-9.
BORZECKI, A. M., CHRISTIANSEN, C. L., CHEW, P., LOVELAND, S. & ROSEN, A. K. 2010. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. *Med Care,* 48**,** 1117-21.
BOTTLE, A., MIDDLETON, S., KALKMAN, C. J., LIVINGSTON, E. H. & AYLIN, P. 2013. Global comparators project: international comparison of hospital outcomes using administrative data. *Health Serv Res,* 48**,** 2081-100.
BRASEL, K. J., LIM, H. J., NIRULA, R. & WEIGELT, J. A. 2007. Length of stay: an appropriate quality measure? *Arch Surg,* 142**,** 461-5; discussion 465-6.
BRISINDA, G., VANELLA, S., CADEDDU, F. & MAZZEO, P. 2009. Colonic Anastomotic Leak: Risk Factors, Diagnosis, and Treatment. *Journal of the American College of Surgeons,* 208**,** 1152-1153.
BROMAGE, S. J. & CUNLIFFE, W. J. 2007. Validation of the CR-POSSUM risk-adjusted scoring system for major colorectal cancer surgery in a single center. *Dis Colon Rectum,* 50**,** 192-6.
BROOK, R. H., MCGLYNN, E. A. & CLEARY, P. D. 1996. Quality of health care. Part 2: measuring quality of care. *N Engl J Med,* 335**,** 966-70.
BRUNELLI, A., XIUME, F., AL REFAI, M., SALATI, M., MARASCO, R. & SABBATINI, A. 2006. Risk-adjusted morbidity, mortality and failure-to-rescue models for internal provider profiling after major lung resection. *Interact Cardiovasc Thorac Surg,* 5**,** 92-6.
BURNS, E. M., BOTTLE, A., AYLIN, P., CLARK, S. K., TEKKIS, P. P., DARZI, A., NICHOLLS, R. J. & FAIZ, O. 2011a. Volume analysis of outcome following restorative proctocolectomy. *Br J Surg,* 98**,** 408-17.
BURNS, E. M., BOTTLE, A., AYLIN, P., DARZI, A., NICHOLLS, R. J. & FAIZ, O. 2011b. Variation in reoperation after colorectal surgery in England as an indicator of surgical performance: retrospective analysis of Hospital Episode Statistics. *BMJ,* 343**,** d4836.
BURNS, E. M., RIGBY, E., MAMIDANNA, R., BOTTLE, A., AYLIN, P., ZIPRIN, P. & FAIZ, O. D. 2011c. Systematic review of discharge coding accuracy. *J Public Health (Oxf)*.
BURNS, E. M., RIGBY, E., MAMIDANNA, R., BOTTLE, A., AYLIN, P., ZIPRIN, P. & FAIZ, O. D. 2012. Systematic review of discharge coding accuracy. *Journal of Public Health,* 34**,** 138-148.
BUTLER-WU, S. M., BURNS, E. M., POTTINGER, P. S., MAGARET, A. S., RAKEMAN, J. L., MATSEN, F. A., 3RD & COOKSON, B. T. 2011. Optimization of periprosthetic culture for diagnosis of Propionibacterium acnes prosthetic joint infection. *J Clin Microbiol,* 49**,** 2490-5.
CAMPBELL, S. E., CAMPBELL, M. K., GRIMSHAW, J. M. & WALKER, A. E. 2001. A systematic review of discharge coding accuracy. *J Public Health Med,* 23**,** 205-11.
CARE QUALITY COMMISSION. 2009a. Better Metrics Project 2007 v8. Available: http://www.cqc.org.uk/_db/_documents/List_of_metrics-The_Better_Metrics_project_v8-Nov07.pdf [Accessed 9/12/2009].
CARE QUALITY COMMISSION. 2009b. *http://www.cqc.org.uk/* [Online]. Available: http://www.cqc.org.uk/ [Accessed 20/12/2009.

CARE., N. R. 2010. *NHS Atlas of Variation Series* [Online]. Available: http://www.rightcare.nhs.uk/index.php/nhs-atlas/ [Accessed 28/08/2012 2012].
CENTER FOR HEALTH POLICY STUDIES 1995. Understanding and choosing clinical performance measures for quality improvement: development of typology: final report. *In:* (AHRQ), A. F. H. R. A. Q. (ed.). Harvard School of Public Health.
CHAMBERS, M. & CLARKE, A. 1990. Measuring readmission rates. *BMJ,* 301**,** 1134-6.
CHANG, G. J., RODRIGUEZ-BIGAS, M. A., SKIBBER, J. M. & MOYER, V. A. 2007. Lymph node evaluation and survival after curative resection of colon cancer: systematic review. *J Natl Cancer Inst,* 99**,** 433-41.
CHARLSON, M. E., POMPEI, P., ALES, K. L. & MACKENZIE, C. R. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis,* 40**,** 373-83.
CHASSIN, M. R., HANNAN, E. L. & DEBUONO, B. A. 1996. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med,* 334**,** 394-8.
CHOWDHURY, M. M., DAGASH, H. & PIERRO, A. 2007. A systematic review of the impact of volume of surgery and specialization on patient outcome. *Br J Surg,* 94**,** 145-61.
CHUWA, E. W. & SEOW-CHOEN, F. 2006. Outcomes for abdominoperineal resections are not worse than those of anterior resections. *Dis Colon Rectum,* 49**,** 41-9.
CLAVIEN, P. A., SANABRIA, J. R. & STRASBERG, S. M. 1992. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery,* 111**,** 518-26.
COHEN, M. E., BILIMORIA, K. Y., KO, C. Y., RICHARDS, K. & HALL, B. L. 2009a. Variability in length of stay after colorectal surgery: assessment of 182 hospitals in the national surgical quality improvement program. *Ann Surg,* 250**,** 901-7.
COHEN, M. E., DIMICK, J. B., BILIMORIA, K. Y., KO, C. Y., RICHARDS, K. & HALL, B. L. 2009b. Risk adjustment in the American College of Surgeons National Surgical Quality Improvement Program: a comparison of logistic versus hierarchical modeling. *J Am Coll Surg,* 209**,** 687-93.
COLLEGE., T. T. O. D. 1996. The Dartmouth Atlas of Health Care . Available: http://www.dartmouthatlas.org/downloads/atlases/96Atlas.pdf.
COMISSION., I. 2012. Patient Reported Outcome Measures Final report.
COMMISSION, A. 2009. *PbR data assurance framework 2008/9* [Online]. Available: http://www.audit-commission.gov.uk/health/nationalstudies/pbr/pbrdataassuranceframework200809/Pages/default_copy.aspx [Accessed 10/11/2009 2009].
COOPERBERG, M. R., BIRKMEYER, J. D. & LITWIN, M. S. 2009. Defining high quality health care. *Urol Oncol,* 27**,** 411-6.
COPELAND, G. P., JONES, D. & WALTERS, M. 1991. POSSUM: a scoring system for surgical audit. *Br J Surg,* 78**,** 355-60.
COPYRIGHT, C. 2001. *The Bristol Royal Infirmary Inquiry* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20090811143745/http://www.bristol-inquiry.org.uk/final_report/report/index.htm.
CORRIGAN, J. M. & KAZANDJIAN, V. A. 1991. Characteristics of multiple admissions. *J Am Med Rec Assoc,* 62**,** 37-47.
COURTNEY, E. D., ANKRETT, S. & MCCOLLUM, P. T. 2003. 28-Day emergency surgical re-admission rates as a clinical indicator of performance. *Ann R Coll Surg Engl,* 85**,** 75-8.

