Division of Surgery & Interventional Science

Development, implementation, and evaluation of postoperative morbidity monitoring in a multicentre UK cohort undergoing major surgery

Dr James R. Bedford

A thesis submitted for the degree of Doctor of Philosophy

University College London

2025

Declaration

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

Chapter 1,6 — Nil

Chapter 2 – Dr Aleksandra Ignaka (AI) assisted with screening of abstracts alongside Dr James Bedford (JB). Dr Ignaka also served as the second reviewer for the full text reviews.

Chapter 3–5 — Study conception of the Perioperative Quality Improvement Programme and development of a morbidity monitoring tool pilot within the wider study: S. Ramani Moonesinghe; Feedback and oversight of the development of study protocols and documents: Sonya Crowe, PQIP project team at the Royal College of Anaesthetists, including Duncan Wagstaff, Arun Sahni, Cecilia Vindrola-Padros, Dermot McGuckin, Jonathan Wilson, Mike Swart, Jenny Dorey, Jose Lourtie and James Goodwin. Electronic database and web development for the patient cohort study and pomVLAD dashboard: Net Solving Limited. Collection of source data and data entry for the Perioperative Quality Improvement Programme patient cohort study: site investigators and collaborators at all participating hospitals.

Acknowledgments

A lot has happened in the past seven years: five Prime Ministers, Britain leaving the European Union, a global pandemic, and a new monarch. On a personal note, these years have also brought a marriage, the arrival of two children, completion of training, and my appointment as a consultant anaesthetist. This thesis has been a constant companion throughout that period, and I owe a huge debt of gratitude to all those who have given their time to support me.

I am enormously grateful to Professor Ramani Moonesinghe, whose kindness, patience, insight, and encouragement have been unfailing, and to Professor Sonya Crowe, whose knowledge and gentle guidance have always nudged me towards better and clearer ideas.

This research would not have been possible without the generous contributions of the collaborators and patients who supported the PQIP study. The Princess Grace Hospital provided flexible employment during my time out of training, and I am also grateful to my educational supervisors and training programme directors who supported me to continue my research after returning to training. The Royal College of Anaesthetists - Centre for Research and Improvement provided administrative support and infrastructure to run both the PQIP and pomVLAD studies. I also wish to thank the Health Foundation who enabled the pomVLAD pilot study through grant funding.

Thank you to all my colleagues and friends who have offered support, advice, listened to my challenges, and helped fix my broken code.

To my wife and family, I owe more than I can put into words—for your patience, your love, and the sacrifices you made so I could spend time in libraries across South London completing this thesis.

Abstract

This thesis investigates the implementation and use of the pomVLAD dashboard within the Perioperative Quality Improvement Programme (PQIP) to report near real-time, risk-adjusted postoperative morbidity outcomes and perioperative care recommendations in NHS hospitals. The aim was to determine whether this intervention improved clinical outcomes, enhanced adherence to care processes, and engaged healthcare professionals in quality improvement (QI) initiatives.

Quality in healthcare is a multifaceted concept, encompassing safety, effectiveness, and patient-centred care. Large-scale QI initiatives, such as PQIP, aim to harness data-driven strategies to monitor and enhance perioperative care. However, the complexity of healthcare delivery and the variability of local contexts present challenges to the effective use of such interventions.

A bespoke risk-adjustment model for morbidity at postoperative day-7 was developed and validated, demonstrating superior performance compared to existing models. This model was integrated into the pomVLAD dashboard to provide timely, visual feedback on clinical performance. Despite its potential, the dashboard did not lead to significant reductions in postoperative morbidity or improved compliance with care processes during a 12-month follow-up.

Engagement with the pomVLAD dashboard varied across sites. While staff valued the dashboard's visualisation of performance metrics and real-time data, its use was often sporadic and secondary to other static quarterly reports. Facilitators of engagement included local champions, structured QI processes, senior management support, and alignment with national improvement priorities. However, barriers such as limited capacity of clinical teams to respond to data, competing clinical demands, and limited infrastructure for QI activities, and organisational inertia hindered its widespread adoption.

This research highlights the importance of tailoring national QI interventions to local contexts, fostering multidisciplinary collaboration, and providing adequate resources for sustained improvement. Future work should explore strategies to integrate data collection into routine care, develop customisable tools for local use, and address systemic challenges to ensure meaningful and sustainable improvements in patient outcomes.

Impact statement

The findings from my thesis, which explored the implementation and effectiveness of Variable Life Adjusted Displays (VLADs) in monitoring postoperative outcomes, offer potential benefits both within and beyond academia.

This work provides a detailed mixed methods evaluation of the role VLADs and other continuous monitoring tools can play in healthcare, particularly highlighting the challenges and limitations that exist when implemented into clinical practice. The findings will help refine methodologies and our understanding how contextual factors influence the successful implementation of such tools.

Beyond the academic environment, the knowledge gained has practical applications both at the `macro` system level and `meso` hospital level. The qualitative workstream identified several factors acting as facilitators and barriers to quality improvement activity based on data presented in continuous monitoring tools. By addressing these, future implementations of continuous tools may obtain more clinician engagement and lead to a greater improvement in patient care and safety.

For policymakers and those involved in health systems design, the thesis offers observations to guide the development of policies promoting effective monitoring tools. In an increasingly stretched healthcare system, the opportunity to implement monitoring tools capable of detecting potential harm early, allowing rapid intervention to improve outcomes, offers significant potential from a systemwide quality and safety perspective.

Following the pilot of the pomVLAD study detailed in this thesis, a wider roll out of a VLAD monitoring tool has been delivered, monitoring postoperative morbidity outcomes for patients undergoing major colorectal surgery. This tool is now available to over 150 NHS hospitals participating in the Perioperative Quality

Improvement Programme and will support clinicians to understand and improve the quality of care their institutions deliver. Around 19,000 patients undergo major colorectal surgery for cancer each year in the NHS. Reducing the incidence of postoperative complications, even by a small proportion will have a significant impact on healthcare costs and resource utilisation.

Although not the topic of the thesis, I worked on the wider PQIP programme which is active in 173 NHS hospitals, having recruited over 53,000 patients. I developed the code and reporting structures that now provide around 400 specialty level reports to NHS hospitals every 3 months, supporting quality improvement. I also helped deliver several national PQIP reports.

Work from this thesis has also been built into a successful NIHR Patient Safety Research Collaborative application with plans to further develop the use of pomVLAD to evaluate implementation of other innovations.

Research paper declaration forms: referencing my own published works

1. Wagstaff DT, Bedford J, Moonesinghe SR. Improvement Science in Anaesthesia. *Current Anesthesiology Reports* 2017; 7: 432–9.

1. For a research manuscript that has already been published
a) **What is the title of the manuscript?**
Improvement Science in Anaesthesia
b) **Please include a link to or doi for the work**
10.1007/s40140-017-0234-5
c) **Where was the work published?**
Current Anaesthesiology Reports
d) **Who published the work?** (e.g. OUP)
Springer
e) **When was the work published?**
September 2017
f) **List the manuscript's authors in the order they appear on the publication**
D Wagstaff, **J Bedford**, S.R. Moonesinghe
g) **Was the work peer reviewed?** Yes
2. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)
SRM was invited to provide the review. DW and JB co-wrote the article with senior oversight from SRM. JB led 'Measurement for Improvement' section. DW led 'Features of Improvement Science' section. All authors took part in revisions and were responsible for final submission.
3. In which chapter(s) of your thesis can this material be found?

Elements of above work can be found in Chapter 1

2. Bedford J, Martin P, Crowe S et al. Development and internal validation of a model for postoperative morbidity in adults undergoing major elective colorectal surgery: the peri-operative quality improvement programme (PQIP) colorectal risk model. *Anaesthesia* 2022; **672**: 1–12.