CROWN 2002. High quality care for all: NHS Next Stage Review final report. *In:* KBE, P. T. L. D. O. D. (ed.). London.
DALEY, J., HENDERSON, W. G. & KHURI, S. F. 2001. Risk-adjusted surgical outcomes. *Annu Rev Med,* 52**,** 275-87.
DAVENPORT, D. L., BOWE, E. A., HENDERSON, W. G., KHURI, S. F. & MENTZER, R. M., JR. 2006. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists Physical Status Classification (ASA PS) levels. *Ann Surg,* 243**,** 636-41; discussion 641-4.
DAWSON, D. G., H. 2005. Delveloping new approaches to measureing NHS outputs and productivity. *Centre for Health Economics* [Online]. [Accessed October 2009].
DEAPRTMENT OF HEALTH 2010. Report from the Steering Group for the National Review of the Hospital Standardised Mortality Ratio. London.
DEN DULK, M., MARIJNEN, C. A. M., COLLETTE, L., PUTTER, H., PAAHLMAN, L., FOLKESSON, J., BOSSET, J. F., RODEL, C., BUJKO, K. & VAN DE VELDE, C. J. H. 2009. Multicentre analysis of oncological and survival outcomes following anastomotic leakage after rectal cancer surgery. *British Journal of Surgery,* 96**,** 1066-1075.
DEPARTMENT OF HEALTH. 2010. *Hospital Activity Statistics* [Online]. Available: http://www.performance.doh.gov.uk/hospitalactivity/data_requests/index.htm [Accessed 29/09/2010.
DEPARTMENT OF HEALTH SOCIAL SERVICES AND PUBLIC SAFETY (NORTHERN IRELAND). 2011. *Hospital Statistics* [Online]. Available: http://www.dhsspsni.gov.uk/index/stats_research/stats-activity_stats-2.htm [Accessed 04/05/2011 2011].
DESCH, C. E., MCNIFF, K. K., SCHNEIDER, E. C., SCHRAG, D., MCCLURE, J., LEPISTO, E., DONALDSON, M. S., KAHN, K. L., WEEKS, J. C., KO, C. Y., STEWART, A. K. & EDGE, S. B. 2008. American Society of Clinical Oncology/National Comprehensive Cancer Network Quality Measures. *J Clin Oncol,* 26**,** 3631-7.
DILLMAN, R. O., AARON, K., HEINEMANN, F. S. & MCCLURE, S. E. 2009. Identification of 12 or more lymph nodes in resected colon cancer specimens as an indicator of quality performance. *Cancer,* 115**,** 1840-8.
DIMICK, J. B. & BIRKMEYER, J. D. 2008. Ranking hospitals on surgical quality: does risk-adjustment always matter? *J Am Coll Surg,* 207**,** 347-51.
DIMICK, J. B., CHEN, S. L., TAHERI, P. A., HENDERSON, W. G., KHURI, S. F. & CAMPBELL, D. A., JR. 2004. Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement Program. *J Am Coll Surg,* 199**,** 531-7.
DIMICK, J. B., COWAN, J. A., JR., UPCHURCH, G. R., JR. & COLLETTI, L. M. 2003. Hospital volume and surgical outcomes for elderly patients with colorectal cancer in the United States. *J Surg Res,* 114**,** 50-6.
DIMICK, J. B., OSBORNE, N. H., HALL, B. L., KO, C. Y. & BIRKMEYER, J. D. 2010a. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg,* 210**,** 503-8.
DIMICK, J. B., STAIGER, D. O. & BIRKMEYER, J. D. 2010b. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res,* 45**,** 1614-29.
DINDO, D., DEMARTINES, N. & CLAVIEN, P. A. 2004. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg,* 240**,** 205-13.

DIXON, E., DATTA, I., SUTHERLAND, F. R. & VAUTHEY, J. N. 2009. Blood loss in surgical oncology: Neglected quality indicator? *Journal of Surgical Oncology,* 99**,** 508-512.

DIXON, J., SANDERSON, C., ELLIOTT, P., WALLS, P., JONES, J. & PETTICREW, M. 1998. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med,* 20**,** 63-9.

DOLAN, S. A., JARMAN, B., BAJEKAL, M., DAVIES, P. M. & HART, D. 1995. Measuring disadvantage: changes in the underprivileged area, Townsend, and Carstairs scores 1981-91. *J Epidemiol Community Health,* 49 Suppl 2**,** S30-3.

DONABEDIAN, A. Explorations in quality assessment and monitoring. *Health Administration Press,* Vol. 1.

DONABEDIAN, A. 1966. Evaluating the quality of medical care. *Milbank Mem Fund Q,* 44**,** Suppl:166-206.

DREISLER, E., SCHOU, L. & ADAMSEN, S. 2001. Completeness and accuracy of voluntary reporting to a national case registry of laparoscopic cholecystectomy. *Int J Qual Health Care,* 13**,** 51-5.

DRESEN, R. C., PETERS, E. E., RUTTEN, H. J., NIEUWENHUIJZEN, G. A., DEMEYERE, T. B., VAN DEN BRULE, A. J., KESSELS, A. G., BEETS-TAN, R. G., VAN KRIEKEN, J. H. & NAGTEGAAL, I. D. 2009. Local recurrence in rectal cancer can be predicted by histopathological factors. *Eur J Surg Oncol,* 35**,** 1071-7.

EBERL, T., JAGODITSCH, M., KLINGLER, A. & TSCHMELITSCH, J. 2008. Risk factors for anastomotic leakage after resection for rectal cancer. *American Journal of Surgery,* 196**,** 592-598.

ELIXHAUSER, A. & HALPERN, M. T. 1999. Economic evaluations of gastric and pancreatic cancer. *Hepatogastroenterology,* 46**,** 1206-13.

ELSEVIER. 2010. *The crude and adjusted rates in epidemiology* [Online]. Available: http://www.scitopics.com/The_crude_and_adjusted_rates_in_epidemiology.html

ENGEL, J., KERR, J., ECKEL, R., GUNTHER, B., HEISS, M., HEITLAND, W., SIEWERT, J. R., JAUCH, K. W. & HOLZEL, D. 2005. Influence of hospital volume on local recurrence and survival in a population sample of rectal cancer patients. *European Journal of Surgical Oncology,* 31**,** 512-520.

EORTC. 2009. *EORTC QLQ CR38* [Online]. Available: http://groups.eortc.be/qol/.

ESCOBAR, G. J., GREENE, J. D., SCHEIRER, P., GARDNER, M. N., DRAPER, D. & KIPNIS, P. 2008. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care,* 46**,** 232-9.

EUROQOL. 2009. *http://www.euroqol.org/* [Online]. Available: http://www.euroqol.org/ [Accessed 20/12/2009.

EVANS, M. D., BARTON, K., REES, A., STAMATAKIS, J. D. & KARANDIKAR, S. S. 2008. The impact of surgeon and pathologist on lymph node retrieval in colorectal cancer and its impact on survival for patients with Dukes' stage B disease. *Colorectal Dis,* 10**,** 157-64.

FAIZ, O., CLARK, J., BROWN, T., BOTTLE, A., ANTONIOU, A., FARRANDS, P., DARZI, A. & AYLIN, P. 2008a. Traditional and laparoscopic appendectomy in adults: outcomes in English NHS hospitals between 1996 and 2006. *Ann Surg,* 248**,** 800-6.

FAIZ, O., HAJI, A., BOTTLE, A., CLARK, S., DARZI, A. & AYLIN, P. 2010a. Elective colonic surgery for cancer in the elderly: an investigation into

postoperative mortality in English NHS hospitals between 1996 and 2007. *Colorectal Dis*.
FAIZ, O., HAJI, A., BURNS, E., BOTTLE, A., KENNEDY, R. & AYLIN, P. 2010b. Hospital stay amongst patients undergoing major elective colorectal surgery: predicting prolonged stay and readmissions in NHS hospitals. *Colorectal Dis*.
FAIZ, O., SMITH, J. & CLARK, S. 2009a. Unacceptable variation in abdominoperineal excision rates for rectal cancer. *Gut,* 58**,** 150; author reply 150-2.
FAIZ, O., WARUSAVITARNE, J., BOTTLE, A., TEKKIS, P. P., CLARK, S. K., DARZI, A. W. & AYLIN, P. 2010c. Nonelective excisional colorectal surgery in English National Health Service Trusts: a study of outcomes from Hospital Episode Statistics Data between 1996 and 2007. *J Am Coll Surg,* 210**,** 390-401.
FAIZ, O., WARUSAVITARNE, J., BOTTLE, A., TEKKIS, P. P., DARZI, A. W. & KENNEDY, R. H. 2009b. Laparoscopically assisted vs. open elective colonic and rectal resection: a comparison of outcomes in English National Health Service Trusts between 1996 and 2006. *Dis Colon Rectum,* 52**,** 1695-704.
FAIZ, O. D., BROWN, T. J., COLUCCI, G., GROVER, M. & CLARK, S. K. 2008b. Trends in colorectal day case surgery in NHS Trusts between 1998 and 2005. *Colorectal Dis,* 10**,** 935-42.
FEARON, K. C., LJUNGQVIST, O., VON MEYENFELDT, M., REVHAUG, A., DEJONG, C. H., LASSEN, K., NYGREN, J., HAUSEL, J., SOOP, M., ANDERSEN, J. & KEHLET, H. 2005. Enhanced recovery after surgery: a consensus review of clinical care for patients undergoing colonic resection. *Clin Nutr,* 24**,** 466-77.
FEDELI, U., BROCCO, S., ALBA, N., ROSATO, R. & SPOLAORE, P. 2007. The choice between different statistical approaches to risk-adjustment influenced the identification of outliers. *J Clin Epidemiol,* 60**,** 858-62.
FISHER, W. H. & ALTAFFER, F. B. 1992. Inpatient Length of Stay Measures - Statistical and Conceptual Issues. *Administration and Policy in Mental Health,* 19**,** 311-320.
GARCIA-GRANERO, E., FAIZ, O., MUNOZ, E., FLOR, B., NAVARRO, S., FAUS, C., GARCIA-BOTELLO, S. A., LLEDO, S. & CERVANTES, A. 2009. Macroscopic assessment of mesorectal excision in rectal cancer: a useful tool for improving quality control in a multidisciplinary team. *Cancer,* 115**,** 3400-11.
GAROUT, M., TILNEY, H. S., TEKKIS, P. P. & AYLIN, P. 2008. Comparison of administrative data with the Association of Coloproctology of Great Britain and Ireland (ACPGBI) colorectal cancer database. *Int J Colorectal Dis,* 23**,** 155-63.
GASTINGER, I., MARUSCH, F., STEINERT, R., WOLFF, S., KOECKERLING, F. & LIPPERT, H. 2005. Protective defunctioning stoma in low anterior resection for rectal carcinoma. *British Journal of Surgery,* 92**,** 1137-1142.
GENDALL, K. A., KENNEDY, R. R., WATSON, A. J. & FRIZELLE, F. A. 2007. The effect of epidural analgesia on postoperative outcome after colorectal surgery. *Colorectal Dis,* 9**,** 584-98; discussion 598-600.
GENERAL MEDICAL COUNCIL. 2011. *Revalidation* [Online]. Available: http://www.gmc-uk.org/doctors/revalidation.asp [Accessed 03/05/2011 2011].
GHAFERI, A. A., BIRKMEYER, J. D. & DIMICK, J. B. 2009a. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg,* 250**,** 1029-34.
GHAFERI, A. A., BIRKMEYER, J. D. & DIMICK, J. B. 2009b. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med,* 361**,** 1368-75.

GHAFERI, A. A., OSBORNE, N. H. & DIMICK, J. B. 2009c. Does Voluntary Reporting Bias Hospital Quality Rankings? *J Surg Res*.
GILLION, J. F. 2005. [Is crude postoperative mortality rate a relevant criterion of the efficiency of a surgical team? Prospective study of 11,756 patients' postoperative course]. *Ann Chir,* 130**,** 400-6.
GLICKMAN, S. W., BAGGETT, K. A., KRUBERT, C. G., PETERSON, E. D. & SCHULMAN, K. A. 2007. Promoting quality: the health-care organization from a management perspective. *Int J Qual Health Care,* 19**,** 341-8.
GLOVER, J. A. 1938. The Incidence of Tonsillectomy in School Children: (Section of Epidemiology and State Medicine). *Proc R Soc Med,* 31**,** 1219-36.
GMC. 2009. *Revalidation: Information for Doctors*

[Online]. London: GMC. Available: http://www.gmc-uk.org/GMC_Revalidation_FAQs.pdf_snapshot.pdf 2009].
GOLDMAN, D. A. & BRENDER, J. D. 2000. Are standardized mortality ratios valid for public health data analysis? *Stat Med,* 19**,** 1081-8.
GOMEZ-OUTES, A., TERLEIRA-FERNANDEZ, A. I., SUAREZ-GEA, M. L. & VARGAS-CASTRILLON, E. 2012. Dabigatran, rivaroxaban, or apixaban versus enoxaparin for thromboprophylaxis after total hip or knee replacement: systematic review, meta-analysis, and indirect treatment comparisons. *BMJ,* 344**,** e3675.
GOUVAS, N., TAN, E., WINDSOR, A., XYNOS, E. & TEKKIS, P. P. 2009a. Fast-track vs standard care in colorectal surgery: a meta-analysis update. *Int J Colorectal Dis,* 24**,** 1119-31.
GOUVAS, N., TSIAOUSSIS, J., PECHLIVANIDES, G., TZORTZINIS, A., DERVENIS, C., AVGERINOS, C. & XYNOS, E. 2009b. Quality of surgery for rectal carcinoma: comparison between open and laparoscopic approaches. *American Journal of Surgery,* 198**,** 702-708.
HANNAN, E. L., KILBURN, H., JR., RACZ, M., SHIELDS, E. & CHASSIN, M. R. 1994. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA,* 271**,** 761-6.
HANNAN, E. L., WU, C., RYAN, T. J., BENNETT, E., CULLIFORD, A. T., GOLD, J. P., HARTMAN, A., ISOM, O. W., JONES, R. H., MCNEIL, B., ROSE, E. A. & SUBRAMANIAN, V. A. 2003. Do hospitals and surgeons with higher coronary artery bypass graft surgery volumes still have lower risk-adjusted mortality rates? *Circulation,* 108**,** 795-801.
HARLING, H., BULOW, S., MOLLER, L. N., JORGENSEN, T. & DANISH COLORECTAL CANCER, G. 2005. Hospital volume and outcome of rectal cancer surgery in Denmark 1994-99. *Colorectal Disease,* 7**,** 90-5.
HAWKES, N. 2010. Patient coding and the ratings game. *BMJ,* 340**,** c2153.
HAYNES, A. B., WEISER, T. G., BERRY, W. R., LIPSITZ, S. R., BREIZAT, A. H., DELLINGER, E. P., HERBOSA, T., JOSEPH, S., KIBATALA, P. L., LAPITAN, M. C., MERRY, A. F., MOORTHY, K., REZNICK, R. K., TAYLOR, B. & GAWANDE, A. A. 2009. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med,* 360**,** 491-9.
HEALTH, D. O. 1999. *Department of Health. Health Act 1999.* [Online]. London. Available: http://www.rcpath.org/resources/pdf/clinical.pdf [Accessed 20/11/2009 2009].
HEALTH, D. O. 2010. Liberating the NHS: Trasparancy in outcomes – a framework for the NHS.
HEALTH., D. O. BRI Inquiry Final Report Summary. Available: http://www.bristol-inquiry.org.uk/final_report/Summary.pdf.

HERNANZ, F., GARCIA-SOMACARRERA, E. & FERNANDEZ, F. 2009. The assessment of lymph nodes missed in mesenteric tissue after standard dissection of colorectal cancer specimens. *Colorectal Dis*.

HESONLINE. 2010. *Hospital Episode Statistic* [Online]. Available: http://www.hesonline.nhs.uk/.

HIBBARD, J. H., STOCKARD, J. & TUSLER, M. 2005. Hospital performance reports: impact on quality, market share, and reputation. *Health Aff (Millwood),* 24**,** 1150-60.

HOFMANN, D. A., JACOBS, R. & LANDY, F. 1995. High reliability process industries: Individual, micro, and macro organizational influences on safety performance. *Journal of Safety Research,* 26**,** 131-149.

HORWITZ, L. I., CUNY, J. F., CERESE, J. & KRUMHOLZ, H. M. 2007. Failure to rescue: validation of an algorithm using administrative data. *Med Care,* 45**,** 283-7.

HOSPITAL, S. 2012. *Shouldice Hospital home page* [Online]. Available: http://www.shouldice.com/our_history.htm [Accessed 28/08/2012 2012].

HUH, J. W., PARK, Y. A. & SOHN, S. K. 2007. A diverting stoma is not necessary when performing a handsewn coloanal anastomosis for lower rectal cancer. *Diseases of the Colon and Rectum,* 50**,** 1040-1046.

HYMAN, N. H., OSLER, T., CATALDO, P., BURNS, E. H. & SHACKFORD, S. R. 2009. Anastomotic Leaks after Bowel Resection: What Does Peer Review Teach Us about the Relationship to Postoperative Mortality? *Journal of the American College of Surgeons,* 208**,** 48-52.

HYUK BAIK, S., KYU KIM, N., YOUNG LEE, K., KOOK SOHN, S., HWAN CHO, C., JIN KIM, M., KIM, H. & SHINN, R. K. 2008. Factors influencing pathologic results after total mesorectal excision for rectal cancer: Analysis of consecutive 100 cases. *Annals of Surgical Oncology,* 15**,** 721-728.

IEZONNI, L. 2003. Risk Adjustment for Measuring Health Care Outcomes. *AcademyHealth/HAP,* 3rd ed:.

INSTITUTE OF MEDICINE 1990. Committee to Design a Strategy for Quality Review and Assurance in Medicine. *Medicare: a strategy for quality assurance / Committee to Design a Strategy for Quality Review and Assurance in Medicare.*

IVERSEN, L. H., HARLING, H., LAURBERG, S. & WILLE-JORGENSEN, P. 2007. Influence of caseload and surgical speciality on outcome following surgery for colorectal cancer: A review of evidence. Part 1: Short-term outcome. *Colorectal Disease,* 9**,** 28-37.

JARMAN, B., BOTTLE, A., AYLIN, P. & TAYLOR, R. 2002. Mortality rates. Dead confusing. *Health Serv J,* 112**,** 26-7.