1. For a research manuscript that has already been published
h) **What is the title of the manuscript?**
Development and internal validation of a model for postoperative morbidity in adults undergoing major elective colorectal surgery: the peri-operative quality improvement programme (PQIP) colorectal risk model
i) **Please include a link to or doi for the work**
10.1111/anae.15858
j) **Where was the work published?**
Anaesthesia
k) **Who published the work?** (e.g. OUP)
John Wiley & Sons Ltd
l) **When was the work published?**
2022
m) **List the manuscript's authors in the order they appear on the publication**
J Bedford, P Martin, S Crowe, D Wagstaff, C Santos, G Singleton, R Baumber, C Vindrola-Padros, R Vohra, M Swart, CM Oliver, J Dorey, I Leeman, and SR Moonesinghe, for the peri-operative quality improvement programme (PQIP) collaborative
n) **Was the work peer reviewed?** Yes
2. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

JB, SC & SRM conceptualised and designed the study. PM provided statistical advice and support. JB drafted the manuscript which was reviewed, revised and approved by all the authors.
3. In which chapter(s) of your thesis can this material be found?
This work leans on and borrows significantly from Chapter 3, and should be considered a direct extension of the work presented in Chapter 3

Table of tables

Table of figures

Abbreviations

ACS	American College of Surgeons
AIC	Akaike information criterion
ASA	American Society of Anaesthesiologists
ASA-PS	American College of Anaesthesiologists - Physical status
AUROC	Area under receiver operating characteristic curve
BMI	Body mass index
BPT	Best practice tariff
CABG	Coronary artery bypass graft
CI	Confidence interval
CITS	Controlled interrupted time series
COPD	Chronic obstructive pulmonary disease
CQC	Care Quality Commission
CQUIN	Commissioning for Quality and Innovation
CRAM	Cumulative risk adjusted mortality
CR&I	Centre for Research and Improvement
CRF	Case record form
CRN	Clinical research network
CUSUM	Cumulative sum chart
DGH	District general hospital
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
ENT	Ear, Nose & Throat
ERAS	Enhanced Recovery After Surgery
EWMA	Exponentially weighted moving average
GI	Gastrointestinal
GIRFT	Getting it right first time
HES	Hospital Episode Statistics
HQIP	Healthcare Quality Improvement Partnership
ICNARC	Intensive Care National Audit and Research Centre
INT	Interviewee
IQR	Interquartile range
ITS	Interrupted time series
JVP	Jugular venous pressure
LASSO	Least absolute shrinkage and selection operator
LOS	Length of stay
MDT	Multidisciplinary team
MLE	Maximum likelihood estimation
MMAT	Mixed methods appraisal tool
NCEPOD	National Confidential Enquiry into Patient Outcome and Death
NELA	National Emergency Laparotomy Audit
NGT	Nasogastric tube

NHS	National Health Service
NIAA-HSRC	National Institute of Academic Anaesthesia - Health Services Research Centre
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
NSQIP	National Surgical Quality Improvement Programme
NYHA	New York Health Association
PDF	Portable document format
PDSA	Plan-do-study-act
PMLE	Penalised maximum likelihood estimation
PMR	Postoperative Morbidity Ratio
POMS	Postoperative Morbidity Survey
POSSUM	Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity
PQIP	Perioperative Quality Improvement Programme
PROM	Patient reported outcome measure
QA	Quality assurance
QI	Quality improvement
RA	Risk-adjusted
REC	Research and ethics committee
RFE	Rapid feedback evaluation
ROC	Receiver Operating Characteristic
SNAP2-EPICCS	Sprint National Audit Project 2 - Epidemiology of Critical Care after Surgery
SOP	Standardised operating procedure(s)
SORT	Surgical Outcome Risk Tool
TPN	Total parenteral nutrition
UCL	University College London
UK	United Kingdom
USA	United States of America
VATS	Video-assisted thoracoscopic surgery
VLAD	Variable life-adjusted display
WHO	World Health Organisation

Chapter 1 Introduction

1.1 Quality in healthcare

1.1.1 Defining quality

Quality in healthcare is a multidimensional concept incorporating the expectations and needs of patients, providers, and the wider society. The Institute of Medicine in the United States of America proposed six dimensions of quality in healthcare: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity.[1] The World Health Organisation (WHO) include these domains in their definition of care quality with one additional domain; providing integrated care that makes available the full range of health services throughout the life course.[2,3]

In the UK, universal healthcare is provided through the National Health Service (NHS), with independent bodies responsible for regulating health and social care. In England, this is the Care Quality Commission (CQC), with the devolved nations having their own care inspectorate services. The CQC consider five key questions, which reflect dimensions of quality, when assessing care delivery: Are they safe? Are they effective? Are the caring? Are they responsive to people's needs? Are they well-led?[4]

Safe healthcare services provide care whilst avoiding harm, minimising medical errors, preventing hospital-acquired infections, and promoting a culture of safety within institutions.[2] Effective care provides evidence-based care, aligned with current scientific knowledge and clinical guidelines.[5] This leads to better health outcomes by using treatments and interventions proven to work, while avoiding unnecessary or potentially harmful treatments. People-centred care is respectful, responsive, and individualised to patient needs, preferences, and values.[5] Examples of patient-centred care include involving patients in decision making,

providing clear communication, and ensuring patients feel supported and listened to throughout their care journey.

Timely care reduces waiting times and delays to provide prompt access to healthcare services. It can reduce complications, improve outcomes, and enhance patient satisfaction. Efficient care maximises the use of resources to deliver high-quality care at the lowest possible cost. This includes reducing waste, optimising processes and the use of healthcare staff, facilities, and technology. Equitable care is fair and does not discriminate based on characteristics such as age, disability, race, gender or sexual orientation, socioeconomic status, or geographical location. Equitable healthcare ensures all individuals have equal access to high-quality services and experience similar outcomes, regardless of their background or circumstances.[5]

1.1.2 Evaluating quality – The Donabedian framework

The Donabedian framework is a conceptual model for evaluating the quality of healthcare delivery.[6,7] It incorporates three related components: structure, process, and outcome. These components provide an approach for assessing and understanding factors that influence healthcare quality.

Structure refers to the organisational and environmental aspects of healthcare delivery, such as the hospital size, availability of resources, staffing levels and training, and organisational policies that support healthcare delivery. Structural factors influence quality by affecting the capacity of healthcare providers to deliver effective services. For example, a well-equipped institution with highly trained staff is more likely to provide high-quality care than an under-resourced one with less qualified personnel.

Process refers to activities and procedures involved in delivering healthcare services. This component includes clinical processes, such as investigations, treatments, the use of protocolised care pathways, and preventive care, as well as

non-clinical processes, like communication, coordination, and patient education. The process aspect of the Donabedian framework focuses on how care is provided and whether it adheres to established standards, guidelines, and best practices. High-quality processes are associated with better patient outcomes and are an important part of effective healthcare delivery.

Outcome refers to the end results or consequences of healthcare services, including changes in patients' health status, quality of life, and satisfaction with care. Common outcome measures in healthcare include mortality rates, morbidity or complication rates, readmission rates, and increasingly patient-reported outcomes. The outcome component of the Donabedian framework helps evaluate the effectiveness of healthcare delivery by assessing its impact on patients. High-quality healthcare delivery should lead to better patient outcomes, such as improved health, reduced complications, and higher levels of patient satisfaction.

1.1.3 Quality assurance

Quality assurance in healthcare refers to the systematic monitoring and evaluation of healthcare services to ensure they meet established standards of quality and safety. Quality assurance involves a range of activities, including the development of standards and guidelines, such as those produced by the National Institute for Health and Care Excellence (NICE).[8] Where national or local guidelines are in place for the investigation and management of certain conditions, performance of hospital trusts can be measured against the guidelines.

National clinical audit is designed to engage healthcare professionals in systematic evaluation of their clinical practice against standards and to encourage improvement in the quality of care.[9] Over 20 national audit projects have been established by the Healthcare Quality Improvement Partnership across a range of specialties in the United Kingdom.[10] Many of these projects are supported by the Royal Colleges, such as the National Emergency Laparotomy Audit (NELA), which

is delivered by the Health Services Research Centre, National Institute for Academic Anaesthesia in association with the Royal College of Anaesthetists. National audit projects use a range of data sources including administrative datasets such as the NHS Digital Hospital Episode Statistics (HES) dataset and prospectively collected clinical registries.[11–13] National audits do not necessarily lead to continuous quality improvement however. Concerns about the quality and robustness of administrative data can reduce clinician engagement compared to where clinical registries are used to support quality improvement.[9,14]

Publication of national audit data allows healthcare providers to compare their performance against that of other institutions and may promote accountability and transparency within the healthcare system. This 'benchmarking' may help identify variation in outcomes, facilitate knowledge sharing, and encourage improvement in care delivery. However, where accepted standards are being met there may be less incentive for clinicians, managers, and institutions to try and improve further. Public reporting has also been cited as a barrier that may reduce clinician engagement.[9] Involvement of clinicians relies on them having confidence in the results of national audits and actively wanting to compare their results.[15] Quality assurance may be considered a reactive approach that identifies deviation from the expected standards and takes actions to address these deviations.

1.1.4 Quality improvement

Quality improvement in healthcare refers to the systematic and continuous efforts aimed at enhancing the quality and safety of healthcare services, patient outcomes, and overall patient experience.[16] Improvement science, a multidisciplinary field that emerged in the late 20th century, seeks to generate evidence-based knowledge to drive effective quality improvement interventions and has since been adopted in various healthcare settings, including perioperative medicine.[17,18]

The field of improvement science has its roots in the application of industrial quality management principles, such as those by Edward Deming and Walter Shewhart, to healthcare settings.[19] The Institute of Medicine's report "To Err is Human" in 1999 highlighted the prevalence of medical errors and patient harm, emphasising the urgent need for quality improvement in healthcare.[20] This report served as a catalyst for the growth of improvement science and its application to various healthcare domains.