JARMAN, B., GAULT, S., ALVES, B., HIDER, A., DOLAN, S., COOK, A., HURWITZ, B. & IEZZONI, L. I. 1999. Explaining differences in English hospital death rates using routinely collected data. *BMJ,* 318**,** 1515-20.

JEYARAJAH, S., SUTTON, C. D., MILLER, A. S. & HEMINGWAY, D. 2007. Factors that influence the adequacy of total mesorectal excision for rectal cancer. *Colorectal Dis,* 9**,** 808-15.

JOHNSON, P. M., MALATJALIAN, D. & PORTER, G. A. 2002. Adequacy of nodal harvest in colorectal cancer: a consecutive cohort study. *J Gastrointest Surg,* 6**,** 883-88; discussion 889-90.

JONES, H. J., COGGINS, R., LAFUENTE, J. & DE COSSART, L. 1999. Value of a surgical high-dependency unit. *Br J Surg,* 86**,** 1578-82.

KAPITEIJN, E., MARIJNEN, C. A., NAGTEGAAL, I. D., PUTTER, H., STEUP, W. H., WIGGERS, T., RUTTEN, H. J., PAHLMAN, L., GLIMELIUS, B., VAN

KRIEKEN, J. H., LEER, J. W. & VAN DE VELDE, C. J. 2001. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med,* 345**,** 638-46.

KAPITEIJN, E., PUTTER, H. & VAN DE VELDE, C. J. 2002. Impact of the introduction and training of total mesorectal excision on recurrence and survival in rectal cancer in The Netherlands. *Br J Surg,* 89**,** 1142-9.

KATZ, J. N., LOSINA, E., BARRETT, J., PHILLIPS, C. B., MAHOMED, N. N., LEW, R. A., GUADAGNOLI, E., HARRIS, W. H., POSS, R. & BARON, J. A. 2001. Association between hospital and surgeon procedure volume and outcomes of total hip replacement in the United States medicare population. *J Bone Joint Surg Am,* 83-A**,** 1622-9.

KEOGH, B., SPIEGELHALTER, D., BAILEY, A., ROXBURGH, J., MAGEE, P. & HILTON, C. 2004. The legacy of Bristol: public disclosure of individual surgeons' results. *BMJ,* 329**,** 450-4.

KHOO, C. K., VICKERY, C. J., FORSYTH, N., VINALL, N. S. & EYRE-BROOK, I. A. 2007. A prospective randomized controlled trial of multimodal perioperative management protocol in patients undergoing elective colorectal resection for cancer. *Ann Surg,* 245**,** 867-72.

KHURI, S. F., DALEY, J., HENDERSON, W., HUR, K., GIBBS, J. O., BARBOUR, G., DEMAKIS, J., IRVIN, G., 3RD, STREMPLE, J. F., GROVER, F., MCDONALD, G., PASSARO, E., JR., FABRI, P. J., SPENCER, J., HAMMERMEISTER, K. & AUST, J. B. 1997. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg,* 185**,** 315-27.

KHWANNIMIT, B. & BHURAYANONTACHAI, R. 2009a. The performance and customization of SAPS 3 admission score in a Thai medical intensive care unit. *Intensive Care Med.*

KHWANNIMIT, B. & BHURAYANONTACHAI, R. 2009b. The performance of customised APACHE II and SAPS II in predicting mortality of mixed critically ill patients in a Thai medical intensive care unit. *Anaesth Intensive Care,* 37**,** 784-90.

KILLEEN, S. D., ANDREWS, E. J., REDMOND, H. P. & FULTON, G. J. 2007. Provider volume and outcomes for abdominal aortic aneurysm repair, carotid endarterectomy, and lower extremity revascularization procedures. *J Vasc Surg,* 45**,** 615-26.

KING, P. M., BLAZEBY, J. M., EWINGS, P., FRANKS, P. J., LONGMAN, R. J., KENDRICK, A. H., KIPLING, R. M. & KENNEDY, R. H. 2006a. Randomized clinical trial comparing laparoscopic and open surgery for colorectal cancer within an enhanced recovery programme. *Br J Surg,* 93**,** 300-8.

KING, P. M., BLAZEBY, J. M., EWINGS, P., LONGMAN, R. J., KIPLING, R. M., FRANKS, P. J., SHEFFIELD, J. P., EVANS, L. B., SOULSBY, M., BULLEY, S. H. & KENNEDY, R. H. 2006b. The influence of an enhanced recovery programme on clinical outcomes, costs and quality of life after surgery for colorectal cancer. *Colorectal Dis,* 8**,** 506-13.

KLUGMAN, R., ALLEN, L., BENJAMIN, E. M., FITZGERALD, J. & ETTINGER, W., JR. 2010. Mortality rates as a measure of quality and safety, "caveat emptor". *Am J Med Qual,* 25**,** 197-201.

KO, C. Y., CHANG, J. T., CHAUDHRY, S. & KOMINSKI, G. 2002. Are high-volume surgeons and hospitals the most important predictors of in-hospital outcome for colon cancer resection? *Surgery,* 132**,** 268-73.

KROON, H. M., BRESLAU, P. J. & LARDENOYE, J. W. 2007. Can the incidence of unplanned reoperations be used as an indicator of quality of care in surgery? *Am J Med Qual,* 22**,** 198-202.

LA PORTE, T. R. 1996. High Reliability Organizations: Unlikely, Demanding and At Risk. *Journal of Contingencies and Crisis Management,* 4**,** 60-71.

LANGE, M. M., DEN DULK, M., BOSSEMA, E. R., MAAS, C. P., PEETERS, K. C. M. J., RUTTEN, H. J., KLEIN KRANENBARG, E., MARIJNEN, C. A. M., VAN DE VELDE, C. J. H. & COOPERATIVE CLINICAL INVESTIGATORS OF THE DUTCH TOTAL MESORECTAL EXCISION, T. 2007. Risk factors for faecal incontinence after rectal cancer treatment. *British Journal of Surgery,* 94**,** 1278-84.

LARSON, D. W., MARCELLO, P. W., LARACH, S. W., WEXNER, S. D., PARK, A., MARKS, J., SENAGORE, A. J., THORSON, A. G., YOUNG-FADOK, T. M., GREEN, E., SARGENT, D. J. & NELSON, H. 2008. Surgeon volume does not predict outcomes in the setting of technical credentialing: results from a randomized trial in colon cancer. *Ann Surg,* 248**,** 746-50.

LASSEN, K., SOOP, M., NYGREN, J., COX, P. B., HENDRY, P. O., SPIES, C., VON MEYENFELDT, M. F., FEARON, K. C., REVHAUG, A., NORDERVAL, S., LJUNGQVIST, O., LOBO, D. N. & DEJONG, C. H. 2009. Consensus review of optimal perioperative care in colorectal surgery: Enhanced Recovery After Surgery (ERAS) Group recommendations. *Arch Surg,* 144**,** 961-9.

LAW, W. L., CHOI, H. K., LEE, Y. M., HO, J. W. C. & SETO, C. L. 2007. Anastomotic leakage is associated with poor long-term outcome in patients after curative colorectal resection for malignancy. *Journal of Gastrointestinal Surgery,* 11**,** 8-15.

LEAPFROG GROUP. 2009. *The Leapfrog Group Survey* [Online]. Available: http://www.leapfroggroup.org/ [Accessed 20/12/2009.

LEBEAU, G., YANAGA, K., MARSH, J. W., TZAKIS, A. G., MAKOWKA, L., GORDON, R. D., TODO, S., STIEBER, A. C., IWATSUKI, S. & STARZL, T. E. 1990. Analysis of surgical complications after 397 hepatic transplantations. *Surg Gynecol Obstet,* 170**,** 317-22.

LEE, J. H., PARK, D. J., KIM, H. H., LEE, H. J. & YANG, H. K. 2011. Comparison of complications after laparoscopy-assisted distal gastrectomy and open distal gastrectomy for gastric cancer using the Clavien-Dindo classification. *Surg Endosc*.

LEONARDI, M. J., MCGORY, M. L. & KO, C. Y. 2007. Publicly available hospital comparison web sites: determination of useful, valid, and appropriate information for comparing surgical quality. *Arch Surg,* 142**,** 863-8; discussion 868-9.

LILFORD, R., MOHAMMED, M. A., SPIEGELHALTER, D. & THOMSON, R. 2004. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet,* 363**,** 1147-54.

LINDENAUER, P. K., REMUS, D., ROMAN, S., ROTHBERG, M. B., BENJAMIN, E. M., MA, A. & BRATZLER, D. W. 2007. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med,* 356**,** 486-96.

LONDON SCHOOL OF HYGEINE AND TROPICAL MEDICINE. 2005. Available: http://www.lshtm.ac.uk/php/documents/PROMs%20Final%20Report%20Dec%2005.pdf 6

LONGO, W. E., VIRGO, K. S., JOHNSON, F. E., OPRIAN, C. A., VERNAVA, A. M., WADE, T. P., PHELAN, M. A., HENDERSON, W. G., DALEY, J. &

KHURI, S. F. 2000. Risk factors for morbidity and mortality after colectomy for colon cancer. *Dis Colon Rectum,* 43**,** 83-91.
LUFT, H. S., BUNKER, J. P. & ENTHOVEN, A. C. 1979. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med,* 301**,** 1364-9.
LUSTOSA, S. A. D. S., MATOS, D., ATALLAH, A. N. & CASTRO, A. A. 2002. Stapled versus handsewn methods for colorectal anastomosis surgery: a systematic review of randomized controlled trials. *Sao Paulo Medical Journal = Revista Paulista de Medicina,* 120**,** 132-6.
LYTLE, B. L., FRAULO, E. S., MULGUND, J., MILLER, V. A., ROE, M. T., SMITH, S. C., JR., GIBLER, W. B., OHMAN, E. M. & PETERSON, E. D. 2007. What aspects of hospital culture influence quality? *Crit Pathw Cardiol,* 6**,** 145-9.
MAMIDANNA, R., ALMOUDARIS, A. M. & FAIZ, O. 2010. Variability in length of stay after colorectal surgery: assessment of 182 hospitals in the national surgical quality improvement program. *Ann Surg,* 252**,** 891-2; author reply 892.
MAMIDANNA, R., ALMOUDARIS, A. M. & FAIZ, O. 2012. Is 30-day mortality an appropriate measure of risk in elderly patients undergoing elective colorectal resection? *Colorectal Disease,* 14**,** 1175-1182.
MANGANO, D. T., LAYUG, E. L., WALLACE, A. & TATEO, I. 1996. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. Multicenter Study of Perioperative Ischemia Research Group. *N Engl J Med,* 335**,** 1713-20.
MARKAR, S. R., KARTHIKESALINGAM, A., THRUMURTHY, S. & LOW, D. E. 2011. Volume-Outcome Relationship in Surgery for Esophageal malignancy: Systematic Review and Meta-analysis 2000-2011. *J Gastrointest Surg*.
MARKL, B., KERWEL, T. G., WAGNER, T., ANTHUBER, M. & ARNHOLDT, H. M. 2007. Methylene blue injection into the rectal artery as a simple method to improve lymph node harvest in rectal cancer. *Mod Pathol,* 20**,** 797-801.
MARR, R., BIRBECK, K., GARVICAN, J., MACKLIN, C. P., TIFFIN, N. J., PARSONS, W. J., DIXON, M. F., MAPSTONE, N. P., SEBAG-MONTEFIORE, D., SCOTT, N., JOHNSTON, D., SAGAR, P., FINAN, P. & QUIRKE, P. 2005. The modern abdominoperineal excision: the next challenge after total mesorectal excision. *Ann Surg,* 242**,** 74-82.
MARSHALL, M. N., SHEKELLE, P. G., DAVIES, H. T. & SMITH, P. C. 2003. Public reporting on quality in the United States and the United Kingdom. *Health Aff (Millwood),* 22**,** 134-48.
MASLEKAR, S., SHARMA, A., MACDONALD, A., GUNN, J., MONSON, J. R. & HARTLEY, J. E. 2007. Mesorectal grades predict recurrences after curative resection for rectal cancer. *Dis Colon Rectum,* 50**,** 168-75.
MAUGHAN, N. J., MORRIS, E., FORMAN, D. & QUIRKE, P. 2007. The validity of the Royal College of Pathologists' colorectal cancer minimum dataset within a population. *Br J Cancer,* 97**,** 1393-8.
MAYER, E. K., BOTTLE, A., RAO, C., DARZI, A. W. & ATHANASIOU, T. 2009a. Funnel plots and their emerging application in surgery. *Ann Surg,* 249**,** 376-83.
MAYER, E. K., CHOW, A., VALE, J. A. & ATHANASIOU, T. 2009b. Appraising the quality of care in surgery. *World J Surg,* 33**,** 1584-93.
MCARDLE, C. S. & HOLE, D. 1991. Impact of variability among surgeons on postoperative morbidity and mortality and ultimate survival. *BMJ,* 302**,** 1501-5.

MCGORY, M. L., SHEKELLE, P. G. & KO, C. Y. 2006. Development of quality indicators for patients undergoing colorectal cancer surgery. *J Natl Cancer Inst,* 98**,** 1623-33.
MCPHERSON, K. 1984. Length of stay and health outcome. *Br Med J (Clin Res Ed),* 288**,** 1854-5.
MEDICAL OUTCOMES TRUST 2009. http://www.sf-36.org/tools/sf36.shtml.
MEKENKAMP, L. J., VAN KRIEKEN, J. H., MARIJNEN, C. A., VAN DE VELDE, C. J. & NAGTEGAAL, I. D. 2009. Lymph node retrieval in rectal cancer is dependent on many factors--the role of the tumor, the patient, the surgeon, the radiotherapist, and the pathologist. *Am J Surg Pathol,* 33**,** 1547-53.
MERKOW, R. P., BILIMORIA, K. Y., COHEN, M. E., RICHARDS, K., KO, C. Y. & HALL, B. L. 2009a. Variability in Reoperation Rates at 182 Hospitals: A Potential Target for Quality Improvement. *Journal of the American College of Surgeons,* 209**,** 557-564.
MERKOW, R. P., BILIMORIA, K. Y., COHEN, M. E., RICHARDS, K., KO, C. Y. & HALL, B. L. 2009b. Variability in reoperation rates at 182 hospitals: a potential target for quality improvement. *J Am Coll Surg,* 209**,** 557-64.
MIDWINTER, M. J., TYTHERLEIGH, M. & ASHLEY, S. 1999. Estimation of mortality and morbidity risk in vascular surgery using POSSUM and the Portsmouth predictor equation. *Br J Surg,* 86**,** 471-4.
MOHIL, R. S., SINGH, T., ARYA, S. & BHATNAGAR, D. 2008. Risk adjustment is crucial in comparing outcomes of various surgical modalities in patients with ileal perforation. *Patient Saf Surg,* 2**,** 31.
MORGAN, O. & BAKER, A. 2006. Measuring deprivation in England and Wales using 2001 Carstairs scores. *Health Stat Q***,** 28-33.
MORIARTY, J. P., FINNIE, D. M., JOHNSON, M. G., HUDDLESTON, J. M. & NAESSENS, J. M. 2010. Do pre-existing complications affect the failure to rescue quality measures? *Qual Saf Health Care,* 19**,** 65-8.
MORINO, M., PARINI, U., ALLAIX, M. E., MONASTEROLO, G., BRACHET CONTUL, R. & GARRONE, C. 2009. Male sexual and urinary function after laparoscopic total mesorectal excision. *Surg Endosc,* 23**,** 1233-40.
MORIYA, Y. 2006. Function preservation in rectal cancer surgery. *Int J Clin Oncol,* 11**,** 339-43.
MORRIS, A. M., BALDWIN, L. M., MATTHEWS, B., DOMINITZ, J. A., BARLOW, W. E., DOBIE, S. A. & BILLINGSLEY, K. G. 2007a. Reoperation as a quality indicator in colorectal surgery: a population-based analysis. *Ann Surg,* 245**,** 73-9.
MORRIS, A. M., BALDWIN, L. M., MATTHEWS, B., DOMINITZ, J. A., BARLOW, W. E., DOBIE, S. A. & BILLINGSLEY, K. G. 2007b. Reoperation as a quality indicator in colorectal surgery: A population-based analysis. *Annals of Surgery,* 245**,** 73-79.
MORRIS, E., QUIRKE, P., THOMAS, J. D., FAIRLEY, L., COTTIER, B. & FORMAN, D. 2008. Unacceptable variation in abdominoperineal excision rates for rectal cancer: time to intervene? *Gut,* 57**,** 1690-7.
MORRIS, E. J., MAUGHAN, N. J., FORMAN, D. & QUIRKE, P. 2007c. Identifying stage III colorectal cancer patients: the influence of the patient, surgeon, and pathologist. *J Clin Oncol,* 25**,** 2573-9.
MORRIS, E. J., TAYLOR, E. F., THOMAS, J. D., QUIRKE, P., FINAN, P. J., COLEMAN, M. P., RACHET, B. & FORMAN, D. 2011. Thirty-day postoperative mortality after colorectal cancer surgery in England. *Gut*.
NATIONAL SERVICES SCOTLAND. 2011. *About ISD* [Online]. Available: http://www.isdscotland.org/isd/846.html [Accessed 04/05/2011 2011].