Improvement science seeks to understand how to promote the adoption of effective practices, policies, and interventions in complex, real-world healthcare settings.[21] It employs a range of methodologies, including the Plan-Do-Study-Act (PDSA) cycles, Lean Six Sigma, and the Model for Improvement, to facilitate the systematic identification, testing, and implementation of change ideas that lead to better patient care.[22]

In perioperative medicine, the adoption of improvement science has led to significant advancements in patient safety, clinical outcomes, and care efficiency. For example, the Enhanced Recovery After Surgery (ERAS) protocols, which integrate evidence-based best practices in perioperative care, have been shown to reduce complications, shorten hospital stays, and improve patient satisfaction.[23] Improvement science has also informed the development of surgical safety checklists, such as the World Health Organization's (WHO) Surgical Safety Checklist, which has been associated with substantial reductions in postoperative complications and mortality.[24]

Additionally, improvement science has played a crucial role in fostering a culture of safety and learning in perioperative medicine. Initiatives like the National Surgical Quality Improvement Program (NSQIP) and the Perioperative Quality Improvement Programme (PQIP) have facilitated the collection, analysis, and reporting of surgical outcome data, allowing healthcare providers to benchmark their

performance, identify areas for improvement, and implement targeted interventions.[25,26]

1.2 The role of context in quality improvement

The terms macro, meso, and micro context are widely used in healthcare to describe different levels of analysis within a system. These levels provide valuable insights into the factors that influence care delivery, outcomes, and quality, ranging from broader systemic factors to individual interactions. Understanding the interplay between these contexts is essential for effective quality improvement in perioperative care.

At the macro level, the focus is on the broader systems and structures that shape healthcare. This includes national and international policies, regulations, healthcare financing, public health initiatives, healthcare workforce trends, and social determinants of health. The macro context significantly impacts healthcare access, equity, and overall health outcomes.[27] For example, a country's healthcare policies and funding models play a crucial role in determining the availability and affordability of services. Social determinants of health, such as socioeconomic status and access to healthcare resources, contribute to disparities in perioperative care.[28–30] Addressing these macro-level factors is vital to create an environment that supports equitable and high-quality perioperative care.

At the meso level, the focus shifts to organisational and institutional factors that directly influence care delivery and quality. Resource allocation, care coordination, communication, and the implementation of evidence-based practices are key meso-level considerations.[31] Effective resource allocation ensures that necessary resources, including staffing, equipment, and facilities, are appropriately allocated to support perioperative care. The implementation of evidence-based protocols, like Enhanced Recovery After Surgery (ERAS), has shown significant impact in improving patient outcomes and enhancing care efficiency.[32] Multidisciplinary

collaboration within healthcare organisations plays a critical role in achieving quality improvement. Communication tools such as surgical checklists have been shown to improve outcomes after surgery.[24,33,34] By supporting collaboration between specialties and healthcare professionals, organisations can enhance care coordination, optimise clinical pathways, and improve patient outcomes.

Establishing a culture of safety and continuous learning within organisations is essential for successful quality improvement.[35] This involves creating an environment where all healthcare providers are encouraged to report errors, near misses, and adverse events. Implementing robust reporting systems and safety protocols, such as incident reporting mechanisms and surgical safety checklists, can help identify areas for improvement and facilitate the implementation of targeted interventions.[24,36] Additionally, institutions can foster a culture of continuous learning by promoting regular training and education for healthcare providers, encouraging participation in quality improvement initiatives, and actively involving frontline staff in the decision-making processes.

The micro context refers to the individual interactions and relationships between healthcare providers and patients. It encompasses clinical decision-making, communication, patient education, and the use of technology in care delivery. At the micro level, the interpersonal skills, empathy, and cultural competence of healthcare providers are vital.[37,38] Effective communication and shared decision-making promote patient engagement and positive perioperative outcomes.[39,40] Strong communication skills are essential for healthcare providers to understand patients' needs, preferences, and concerns, and to effectively convey information about procedures, risks, and benefits.[41] Patient education and involvement in care decisions enhance patient satisfaction and adherence to treatment plans.[42] The integration of technology such as electronic health records, telemedicine, and remote monitoring tools into perioperative care can streamline processes, enhance communication and documentation, and facilitate real-time access to patient information.

1.3 The emergence of large-scale quality improvement programmes

1.3.1 American College of Surgeons - National Surgical Quality Improvement Programme

The American College of Surgeons National Surgical Quality Improvement Programme (ACS-NSQIP) was initially developed in the 1990s as the National Veterans Affairs Surgical Quality Improvement Program to address concerns about the quality of surgical care in the Veterans Affairs hospital system in the United States.[43] After demonstrating significant improvements in surgical outcomes within the Veterans Affairs system, the programme was expanded to the private sector in 2001, and in 2004, it became the ACS-NSQIP under the guidance of the American College of Surgeons.[44]

Over the past two decades it has developed into a nationally validated, risk-adjusted, outcomes-based programme that aims to improve the quality of surgical care. The primary aims and objectives of the ACS-NSQIP include:

1. Data collection and risk adjustment through the collection of detailed, high-quality clinical data on surgical patients, including preoperative risk factors, intraoperative variables, and 30-day postoperative outcomes.[45] The programme utilises risk-adjustment models to allow for fair comparisons among participating hospitals, accounting for differences in patient populations and case complexity.
2. Benchmarking and performance feedback allowing them to benchmark their outcomes against national averages and identify areas for improvement.
3. Encourage quality improvement initiatives and support participating hospitals in implementing targeted quality improvement initiatives, based on their specific performance data and identified areas for improvement.

4. Improve collaboration and best practice sharing by fostering a collaborative environment, promoting the sharing of best practices and successful improvement strategies among participating hospitals, with the aim of accelerating the adoption of evidence-based interventions.

Studies have shown significant reductions in morbidity and mortality rates among hospitals participating in ACS-NSQIP.[25,46] Improvements in patient outcomes have also led to substantial cost savings.[25]

1.3.2 The Perioperative Quality Improvement Programme

The Perioperative Quality Improvement Programme (PQIP) is a large-scale, national initiative launched by the National Institute of Academic Anaesthesia - Health Services Research Centre (NIAA-HSRC), part of the Royal College of Anaesthetists (RCoA), in the United Kingdom in December 2016. The primary aims of PQIP are to improve patient outcomes and experiences following major elective surgery by collecting and analysing high-quality data on perioperative care processes and patient outcomes, enabling hospitals to understand their performance and implement evidence-based improvements.[26]

Patients aged 18 and over, who are planned to undergo major elective surgery under the care of selected surgical specialties (see Appendix A-1) are eligible for recruitment to the PQIP study. Patients are required to provide informed consent to be included in the study. The study is open to all NHS sites in England and Wales, on an opt-in basis. During the setup phase for each site local investigators are asked to confirm which surgical specialties they will recruit patients from. They are also asked to confirm if they plan to recruit all eligible patients or a random selection of patients based on an 8-day rolling recruitment cycle.

To support local quality improvement, a range of site-specific quarterly reports are produced outlining specialty and institution level performance against a range of perioperative process and outcome measures. National level reports are also

produced on an annual cycle to describe the wider perioperative landscape and to support the setting of national improvement priorities.[47–50]

The Perioperative Quality Improvement Programme has four main objectives:

1. Develop a comprehensive dataset that includes process and outcome measures relevant to perioperative care, including information on patient demographics, comorbidities, surgical procedures, anaesthetic management, postoperative care, and patient-reported outcomes.
2. Facilitate local quality improvement by collecting and analysing data from participating sites. Through the provision of individualised feedback to hospitals support them to identify areas in need of improvement, benchmark their performance against national averages, and develop targeted quality improvement initiatives.
3. Conduct collaborative research and quality improvement by encouraging participating hospitals to work together to generate new knowledge and share best practices. This collaborative approach aims to accelerate the implementation of evidence-based interventions and drive improvements in perioperative care at a national level.
4. Inform health policy and commissioning through the generation of robust evidence on the effectiveness of various perioperative interventions and care pathways, which can inform health policy and commissioning decisions. This evidence base can support the development of national guidelines, standards, and quality indicators, promoting consistency and excellence in perioperative care across the UK. An example of this is the commissioning for quality and innovation (CQUIN) for successful delivery of postoperative targets for patients drinking, eating and mobilising within 24 hours of surgery.[51,52]

Following a second five-year funding agreement in 2021, the Perioperative Quality Improvement Programme has continued to expand its scale and influence as a

national quality improvement initiative. The fifth national PQIP report, published in September 2024, reported participation from 173 NHS hospitals, representing a marked increase from the approximately 70 hospitals that initially expressed interest at the programme's inception.[53] During this period, patient recruitment has also grown significantly. By the time of the fifth cohort report, a total of 53,478 patients had consented to participate in the study. This increased recruitment has enhanced the representativeness of the dataset, strengthening the generalisability and reliability of PQIP's as a source of national insight into perioperative care.

The expansion in both site and patient recruitment supports more comprehensive comparison across NHS trusts and facilitates the identification of trends and system-wide challenges. This breadth of data enables the generation of more robust quality improvement initiatives and has supported the development of evidence-informed national policy such as the Commisioning for Quality and Innovation (CQUIN) introduced in 2022 promoting early drinking, eating and mobilising (DrEaMing) after surgery.[54] The continued scale-up of the programme reflects an increasing recognition of the value of detailed, patient-level data in supporting local and national improvement efforts

1.3.3 The National Emergency Laparotomy Audit

The National Emergency Laparotomy Audit (NELA) is a large-scale clinical audit in the United Kingdom launched in 2013, aimed at evaluating and improving the care provided to patients undergoing emergency laparotomy surgery. The audit is funded by the Healthcare Quality Improvement Partnership and managed by the RCoA and the NIAA-HSRC.

NELA aims to assess the quality of care provided to patients undergoing emergency laparotomy surgery by collecting and analysing data on patient demographics, surgical and anaesthetic processes, postoperative care, and clinical outcomes, such as mortality and morbidity.[12,55] By benchmarking hospital performance against

national standards and guidelines, NELA seeks to identify areas where care can be improved, thereby promoting a culture of continuous quality improvement. Through provision of individualised feedback on hospital performance the audit facilitates sharing of best practices among participating institutions. This feedback and support enable hospitals to develop targeted quality improvement initiatives and implement evidence-based interventions to improve patient care.[12,56–58] Several improvements in patient care have been achieved through NELA. It has contributed to a significant reduction in 30-day mortality for patients undergoing emergency laparotomy surgery, from 11.8% in 2013 to 9.5% in 2019.[59] The audit has also led to increased adherence to evidence-based care processes, such as preoperative risk assessment, timely surgical intervention, and appropriate postoperative care.[59]

1.4 Outcome measures for monitoring healthcare quality

Outcome measures are crucial tools for assessing the quality-of-care delivered by hospitals and clinical teams. Outcome measures commonly used to monitor the quality of healthcare delivery include mortality rates, morbidity or complication rates, and patient-reported outcome measures (PROMs).

Mortality rates are a widely used outcome measure, as they are objective and easily quantifiable. However, they may not capture the full range of patient experiences, particularly in the context of non-fatal complications or the impact of care on long-term functional status.

The number of patients at high-risk of adverse postoperative outcomes has grown substantially in recent years: this is attributable to a combination of an ageing population, the increased numbers of surgical options available for previously untreatable conditions, and the increasing numbers of patient presenting for surgery with multiple comorbidities. Estimates of inpatient mortality after elective non-cardiac surgery are low, ranging between 0.5 and 3.6% depending on the type of surgery and patient related risks[60–62] Major or prolonged postoperative morbidity

(for example, significant infections, respiratory or renal impairment) occur in a much higher number of patients (up to 15%), and are associated with reduced long-term survival and worse health-related quality of life; this reduced long-term survival associated with postoperative morbidity has been consistently demonstrated across different types of surgery, patient and healthcare system[62,63]. The high prevalence of morbidity in comparison to mortality after major surgery has led a drive to implement improvement strategies focussed on reducing the burden of complications.[64]

Mortality data in isolation, without detailed morbidity or process data provide information which can be used for quality improvement. If the cause of deterioration (e.g., pulmonary or renal morbidity) is not known, it is not possible to target specific interventions to reduce those complications. Rates of morbidity also vary substantially by organ system – for example, cardiac morbidity following non-cardiac surgery is uncommon, whereas pain, renal and gastrointestinal morbidity is much more common.[62,65–67] With the collection and reporting of morbidity data it is also possible to define 'failure to rescue' rates – i.e., the rate of death in hospital after developing a complication (Figure 1-1).[60,68] Data from the US demonstrate wide variation in morbidity and failure to rescue rates between healthcare providers.[69,70] Despite this variation, success in improving patient level outcomes has been variable in both the US and in improvement trials in the UK.[71–74] More recently, mixed method evaluations of improvement programmes have started to provide useful insight into why positive change may not have been delivered.[64,75]

Postoperative morbidity and complications may be considered a modifiable outcome measure which could be a useful measure of quality of care, occurring in a greater proportion of patients compared to failure to rescue or postoperative mortality. However, postoperative morbidity is not a clear binary outcome, like mortality, and is therefore more subjective and difficult to define. To address these potential issues classification systems have been developed to objectively define types of morbidity and postoperative complications. Two commonly used systems

are the Postoperative Morbidity Survey (POMS) and Clavien-Dindo grading of surgical complications.[76,77]

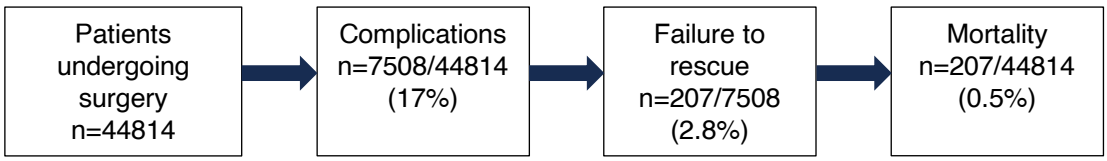


Figure 1-1: Conceptual pathway of mortality after surgery incorporating the concept of 'failure to rescue'. The numbers and percentages are shown to provide context for each stage of the pathway, taken from Pearse et al.[60]

1.4.1 The Postoperative Morbidity Survey

The Postoperative Morbidity Survey (POMS) is a simple, validated tool designed to capture postoperative morbidity which leads to extended hospital stays following surgery.[65,66,76,78] The POMS can be completed at a range of postoperative timepoints, typically between days three and fourteen following surgery. The POMS collects information on nine domains (see Table 1-1). The survey is used as a composite measure, where the presence of defined morbidity in any of the nine domains results in a patient being recorded as having postoperative morbidity.

The POMS has been demonstrated to be a reliable and valid tool for measuring outcomes after major abdominal surgery ensuring accurate and consistent results when assessing patient outcomes.[66] It offers a standardised approach to measuring postoperative morbidity, which allows for easier comparison of outcomes across different hospitals or surgical teams, and by providing a detailed overview of postoperative complications, the POMS can help healthcare professionals better understand the various factors contributing to patient morbidity and prolonged postoperative hospital length of stay.[65] This information can be used to identify trends, patterns, and potential areas for improvement in perioperative care. The POMS has been used in variety of surgical contexts, including routine, moderate-

risk elective surgery, emergency laparotomy, and major abdominal surgery.[66,76,78] The presence of postoperative morbidity identified by the POMS has also been associated with an increased risk of mortality for up to 3 years following surgery.[62]

Table 1-1: Postoperative morbidity survey domains and criteria

Organ system	Morbidity criteria
Pulmonary	New requirement for oxygen
Pulmonary	New requirement for respiratory support
Infectious	Currently on antibiotics
Infectious	Temperature >38°C in the last 24hr
Renal	Urinary catheter in situ
Renal	Increased serum creatinine (>30% from preoperative level)
Renal	Presence of oliguria <500 mL/24hr
Gastrointestinal	Unable to tolerate an enteral diet for any reason
Gastrointestinal	Vomiting or abdominal distension, or use of antiemetics
Cardiovascular	Thrombotic event requiring anticoagulation (new)
Cardiovascular	Atrial or ventricular arrhythmias (new)
Cardiovascular	Hypotension (requiring pharmacological or fluid therapy >200 mL/hr)
Cardiovascular	New myocardial infarction or ischaemia
Cardiovascular	Cardiogenic pulmonary oedema
Neurological	New coma
Neurological	New confusion or delirium
Neurological	New focal neurological deficit
Haematological	Platelet, fresh-frozen plasma, or cryoprecipitate transfusion in last 24hrs
Haematological	Packed erythrocyte transfusion in the last 24hrs
Wound	Wound dehiscence requiring surgical exploration or drainage of pus from the operation wound with or without isolation of organisms
Pain	New pain significant enough to require parenteral opioids
Pain	New pain significant enough to require regional analgesia

Increased collection and use of POMS data may allow risk stratification of patients based on their risk of developing postoperative complications. Such information may improve patient understanding of their risk prior to surgery, leading to clearer consent conversations and improved shared decision making when compared to discussion of mortality risk alone. The perioperative care patients receive could also be tailored to their individual risk, optimising resource allocation to those at the highest risk of poor postoperative outcomes. Risk-adjustment models for postoperative morbidity and complications have been developed to assist with both patient-centred discussions and the monitoring of surgical outcomes.[67,79]

Despite the potential of the POMS to identify important morbidity, its use in clinical practice has been generally limited to research settings. In part, this is due to the detailed information needed across its nine organ domains, which can be time-consuming and resource-intensive for healthcare providers to collect. With increasing availability of electronic healthcare records, there has been attempts to capture morbidity data directly from healthcare records without the need for manual data collection.[80–82] If such methods are proven to be reliable in capturing significant postoperative morbidity, it is likely that a version of POMS becomes more widely used in routine clinical practice.