NATIONAL STATISTICS PUBLICATION 2010. Delayed Discharges in NHSScotland-figures from October 2010 census.
NELSON, R. L., GLENNY, A. M. & SONG, F. 2009. Antimicrobial prophylaxis for colorectal surgery. *Cochrane Database Syst Rev*, CD001181.
NEWSCHAFFER, C. J., BUSH, T. L. & PENBERTHY, L. T. 1997. Comorbidity measurement in elderly female breast cancer patients with administrative and medical records data. *J Clin Epidemiol,* 50**,** 725-33.
NICE. 2004. *Improving outcomes in colorectal cancers: Manual update* [Online]. Available: http://www.nice.org.uk/nicemedia/pdf/CSGCCfullguidance.pdf [Accessed January 2010.
NICKELSEN, T. N., JORGENSEN, T. & KRONBORG, O. 2005. Thirty-day mortality after surgery for colorectal cancer in Denmark. *Colorectal Dis,* 7**,** 500-6.
NOBLETT, S. E., SNOWDEN, C. P., SHENTON, B. K. & HORGAN, A. F. 2006. Randomized clinical trial assessing the effect of Doppler-optimized fluid management on outcome after elective colorectal resection. *Br J Surg,* 93**,** 1069-76.
O'CONNOR, G. T., PLUME, S. K., OLMSTEAD, E. M., COFFIN, L. H., MORTON, J. R., MALONEY, C. T., NOWICKI, E. R., TRYZELAAR, J. F., HERNANDEZ, F., ADRIAN, L. & ET AL. 1991. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. The Northern New England Cardiovascular Disease Study Group. *JAMA,* 266**,** 803-9.
OMOIGUI, N. A., MILLER, D. P., BROWN, K. J., ANNAN, K., COSGROVE, D., 3RD, LYTLE, B., LOOP, F. & TOPOL, E. J. 1996. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation,* 93**,** 27-33.
PAL, N., AXISA, B., YUSOF, S., NEWCOMBE, R. G., WEMYSS-HOLDEN, S., RHODES, M. & LEWIS, M. P. 2008. Volume and outcome for major upper GI surgery in England. *J Gastrointest Surg,* 12**,** 353-7.
PATEL, V. M., WARREN, O., HUMPHRIS, P., AHMED, K., ASHRAFIAN, H., RAO, C., ATHANASIOU, T. & DARZI, A. 2010. What does leadership in surgery entail? *ANZ J Surg,* 80**,** 876-83.
PAULSON, E. C., MITRA, N., SONNAD, S., ARMSTRONG, K., WIRTALLA, C., KELZ, R. R. & MAHMOUD, N. N. 2008. National Cancer Institute designation predicts improved outcomes in colorectal cancer surgery. *Ann Surg,* 248**,** 675-86.
PEETERS, K. C., MARIJNEN, C. A., NAGTEGAAL, I. D., KRANENBARG, E. K., PUTTER, H., WIGGERS, T., RUTTEN, H., PAHLMAN, L., GLIMELIUS, B., LEER, J. W. & VAN DE VELDE, C. J. 2007. The TME trial after a median follow-up of 6 years: increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. *Ann Surg,* 246**,** 693-701.
PERLA, R. J. & PROVOST, L. P. 2012. Judgment sampling: a health care improvement perspective. *Qual Manag Health Care,* 21**,** 169-75.
PLOEG, A. J., LANGE, C. P., LARDENOYE, J. W. & BRESLAU, P. J. 2008. The incidence of unplanned returns to the operating room after peripheral arterial bypass surgery and its value as indicator of quality of care. *Vasc Endovascular Surg,* 42**,** 19-24.
PRYSTOWSKY, J. B., BORDAGE, G. & FEINGLASS, J. M. 2002. Patient outcomes for segmental colon resection according to surgeon's training, certification, and experience. *Surgery,* 132**,** 663-70; discussion 670-2.

PTOK, H., MARUSCH, F., KUHN, R., GASTINGER, I. & LIPPERT, H. 2007a. Influence of hospital volume on the frequency of abdominoperineal resection and long-term oncological outcomes in low rectal cancer. *European Journal of Surgical Oncology,* 33**,** 854-61.

PTOK, H., MARUSCH, F., MEYER, F., SCHUBERT, D., GASTINGER, I. & LIPPERT, H. 2007b. Impact of anastomotic leakage on oncological outcome after rectal cancer resection. *British Journal of Surgery,* 94**,** 1548-1554.

PUBLICATION, D. 1986. Medicare Hospital Mortality Information. *US Department of Health and Human Services.* Washington DC: 01-002.

QUIRKE, P. 2003. Training and quality assurance for rectal cancer: 20 years of data is enough. *Lancet Oncol,* 4**,** 695-702.

REASON, J. 2000. Human error: models and management. *BMJ,* 320**,** 768-770.

RICCIARDI, R., ROBERTS, P. L., READ, T. E., MARCELLO, P. W., HALL, J. F. & SCHOETZ, D. J. 2012. How often do patients return to the operating room after colorectal resections? *Colorectal Dis,* 14**,** 515-21.

RIEKER, R. J., HAMMER, E., EISELE, R., SCHMID, E. & HOGEL, J. 2002. The impact of comorbidity on the overall survival and the cause of death in patients after colorectal cancer resection. *Langenbecks Arch Surg,* 387**,** 72-6.

RIGBY, K. A., PALFREYMAN, S. & MICHAELS, J. A. 2001. Performance indicators from routine hospital data: death following aortic surgery as a potential measure of quality of care. *Br J Surg,* 88**,** 964-8.

RISTVEDT, S. L. & TRINKAUS, K. M. 2009. Trait anxiety as an independent predictor of poor health-related quality of life and post-traumatic stress symptoms in rectal cancer. *Br J Health Psychol,* 14**,** 701-15.

ROBERTS, K. H., MADSEN, P., DESAI, V. & VAN STRALEN, D. 2005. A case of the birth and death of a high reliability healthcare organisation. *Quality and Safety in Health Care,* 14**,** 216-220.

ROGERS, S. O., JR., WOLF, R. E., ZASLAVSKY, A. M., WRIGHT, W. E. & AYANIAN, J. Z. 2006. Relation of surgeon and hospital volume to processes and outcomes of colorectal cancer surgery. *Ann Surg,* 244**,** 1003-11.

SAGAR, P. M., HARTLEY, M. N., MACFIE, J., TAYLOR, B. A. & COPELAND, G. P. 1996. Comparison of individual surgeon's performance. Risk-adjusted analysis with POSSUM scoring system. *Dis Colon Rectum,* 39**,** 654-8.

SAGAR, P. M., HARTLEY, M. N., MANCEY-JONES, B., SEDMAN, P. C., MAY, J. & MACFIE, J. 1994. Comparative audit of colorectal resection with the POSSUM scoring system. *Br J Surg,* 81**,** 1492-4.

SALZ, T. & SANDLER, R. S. 2008. The effect of hospital and surgeon volume on outcomes for rectal cancer surgery. *Clin Gastroenterol Hepatol,* 6**,** 1185-93.

SANABRIA, A., DOMINGUEZ, L. C., VALDIVIESO, E. & GOMEZ, G. 2007. Prophylactic antibiotics for mesh inguinal hernioplasty: a meta-analysis. *Ann Surg,* 245**,** 392-6.

SANFEY, H., MCDOWELL, C., MEIER, A. H. & DUNNINGTON, G. L. 2011. Team training for surgical trainees. *Surgeon,* 9 Suppl 1**,** S32-4.

SANO, T., SASAKO, M., YAMAMOTO, S., NASHIMOTO, A., KURITA, A., HIRATSUKA, M., TSUJINAKA, T., KINOSHITA, T., ARAI, K., YAMAMURA, Y. & OKAJIMA, K. 2004. Gastric cancer surgery: morbidity and mortality results from a prospective randomized controlled trial comparing D2 and extended para-aortic lymphadenectomy--Japan Clinical Oncology Group study 9501. *J Clin Oncol,* 22**,** 2767-73.

SCHRAG, D., CRAMER, L. D., BACH, P. B., COHEN, A. M., WARREN, J. L. & BEGG, C. B. 2000. Influence of hospital procedure volume on outcomes following surgery for colon cancer. *JAMA,* 284**,** 3028-35.

SCHRAG, D., PANAGEAS, K. S., RIEDEL, E., CRAMER, L. D., GUILLEM, J. G., BACH, P. B. & BEGG, C. B. 2002. Hospital and surgeon procedure volume as predictors of outcome following rectal cancer resection. *Ann Surg,* 236**,** 583-92.

SCHRAG, D., PANAGEAS, K. S., RIEDEL, E., HSIEH, L., BACH, P. B., GUILLEM, J. G. & BEGG, C. B. 2003. Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection. *Journal of Surgical Oncology,* 83**,** 68-78.

SEAGROATT, V. & GOLDACRE, M. 1994. Measures of early postoperative mortality: beyond hospital fatality rates. *BMJ,* 309**,** 361-5; discussion 365-6.