Another potential limitation of the POMS is whether it is applicable to all surgical specialties. Some elements of morbidity, such as the presence of a urinary catheter, may be considered routine following major urological surgery, for example. This has led some to question whether the POMS is too sensitive in the setting of major surgery, and whether only certain subgroups of morbidity should be included when used in certain surgical cohorts.[67]

1.4.2 Clavien-Dindo Grading of Surgical Complications

The Clavien-Dindo system of measuring surgical complications categorises adverse postoperative events into five grades based on their severity and the level of intervention required to treat them (see Table 1-2).[77] The system is designed to be simple, reproducible, and applicable to various surgical specialties, allowing comparisons across different studies and institutions.[77] The presence of complications defined by the grading system was strongly correlated with hospital length of stay in the development study and it has been widely adopted as a standard for reporting surgical complications in clinical research.[77,83,84]

Table 1-2: The Clavien-Dindo grading of surgical complications

Grade of complication	Intervention required to manage complication
I	Minor complications that do not require any specific intervention or treatment. These include minor infections, postoperative nausea or vomiting, and mild pain.
II	Complications requiring pharmacological treatment, blood transfusions, or total parenteral nutrition (TPN). These complications are more severe than Grade I but can be managed without invasive procedures.
III	Complications requiring surgical, endoscopic, or radiological interventions. This grade is further subdivided into two categories:
IIIa	Complications requiring intervention under local anesthesia or conscious sedation (e.g., wound debridement).
IIIb	Complications requiring intervention under general anesthesia (e.g., reoperation for bleeding or anastomotic leak).
IV	Life-threatening complications requiring intensive care management. This grade is also subdivided into two categories:
IVa	Single organ dysfunction (e.g., renal failure, respiratory failure).
IVb	Multiple organ dysfunction (e.g., septic shock, multiple organ failure).
V	Death of the patient due to a complication.

A strength of the Clavien-Dindo grading system is the relatively objective way in which complication, therefore supporting interobserver reliability. It also encompasses a wide range of complications from minor to major. Data are usually recorded at hospital discharge meaning that complications occurring at any timepoint in the postoperative period are recorded. This contrasts with the POMS, which may miss morbidity that was not present at the timepoint the survey was performed but instead occurred before or after.

The standardised format and reproducible structure of the Clavien-Dindo system means that outcomes can be collected and compared between institutions, supporting collaboration and quality improvement work.[60] The applicability of the Clavien-Dindo system to low- and middle-income countries may be questioned however. A large study using the Clavien-Dindo system found that complications in low- and middle-income countries were under reported compared to when a different classification system was used.[85] This may be related to the availability of resources to manage complications. For example, if a country has a low provision

of intensive care services, fewer patients are likely to be managed in an intensive care unit. Using the Clavien-Dindo grading system this may result in complications being graded differently compared to a resource rich setting where more patients may be admitted to an intensive care area.

Further difficulties arise when using the system in emergency surgery, where patients may require organ support in the preoperative period and therefore the postoperative need for organ support may not be considered a direct complication of surgery.[86] There is also limited evidence on how complications graded within the Clavien-Dindo system impact on patients' long-term outcomes and quality of life, potentially limiting its prognostic value.[46]

Finally, the system also grades complications based on the intervention needed to treat them but does not offer more detailed information about the type of complications experience by patients. For example, if all patients undergoing colorectal surgery at an institution required postoperative ventilation for a hospital acquired pneumonia (grade IVa complication), this would be recorded in the same way as patients requiring postoperative ventilation due to a major cerebrovascular accident. Clearly, it is important and relevant for clinicians and institutions to understand what type of complications are occurring, not just the treatment required to manage them. Quality improvement initiatives to improve the outcomes for these two complications would focus on very different care processes.

1.5 Risk-adjustment of healthcare outcomes

To compare perioperative outcomes between clinicians, institutions, and healthcare systems it is important to take patient and surgical risk factors into consideration, both of which influence the likelihood of adverse postoperative outcomes. An elderly patient with multiple medical comorbidities is understandably at greater risk of complications following major surgery compared to a younger, fit and well patient.

Similarly, a patient undergoing minor surgery would be expected to be at lower risk of postoperative complications than the same patient undergoing major surgery.

Risk adjustment is the process of trying to account for differences in patient and surgical characteristics that could potentially influence outcomes. The concept of risk-adjusted outcome monitoring has been present in healthcare since the early 20th century, with early efforts focused on adjusting for patient characteristics such as age and comorbidities to compare outcomes across different populations or institutions. In the 1980s and 1990s, risk adjustment methodologies became more sophisticated with the development of statistical techniques that could account for numerous factors simultaneously.[87] Multivariable regression modelling is a common method used to adjust for patient risk with results often presented in funnel plots or using the standardised mortality ratio chart.[88,89]

Figure 1-2 shows an example funnel plot. Each point on the plot represents a hospital with the number of operations performed during the time frame under consideration shown on the x-axis. The risk-adjusted mortality is shown on the y-axis. Curved lines represent two and three standard deviations from the mean risk-adjusted mortality rate.

A strength of funnel plots is that they allow for greater variation in outcomes in smaller volume centres compared to large volume centres before an institution is classed as an outlier (points above or below the standard deviation curves). In a small volume centre, chance may play a greater role in variation of outcomes than that of larger centres with a higher number of patients. Traditional cross-sectional analyses, however, condense information to a single value or point on a graph, such as the overall mortality rate for a hospital unit over a one-year period. It is important to note, the identification of a hospital as an outlier may vary depending on the time considered. An institution may be identified based on analysis that includes a 6-month cohort of patients, but this may not be true in the preceding or following 6-month period. Variation in outcomes within the period considered is also

not visible in a funnel plot, for example if outcomes were significantly better than expected in the first half of the period and poorer in the second half, this may average out over the whole period. There may or may not be important learning over both halves.

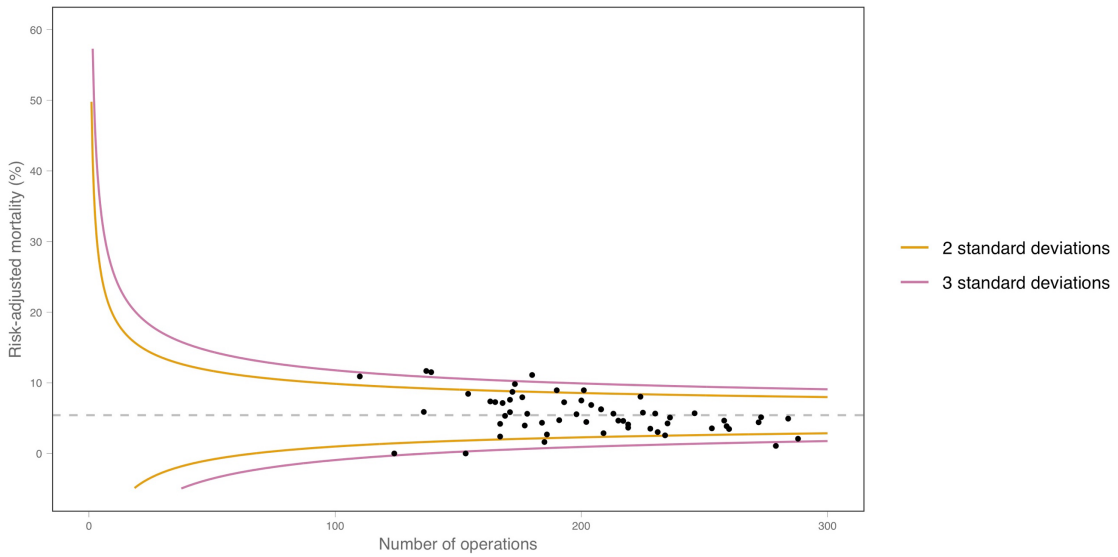


Figure 1-2: Example funnel plot showing the number of operations performed at each institution (x-axis) against the risk-adjusted mortality rate (y-axis). The grey horizontal dashed line shows the mean risk-adjusted mortality across all hospitals in this fictional sample.

1.5.1 Continuous outcome monitoring charts

Continuous monitoring charts have been adopted into healthcare from the manufacturing industry where they have been used for decades. Shewhart control charts were first developed by Walter A. Shewhart in the 1920s, to monitor variation in manufacturing practice at the Western Electric Company.[90] The Shewhart chart, or statistical process control chart is a type of continuous chart used to distinguish assignable or 'common-cause' variation and chance-cause or 'uncommon-cause' variation.[90]

A benefit of continuous analysis over cross-sectional, retrospective analysis is that such methods allow the identification of variation in outcome over time. The

incorporation of risk-adjustment into such charts allows clinicians, administrators, and researchers to assess the quality of care delivered, identify areas for improvement, and track progress in meeting quality and safety targets. Continuous display methods may support the earlier identification of positive or negative trends in outcomes, learning from positive deviance or preventing further deterioration of negative trends.

There are several published methods to incorporate risk-adjustment into continuous monitoring charts, these include but are not limited to: the variable life-adjusted display (VLAD); risk-adjusted exponentially weighted moving average (RA-EWMA) chart; the risk-adjusted cumulative sum chart (RA-CUSUM); and the cumulative risk-adjusted mortality (CRAM) chart.[91–94] The CRAM, VLAD, and RA-CUSUM charts employ similar statistical methods, showing the difference between expected and observed outcomes over time.

1.5.1.1 The variable life-adjusted display

The VLAD is a graphical method for monitoring healthcare outcomes that was first introduced in the 1990s.[91] It is a type of cumulative sum (CUSUM) chart that represents the difference between the observed and expected outcomes over time. The VLAD chart displays this difference on the vertical axis, with the horizontal axis usually representing the chronological order of patients. The chart's slope reflects the performance of the healthcare provider, with an upward slope indicating better-than-expected outcomes and a downward slope representing worse-than-expected outcomes, allowing healthcare providers to identify areas for improvement. VLAD analysis can also be used to compare the performance of different healthcare providers or institutions, helping to identify best practices and promote quality improvement across the healthcare system.[95–97] VLAD charts are particularly useful for monitoring rare events, such as postoperative mortality, and can provide early warnings of changes in performance that may warrant further investigation.

They have been widely used in healthcare settings for monitoring clinical performance, particularly in the field of cardiac surgery.[91,95,98] An advantage of VLAD charts is that they can be easily understood by both medical professionals and non-experts, making them a valuable communication tool for quality improvement initiatives.[95] A potential limitation of the VLAD is its focus on binary outcomes, initially developed to monitor mortality, although their use has been expanded to monitor other binary outcomes.[91,98]

The accuracy of VLAD relies on the precision of risk estimates for the expected mortality rates.[94] If the underlying risk prediction models do not adequately account for patient-specific factors, resulting in patient risk being systematically under- or over-estimated the VLAD may not accurately reflect true clinical performance. While the graphical representation of data in VLAD can facilitate understanding, it may also lead to misinterpretation or overemphasis on short-term fluctuations in performance.[89] This can potentially result in undue focus on temporary variations at the expense of long-term trends.

Figure 1-3 shows the VLAD of a theoretical hospital, showing risk-adjusted mortality after surgery. With data displayed in VLAD form it is possible to see variation in outcomes over the 500-case series. After the first 200 or so cases performance is the same as expected (orange VLAD line is around zero). However, after this there is then and upward trend in the VLAD, climbing to around 10 by case 290 (black circle), representing 10 excess deaths over these 90 cases compared to that expected by the risk-adjustment model. Following this the VLAD returns to zero, before continuing to fluctuate around zero for the remainder of the case series. The overall mortality rate is 8.0% for the entire case series, which may be in keeping with expected results. Using a continuous display however, it is possible to identify variation in outcomes over time that is hidden in a cross-sectional analysis.

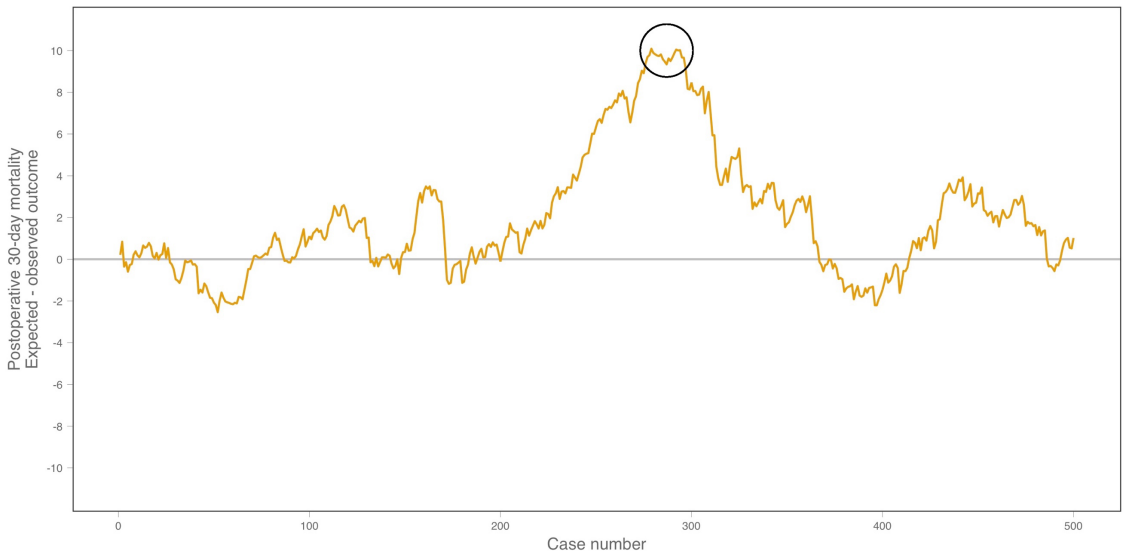


Figure 1-3: Example VLAD showing expected vs. observed mortality of a 500-case series

1.5.1.2 The risk-adjusted exponentially weighted moving average chart

Risk-adjusted exponentially weighted moving average chart (RA-EWMA) charts build upon the principles of traditional EWMA charts, which plot the moving average of a process variable over time. The EWMA chart uses a weighted average of outcome data to estimate current performance, with more recent data points given greater weight than older data points. The EWMA charts can be customised by adjusting the weighting parameter, which determines the degree of responsiveness to recent changes in performance. A higher weighting parameter makes the chart more sensitive to recent trends, while a lower parameter emphasizes long-term performance.[92,99] This method provides a more sensitive and responsive analysis of changes in performance compared to traditional control charts.

In a risk-adjusted EWMA chart, the expected outcome is plotted together with pre-specified control limits around the expected outcome. A EWMA of the observed outcome is also shown. Where the observed outcome crosses the pre-specified control limits the RA-EWMA chart is said to ‘signal’ a shift in the odds ratio of the

outcome. The control limits are set to detect a shift in the odds ratio of the outcome, which is predetermined when the monitoring chart is set up. These charts can be used to monitor a variety of outcomes, from clinical complications to process measures, and can help detect gradual changes in performance that may not be apparent using traditional methods. Risk-adjusted EWMA charts have been adopted by the Intensive Care National Audit and Research Centre (ICNARC) and the National Emergency Laparotomy Audit (NELA) as tools to monitor mortality outcomes in both national audits.[12,100,101]

Figure 1-4 shows an example of a RA-EWMA chart, with the x-axis showing the date and y-axis representing the mortality rate (percent). The EWMA of the expected outcome is shown by the pink line, the blue area around this represents the control limits. The observed outcome is shown by the orange line. Where the observed outcome EWMA line crosses the control limits of the expected outcome, the RA-EWMA chart is said to 'signal', here shown by the black circle.

An advantage of the RA-EWMA methodology is that the charts are more sensitive to detecting small shifts in performance compared to traditional statistical control methods, such as Shewhart control charts.[92,102] The RA-EWMA chart allows for the adjustment of weighting factors based on the specific application, enabling a balance between responsiveness to changes and stability against random variations.[103] This flexibility allows for the customisation of the chart to suit the needs of different healthcare settings and outcome measures. The EWMA chart provides a comprehensive view of performance trends by incorporating information from multiple data points, rather than focusing on individual measurements.[99] This aggregation of data can facilitate the identification of patterns and trends that may not be apparent in isolated data points.

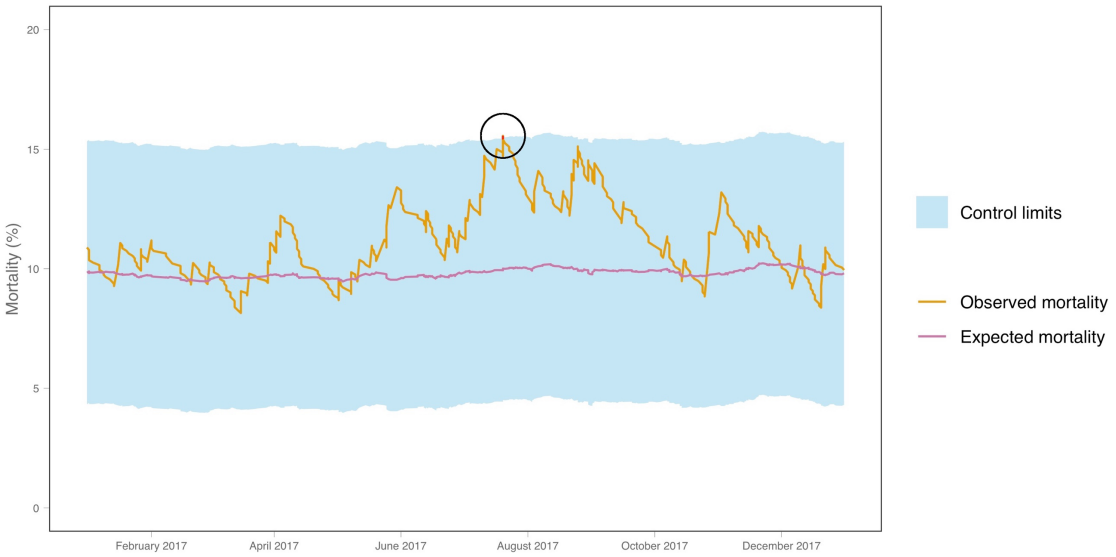


Figure 1-4: Exponentially weighted moving average chart showing observed and expected mortality outcomes

A potential disadvantage of the RA-EWMA chart is its more complex methodology and calculations compared to traditional methods, such as Shewhart control charts or the VLAD.[103] Interpreting EWMA charts may also be more challenging for healthcare professionals due to the additional weighting factors and the need to understand the implications of varying weights on the chart's responsiveness.[92,99] This complexity may hinder its adoption in healthcare settings where simplicity and ease of use aid successful implementation.