SENAGORE, A. J., WARMUTH, A. J., DELANEY, C. P., TEKKIS, P. P. & FAZIO, V. W. 2004. POSSUM, p-POSSUM, and Cr-POSSUM: implementation issues in a United States health care system for prediction of outcome for colon cancer resection. *Dis Colon Rectum,* 47**,** 1435-41.

SERVICES, D. O. H. A. H. 2008. *Medicare Programe; Changes to the Hospital Prospective Payment Systems and Fiscal Year 2008 rates* [Online]. Available: http://www.rwjf.org/files/research/1533cms2007.pdf [Accessed 18/10/2010 2010].

SERVICES, U. S. D. O. H. H. 2010. *Agency for Healthcare Research and Quality* [Online]. Available: http://www.ahrq.gov/ [Accessed 18/10/2010.

SHAHIAN, D. M., WOLF, R. E., IEZZONI, L. I., KIRLE, L. & NORMAND, S. L. 2010. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med,* 363**,** 2530-9.

SHAIKH, A. R., SANGRASI, A. K. & SHAIKH, G. A. 2009. Clinical outcomes of laparoscopic versus open appendectomy. *JSLS,* 13**,** 574-80.

SHEPPERD, S., MCCLARAN, J., PHILLIPS, C. O., LANNIN, N. A., CLEMSON, L. M., MCCLUSKEY, A., CAMERON, I. D. & BARRAS, S. L. 2010. Discharge planning from hospital to home. *Cochrane Database Syst Rev***,** CD000313.

SIBBRITT, D. W. 1995. Validation of a 28 day interval between discharge and readmission for emergency readmission rates. *J Qual Clin Pract,* 15**,** 211-20.

SIEWERT, J. R., BOTTCHER, K., STEIN, H. J. & RODER, J. D. 1998. Relevant prognostic factors in gastric cancer: ten-year results of the German Gastric Cancer Study. *Ann Surg,* 228**,** 449-61.

SILBER, J. H., ROSENBAUM, P. R., SCHWARTZ, J. S., ROSS, R. N. & WILLIAMS, S. V. 1995. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA,* 274**,** 317-23.

SILBER, J. H., WILLIAMS, S. V., KRAKAUER, H. & SCHWARTZ, J. S. 1992. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. *Med Care,* 30**,** 615-29.

SKIPWORTH, R. J., PARKS, R. W., STEPHENS, N. A., GRAHAM, C., BREWSTER, D. H., GARDEN, O. J. & PATERSON-BROWN, S. 2010. The relationship between hospital volume and post-operative mortality rates for upper gastrointestinal cancer resections: Scotland 1982-2003. *Eur J Surg Oncol,* 36**,** 141-7.

SMITH, P. 1995. On the unintended consequences of publishing performance data in the public sector. . *International Journal of Public Administration,* 18**,** 277 - 310.

SOLOMON, M. J., PAGER, C. K., KESHAVA, A., FINDLAY, M., BUTOW, P., SALKELD, G. P. & ROBERTS, R. 2003. What do patients want? Patient preferences and surrogate decision making in the treatment of colorectal cancer. *Dis Colon Rectum,* 46**,** 1351-7.

SOSA, J. A., BOWMAN, H. M., GORDON, T. A., BASS, E. B., YEO, C. J., LILLEMOE, K. D., PITT, H. A., TIELSCH, J. M. & CAMERON, J. L. 1998. Importance of hospital volume in the overall management of pancreatic cancer. *Ann Surg,* 228**,** 429-38.

SPIEGELHALTER, D. J. 2005. Funnel plots for comparing institutional performance. *Stat Med,* 24**,** 1185-202.

STEINBERG, S. M., POPA, M. R., MICHALEK, J. A., BETHEL, M. J. & ELLISON, E. C. 2008. Comparison of risk adjustment methodologies in surgical quality improvement. *Surgery,* 144**,** 662-7; discussion 662-7.

STOCCHI, L., NELSON, H., SARGENT, D. J., O'CONNELL, M. J., TEPPER, J. E., KROOK, J. E. & BEART, R., JR. 2001. Impact of surgical and pathologic variables in rectal cancer: a United States community and cooperative group report. *J Clin Oncol,* 19**,** 3895-902.

SUTCLIFFE, K. M., LEWTON, E. & ROSENTHAL, M. M. 2004. Communication failures: an insidious contributor to medical mishaps. *Acad Med,* 79**,** 186-94.

TALATI, R., REINHART, K. M., WHITE, C. M., PHUNG, O. J., SEDRAKYAN, A., KLUGER, J. & COLEMAN, C. I. 2009. Outcomes of perioperative beta-blockade in patients undergoing noncardiac surgery: a meta-analysis. *Ann Pharmacother,* 43**,** 1181-8.

TALSMA, A., JONES, K., LIU, G. & CAMPBELL, D. A. 2010. Failure to rescue measure: validation of community- and hospital-acquired complications. *J Nurs Adm,* 40**,** 417-23.

TAN, W. S., TANG, C. L., SHI, L. & EU, K. W. 2009. Meta-analysis of defunctioning stomas in low anterior resection for rectal cancer. *British Journal of Surgery,* 96**,** 462-472.

TAYLOR, F. C., ASCIONE, R., REES, K., NARAYAN, P. & ANGELINI, G. D. 2003. Socioeconomic deprivation is a predictor of poor postoperative cardiovascular outcomes in patients undergoing coronary artery bypass grafting. *Heart,* 89**,** 1062-6.

TEKKIS, P. P., POLONIECKI, J. D., THOMPSON, M. R. & STAMATAKIS, J. D. 2003. Operative mortality in colorectal cancer: prospective national study. *BMJ,* 327**,** 1196-201.

THE INFORMATION CENTRE (ENGLAND). 2011. *What is HES* [Online]. Available: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=456 [Accessed 04/05/2011 2011].

THOMPSON, J. W., PINIDIYA, S. D., RYAN, K. W., MCKINLEY, E. D., ALSTON, S., BOST, J. E., BRIEFER FRENCH, J. & SIMPSON, P. 2003. Health plan quality-of-care information is undermined by voluntary reporting. *Am J Prev Med,* 24**,** 62-70.

THOMPSON, M. R., TEKKIS, P. P., STAMATAKIS, J., SMITH, J. J., WOOD, L. F., VON HILDEBRAND, M. & POLONIECKI, J. D. 2010. The National Bowel Cancer Audit: the risks and benefits of moving to open reporting of clinical outcomes. *Colorectal Dis,* 12**,** 783-91.

THORNE, K., HUTCHINGS, H. A. & ELWYN, G. 2006. The effects of the Two-Week Rule on NHS colorectal cancer diagnostic services: a systematic literature review. *BMC Health Serv Res,* 6**,** 43.

TORNROOS, A., GARVIN, S. & OLSSON, H. 2009. The number of identified lymph node metastases increases continuously with increased total lymph node recovery in pT3 colon cancer. *Acta Oncol,* 48**,** 1152-6.

UK, C. R. 2006. *UK cancer incidence statistics for common cancers*.

UK, C. R. 2011a. *Oesophageal cancer - UK incidence statistics* [Online]. London. Available: http://info.cancerresearchuk.org/cancerstats/types/oesophagus/incidence/ [Accessed 29/11/2011 2011].

UK, C. R. 2011b. *Stomach cancer - UK incidence statistics* [Online]. London. Available: http://info.cancerresearchuk.org/cancerstats/types/stomach/incidence/ [Accessed 29/11/2011 2011].

VAN DER SCHAAF, M., LAGERGREN, J. & LAGERGREN, P. 2012. Persisting symptoms after intrathoracic anastomotic leak following oesophagectomy for cancer. *Br J Surg,* 99**,** 95-9.

VARADHAN, K. K., NEAL, K. R., DEJONG, C. H. C., FEARON, K. C. H., LJUNGQVIST, O. & LOBO, D. N. 2010. The enhanced recovery after surgery (ERAS) pathway for patients undergoing major elective open colorectal surgery: A meta-analysis of randomized controlled trials. *Clinical nutrition (Edinburgh, Scotland),* 29**,** 434-440.

VARNAVA, A. M., SEDGWICK, J. E., DEANER, A., RANJADAYALAN, K. & TIMMIS, A. D. 2002. Restricted weekend service inappropriately delays discharge after acute myocardial infarction. *Heart,* 87**,** 216-9.

VATHER, R., ZARGAR-SHOSHTARI, K., ADEGBOLA, S. & HILL, A. G. 2006. Comparison of the possum, P-POSSUM and Cr-POSSUM scoring systems as predictors of postoperative mortality in patients undergoing major colorectal surgery. *ANZ J Surg,* 76**,** 812-6.

VINCENT, C., BENN, J. & HANNA, G. B. 2010. High reliability in health care. *BMJ,* 340.