1.5.2 Applications of risk-adjusted monitoring charts in healthcare settings

By incorporating a measure of patient risk in their production the VLAD, RA-EWMA, and other continuous charts account for differences in patient characteristics that could potentially influence outcomes. In adjusting for these differences, the charts support the comparison of outcomes between healthcare providers and helps ensure that observed differences in performance are truly reflective of the quality of care, rather than being driven by variations in patient populations.

The continuous nature of graphical displays, either presented by case number or date, provides a visual representation of performance trends. Such charts help engage clinicians and administrators in the process of continuous improvement, fostering a culture of accountability and collaboration within healthcare organisations.[9] They enable healthcare providers to identify specific areas where interventions are needed, track the impact of those interventions over time, and refine their approach to quality improvement based on the data-driven insights these charts provide. They may also enable earlier identification of potential outlying institutions, supporting earlier intervention to improve outcomes, and improve transparency and communication.[98,104]

1.6 Conclusion

Quality in healthcare is a multifaceted concept, underpinned by dimensions such as safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity, with global bodies such as the Institute of Medicine and the World Health Organization providing guiding principles. In the UK, the Care Quality Commission employ these dimensions to evaluate healthcare service quality. Donabedian's framework offers a triple approach - structure, process, and outcome - to understanding and assessing the quality of healthcare delivery. Quality assurance practices, such as the national audits established and supported by the Healthcare Quality Improvement Partnership (HQIP), facilitate the regular monitoring of care against predefined standards. However, it is important to recognise that solely relying on these might not always lead to continuous improvement in the quality of care delivered by hospitals and healthcare systems. Quality improvement, however, does seek continuous enhancement in the quality of healthcare delivery, drawing methods of analysis from the field of improvement science. With its origins in industrial quality management principles, improvement science has driven significant advancements in areas like perioperative medicine through methods like Plan-Do-Study-Act cycles and initiatives such as Enhanced Recovery After Surgery

protocols. Such methodologies and tools ensure ongoing efforts to promote safety, optimise patient outcomes, and improve the overall care experience.

Healthcare delivery is complex and multifaceted. Context plays a significant role in the success or failure of improvement initiatives. Understanding the macro, meso, and micro context of a system can support successful implementation and the embedding of new practice into clinical care. The macro context, governed by overarching policies, regulations, and socio-economic determinants, significantly influences healthcare accessibility, equity, and outcomes.[27–30] The meso context delves into the organisational and institutional aspects that influence care, emphasising the significance of resource allocation, multidisciplinary collaboration, and evidence-based practices in achieving quality.[31,32] Fostering a culture prioritising safety, continuous learning, and active frontline staff involvement sets the stage for meaningful quality improvement initiatives.[35,36] At the micro level, individual interactions between healthcare providers and patients come to the forefront. Components such as effective communication, patient education, technological integration, and empathy play integral roles in determining patient engagement and positive outcomes.[37–39,41,42]

The rise of large-scale quality improvement programmes, notably the ACS-NSQIP and PQIP, demonstrates a concentrated effort towards enhancing surgical care quality on both national and international fronts. The ACS-NSQIP, developed to address quality issues in the Veterans Affairs system, has evolved into a large scale initiative that reflects the increasing importance of data-driven strategies, benchmarking, and collaborative practices, with demonstrable reductions in morbidity, mortality, and costs.[43–46] The PQIP reflects a similar drive in the UK a to improve patient outcomes after major surgery. Through comprehensive data collection, collaborative research, and informed policy-making, it aims to achieve national improvements in standards of perioperative care.[26,50,52,105] Such programmes demonstrate an increasing global commitment to elevating surgical care standards through systematic, data-centred approaches.

Outcome measures are vital ways of assessing the quality of healthcare delivered by hospitals and clinical teams. While mortality rates remain a commonly used, unambiguous metric, they do not capture the full scope of patient outcomes and experiences, especially postoperative complications or the longer-term impact of care.[60–62] There has been a notable shift towards recognising the importance of morbidity data, with its incidence significantly higher than mortality, necessitating strategies focused on reducing postoperative complications.[62–64,67] Given this shift, tools such as the POMS and the Clavien-Dindo grading of surgical complications have been developed to provide standardised methods of assessing morbidity and surgical complications.[65,76,77]

The POMS offers a detailed view of postoperative complications, promoting comparisons across various surgical contexts and offering the opportunity to monitor trends in outcomes for quality improvement.[65,66,76,78] Despite its potential, the survey's applicability across surgical specialties remains debated, and its extensive data requirements can be burdensome, hindering its routine clinical adoption.[67,80–82] The Clavien-Dindo system categorises complications by their severity and required interventions, ensuring a consistent and potentially cross-specialty method of recording postoperative complications.[76,83,84] However, the system might not fully reflect the complexities of postoperative complications in different contexts, especially in resource-constrained settings, and lacks granularity in identifying the exact nature of complications.[85,86]

In healthcare, risk-adjustment is crucial when comparing perioperative outcomes across clinicians, institutions, and systems. This approach accounts for patient and surgical factors which can influence postoperative results using statistical techniques such as multivariable regression modelling, which allows for comprehensive adjustments.[87,88]

Funnel plots, for example, visualise risk-adjusted mortality rates for hospitals, but can mask variability within the timeframe analysed. Continuous outcome monitoring

charts originated in the manufacturing sector and were first developed by Shewhart in the 1920s.[90] More recently, risk-adjustment methodology has been incorporated into some types of continuous charts which have been adopted into healthcare settings. These charts, such as the VLAD and RA-EWMA, allow for a dynamic assessment of patient outcomes over time. Both chart types have their advantages and disadvantages. The VLAD is particularly useful in monitoring rare events and provides an intuitive graphical representation.[91,95,97] The RA-EWMA chart is sensitive to small shifts in performance and offers the flexibility of customisation of control limits.[92,99] Incorporating patient risk measurements, these graphical tools facilitate comparisons between healthcare providers by accounting for potential disparities in patient characteristics. By offering a visual trend of performance, these charts encourage a collaborative culture and may support continuous quality improvement in healthcare settings. They support clinicians to identify areas for enhancement and to measure the effects of interventions on patient outcomes. This approach, underpinned by high quality data can enhance transparency, improve communication, and raise the quality of healthcare delivery.[9,98,104]

Chapter 2 Prospective monitoring of risk-adjusted outcomes using variable life-adjusted displays: a systematic review

2.1 Introduction

The variable life-adjusted display (VLAD) was developed to report expected minus observed mortality after cardiac surgery.[91] Other forms of expected minus observed charts have also been reported including the cumulative risk-adjusted mortality (CRAM) chart[94], the expected minus observed (E-O) chart, their reciprocal observed minus expected (O-E) chart, and the risk-adjusted cumulative sum chart (RA-CUSUM). Whilst RA-CUSUM charts traditionally report the log-likelihood ratio[106,107], charts showing expected minus observed outcomes are occasionally also referred to as RA-CUSUM. For consistency all continuous outcome displays which use the expected minus observed methodology will be referred to as VLADs within this chapter, even where a different nomenclature was used in the original study.

The focus of this chapter is to identify where VLADs have been implemented prospectively to monitor one or more outcomes in a healthcare setting, understand the context of these implementations and identify factors that act as facilitators and barriers to implementation. To do this the chapter is divided into a systematic review of the peer-reviewed literature and a targeted grey literature review.

2.2 Aim and research questions

The aim of this review is to summarise the evidence on the use of continuous patient-level outcome monitoring in a healthcare setting. The objectives are to:

1. Conduct a systematic review to identify and summarise studies reporting the prospective implementation of continuous patient-level healthcare outcome monitoring using VLADs
2. Describe methods to embed the use of continuous patient-level healthcare outcome monitoring tools such as VLADs in routine clinical practice

3. Assimilate potential facilitators and barriers to the implementation of prospective monitoring using VLAD

The research questions for this review are:

1. In what healthcare settings have variable life-adjusted displays (VLADs) been used to prospectively monitor an outcome?
2. What outcomes have VLADs been used to monitor?
3. What methods are used to embed VLADs into quality assurance or quality improvement initiatives?
4. What factors act as facilitators and barriers to the implementation of VLADs as a monitoring tool?

2.3 Methods

2.3.1 Definitions

Graphical displays that plot expected (using a risk-adjustment model) minus observed outcomes were considered for inclusion. Charts reporting outcomes in this way are sometimes called VLAD, CRAM, CUSUM or RA-CUSUM. Studies were considered for inclusion regardless of the name used for the chart in the manuscript, as long as they employed the calculation and display of the expected minus observed outcome. Studies that used RA-CUSUM charts showing the log-likelihood ratio of the outcome were excluded as this methodology is distinct from VLADs. To aid consistency throughout this chapter all charts displaying an expected minus observed calculation will be referred to as VLADs. The term VLAD will be used regardless of the term used within the included study.