VINCENT, C. B., SUSAN; CARTHEY, J. 2013. The measurement and monitoring of safety. *In:* FOUNDATION, T. H. (ed.).

VISSER, B. C., KEEGAN, H., MARTIN, M. & WREN, S. M. 2009. Death after colectomy: it's later than we think. *Arch Surg,* 144**,** 1021-7.

VLUG, M. S., WIND, J., VAN DER ZAAG, E., UBBINK, D. T., CENSE, H. A. & BEMELMAN, W. A. 2009. Systematic review of laparoscopic vs open colonic surgery within an enhanced recovery programme. *Colorectal Dis,* 11**,** 335-43.

VOGEL, C., KIRTIL, T., OELLIG, F. & STOLTE, M. 2008. Lymph node preparation in resected colorectal carcinoma specimens employing the acetone clearing method. *Pathol Res Pract,* 204**,** 11-5.

WEISER, T. G., REGENBOGEN, S. E., THOMPSON, K. D., HAYNES, A. B., LIPSITZ, S. R., BERRY, W. R. & GAWANDE, A. A. 2008. An estimation of the global volume of surgery: a modelling strategy based on available data. *The Lancet,* 372**,** 139-144.

WEST, N. P., FINAN, P. J., ANDERIN, C., LINDHOLM, J., HOLM, T. & QUIRKE, P. 2008a. Evidence of the oncologic superiority of cylindrical abdominoperineal excision for low rectal cancer. *J Clin Oncol,* 26**,** 3517-22.

WEST, N. P., MORRIS, E. J., ROTIMI, O., CAIRNS, A., FINAN, P. J. & QUIRKE, P. 2008b. Pathology grading of colon cancer surgical resection and its association with survival: a retrospective observational study. *Lancet Oncol,* 9**,** 857-65.

WIBE, A., SYSE, A., ANDERSEN, E., TRETLI, S., MYRVOLD, H. E. & SOREIDE, O. 2004. Oncological outcomes after total mesorectal excision for cure for cancer of the lower rectum: anterior vs. abdominoperineal resection. *Dis Colon Rectum,* 47**,** 48-58.

WILLIAMS, J. G. & MANN, R. Y. 2002. Hospital episode statistics: time for clinicians to get involved? *Clin Med,* 2**,** 34-7.

WIND, J., POLLE, S. W., FUNG KON JIN, P. H., DEJONG, C. H., VON MEYENFELDT, M. F., UBBINK, D. T., GOUMA, D. J. & BEMELMAN, W. A. 2006. Systematic review of enhanced recovery programmes in colonic surgery. *Br J Surg,* 93**,** 800-9.

WONG, J. H., BOWLES, B. J., BUENO, R. & SHIMIZU, D. 2002. Impact of the number of negative nodes on disease-free survival in colorectal cancer patients. *Dis Colon Rectum,* 45**,** 1341-8.

WU, C., HANNAN, E. L., RYAN, T. J., BENNETT, E., CULLIFORD, A. T., GOLD, J. P., ISOM, O. W., JONES, R. H., MCNEIL, B., ROSE, E. A. & SUBRAMANIAN, V. A. 2004. Is the impact of hospital and surgeon volumes on the in-hospital mortality rate for coronary artery bypass graft surgery limited to patients at high risk? *Circulation,* 110**,** 784-9.

YEH, C. Y., CHANGCHIEN, C. R., WANG, J. Y., CHEN, J. S., CHEN, H. H., CHIANG, J. M. & TANG, R. 2005. Pelvic drainage and other risk factors for leakage after elective anterior resection in rectal cancer patients: A prospective study of 978 patients. *Ann Surg,* 241**,** 9-13.

YOO, H. M., LEE, H. H., SHIM, J. H., JEON, H. M., PARK, C. H. & SONG, K. Y. 2011. Negative impact of leakage on survival of patients undergoing curative resection for advanced gastric cancer. *J Surg Oncol,* 104**,** 734-40.

YOUNG, G. J., CHARNS, M. P., DESAI, K., KHURI, S. F., FORBES, M. G., HENDERSON, W. & DALEY, J. 1998. Patterns of coordination and clinical outcomes: a study of surgical services. *Health Serv Res,* 33**,** 1211-36.

ZOOK, C. J. & MOORE, F. D. 1980. High-cost users of medical care. *N Engl J Med,* 302**,** 996-1002.

APPENDIX 1

Procedures	OPCS-4 codes	ICD-10 codes
right and extended right hemicolectomies,	H06.1-H07.9	
transverse colectomy,	H08.1-H08.9	C18-C21, C26
left hemicolectomy, sigmoid colectomy, Hartmann's procedure,	H09.1-H09.9 H10.1-H10.9 H33.5	(malignant colorectal codes)
subtotal colectomy, panproctocolectomy, total colectomy,	H29 H04.1, H04.2, H04.3, H04.8, H04.9, H05.1, H05.2, H05.3, H05.8, H05.9	
anterior resection (AR),	H33.2, H33.3, H33.4, H33.6,	
excision of rectum unspecified/other	H33.7, H33.8, H33.9,	
abdominoperineal resection (APER)	H33.1	
Laparoscopic procedures	Y50.8,Y75, Y71.4	

Appendix 1- OPCS and ICD-10 codes used in the analyses

APPENDIX 2

High Performance in Surgery (HiPer)

Interviewee Information

Alex Almoudaris, Omar Faiz, Charles Vincent

Dear Colleague,

We are contacting you as **[name of GC contact]** has informed us that you have kindly agreed to consider taking part in a telephone interview being undertaken as part of a quality improvement process that your hospital has signed up to. The overall project is called the High Performance in Surgery project (HiPer).

The enclosed booklet explains the purpose of the study and what sort of questions will be asked. Participation is entirely voluntary and we greatly appreciate your time and contributions in advance.

1. About HiPer

The aim of the programme is to gain a better understanding of what defines a High Performing (HiPer) surgical unit, and furthermore how this is achieved. The essence is 'looking behind the numbers' that are so often used for benchmarking or available to the public. The objective is to use colorectal cancer surgery for this work with the overall aim of developing a framework for quality improvement that might be applied across other specialities.

This specific piece of work has been established under the direction of Professor Charles Vincent and Omar Faiz using participant institutions from the Global Comparators project of Dr. Foster. This work is in collaboration with and co-funded by the National Institute of Health Research (UK). The programme of work was launched in October 2009.

2. About Your Contribution

You have been asked to participate as an interviewee because of your role in colorectal surgery and that you care for such patients. Furthermore you have been specifically chosen as it was felt that you would be the most informative person to discuss how colorectal patients are cared for in your unit.

3. What is expected of me?

We ask for half an hour (30minutes) of your time to speak with one of our trained interviewers. The interviewers have a framework of questions they can ask but it is not a rigid interview by any means. We want to hear what you think is important about how your unit achieves its results. We would like to record the interviews so that we can re-visit them at a later time rather than need to contact you again. The interviews will take place from November 2011 and run into 2012.

4. Who is involved and who are we interviewing?

Through collaboration with Dr Foster, 10-12 Academic Health Science equivalent colorectal units from across the world have agreed to take part. We aim to interview 4 people (a surgeon, nurse, intensivist and surgical manager) from each hospital.

5. So what is the MAIN AIM?

The main aim is to understand from you, how your unit looks after colorectal patients undergoing surgical resections. There are no right or wrong answers. We simply want to know how things are done in your unit.

6. What sort of questions am I likely to be asked?

The types of questions will vary from general questions like 'how is the decision made that a patient needs an operation?' to more specific questions like 'what is the process of asking opinions from different teams/specialities both in and out of hours? What we really would like to hear is what actually happens in your unit. There will be specific questions depending upon whether you are a manager, nurse or doctor as well.

7. Is the information confidential

Everything you say is 100% confidential. No one outside the research team will ever have access to transcripts of the recordings or the interviews. Furthermore the recordings will be coded so neither your institution nor your name will be relatable to the recording. Only the research organiser will have this information stored in an encrypted fashion.

8. So what happens next?

If you agree to take part (and we hope you do) one of our panel of interviewers will contact you shortly. Please liaise with them via email to arrange a mutually beneficial time for you both to undertake the interview. They will know your name, hospital and job role. They will clarify this information with you at the beginning of the interview.

9. How is the project co-ordinated?

The project is co-ordinated by Alex Almoudaris (PhD Surgical Research fellow at the Centre for Patient Safety and Service quality) at Imperial College to whom all queries should be addressed- contact details below.

We thank you for your commitment and contribution.

Further Information

If you have any questions about the process or need further information on the purpose of the research please contact the research team by telephoning or emailing:

Alex Almoudaris, Lead Researcher,

+447970699242: email alex.almoudaris@imperial.ac.uk

HiPer SURGERY