2.3.2 Search strategy

A search of published literature between 1st January 1995 and 17th July 2018 was performed using multiple databases: MEDLINE, EMBASE, Web of Science, CINAHL Plus and the Cochrane Collaborative. The search terms were:

1. ("Variable life-adjusted" or VLAD)
2. ("cumulative sum" or CUSUM)
3. ("cumulative risk-adjusted mortality" or CRAM)
4. (expected AND minus AND observed)
5. 1 or 2 or 3 or 4 AND (monit* or measur* or perform*)

The search was conducted using all fields within each database. Results were combined and uploaded to the online systematic review software Covidence[108]. Duplicates were removed using automatic recognition within the software and manually.

A citation search of the first papers to describe the VLAD chart[91] and the CRAM chart[94] was carried out. The references and citations of all included papers were also screened for possible inclusion and to identify any evaluations of the included studies that were carried out.

2.3.3 Selection

Articles identified through the search were screened in two stages; initially by title/abstract and then a full-text review. The title/abstract screening was performed by one assessor (JB screened ~4250 abstracts, with a second reviewer AI screening the remaining ~900). Full-text screening was performed independently by the two assessors (JB and AI). Any conflicts relating to the inclusion or exclusion of articles was resolved through discussion between assessors.

No restriction was placed on the language studies were written in. Articles published in languages other than English were translated using Google Translate®.[109]

Inclusion criteria:

Studies were included if they met all of the following inclusion criteria:

1. The study reported on the implementation of prospective continuous monitoring of a patient related healthcare outcome;
2. The study used VLAD methodology (or equivalent) to monitor the outcome;
3. Peer reviewed

Exclusion criteria:

Studies which applied VLADs retrospectively or to demonstrate the learning curve of a procedural skill were excluded. Conference abstracts, letters, and case reports were also excluded.

2.3.4 Data extraction, analysis and quality assessment

Included articles were analysed using a data extraction table developed in Microsoft® Excel 2016 (see Appendix B-1). Numerical and categorical data were extracted as raw data. Data that could not be assigned to categories were extracted as free text. The data extraction form was piloted using a random sample of four articles.

Once extraction was complete, data were imported into NVivo qualitative data analysis Software[110]. Free text data was analysed using a framework method[111] based on the research questions. Subcategories were developed based on topics identified in the data and themes were created to illustrate key patterns. Numerical and categorical data were analysed within the Microsoft® Excel worksheet.

The quality of studies was assessed using the evaluation tool for mixed method study design developed by Long et al.[112] and the Mixed Methods Appraisal Tool (MMAT).[113]

The mixed method evaluation tool[112] is based on critical appraisal tools developed for quantitative and qualitative studies.[114] It combines relevant questions from the two evaluation tools and was produced to support systematic reviews of research

literature on effectiveness and outcomes in social care but is applicable to healthcare settings.

The MMAT is an efficient tool to appraise the methodological quality of qualitative, quantitative and mixed methods studies concomittently.[115] It was developed using peer-reviewed mixed methods literature and has been piloted and refined to improve its reliability.[115–117] The MMAT and its user manual are openly available online.[118] An abbreviated version of the MMAT assessment tool showing questions relevant to this review is shown in Appendix B-2.

Following two initial screening questions, the MMAT then appraises each study based on its methodology. There are specific qualitative and quantitative question domains to be completed dependent on the study design being appraised. Appraisal of mixed method studies combines quantitative and qualitative questions.

A summary rating, with a range from one to four stars can be calculated based on the answers given to component questions. It is important to note that for mixed method studies the summary rating is determined by the lower score of the two methodological parts (the qualitative or quantitative part). Hence a mixed method study with a strong quantitative methodology might not achieve a high summary rating due to poor qualitative methodology, and vice versa.

2.4 Results

2.4.1 Search results

The database, citation, and reference searches identified 5156 unique articles (see Figure 2-1: PRISMA flow diagram of systematic review). Initial assessment by title and abstract reduced the number of articles to 222 prior to full-text review. After full-text screening, 22 articles (based on 21 studies) were included in the review. Citation and reference searches of included papers identified a further 585 articles that were screened for possible inclusion. All 585 of the additional articles were excluded, resulting in 22 articles (based on 21 studies) being included in the review.

All studies included in the analysis were prospective observational cohort studies. Table 2-1 summarises the characteristics of studies included.

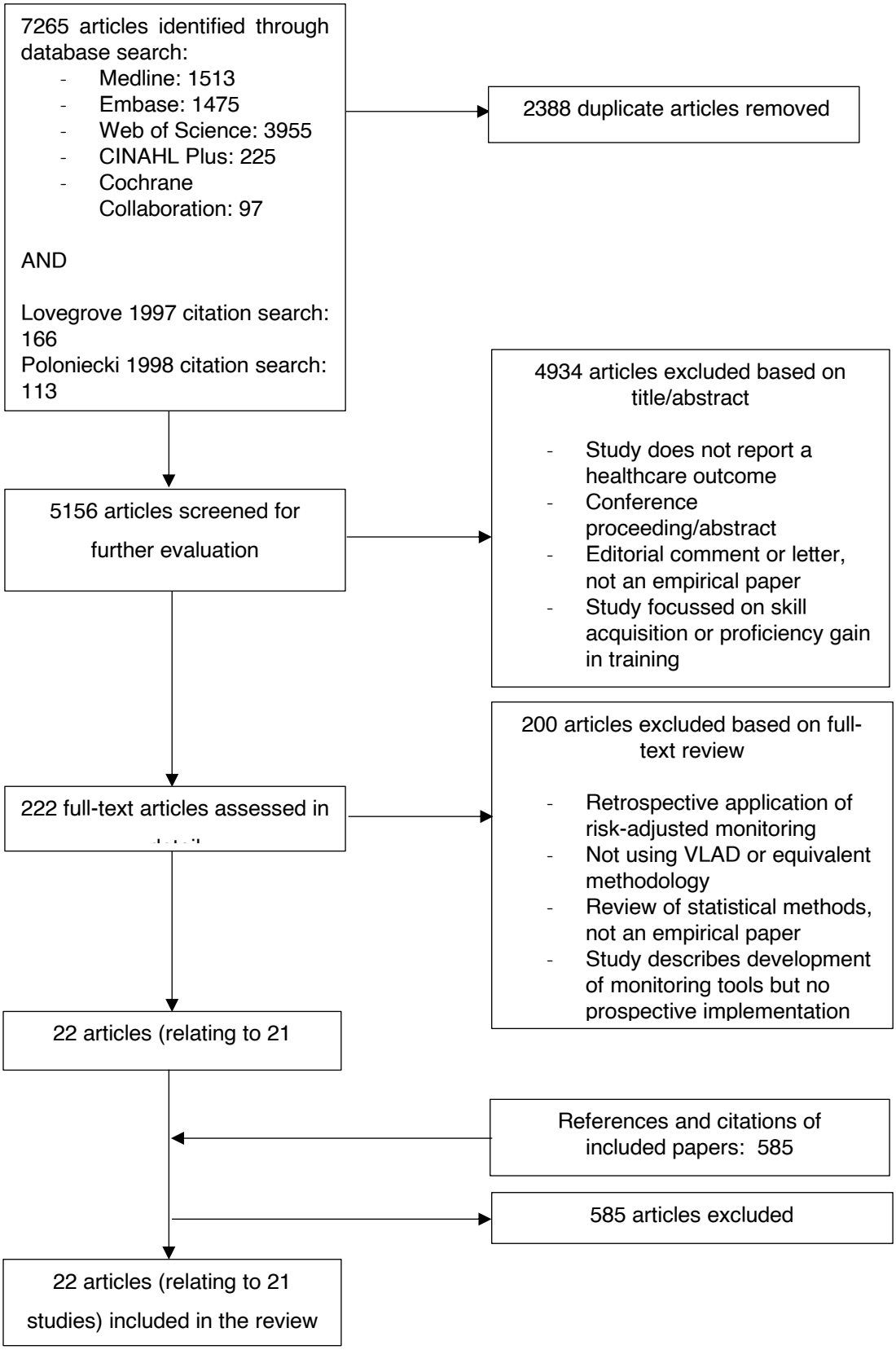


Figure 2-1: PRISMA flow diagram of systematic review

Table 2-1: Summary of studies included in review

Author and year	Specialty	Patient population	Outcome monitored	Data source	Methods used to disseminate data and embed VLADs in routine practice	Level of outcome reporting	Use of flags or alarm limits	QA or QI?	Published or risk-model derived for purpose of monitoring
Albert 2004[119]	Cardiac surgery	Adult	mortality	Clinical database	Online information portal, weekly meetings	Individual clinician level	No	QA	Published model
Albert 2004[120]	Cardiac surgery	Adult	mortality	Clinical database	Online information portal, weekly meetings/complication conferences, annual reports to family physicians and cardiologists	Individual clinician level	No	QA	Published model
Arrowsmith 2006[121]	Cardiac surgery	Adult	mortality	Clinical database	VLAD report produced annually or more frequently on an ad hoc basis to scrutinise certain aspects of practice.	Individual clinician level	Yes	QA	Published model
Belliveau 2012[122]	Cardiac surgery	Paediatric	Mortality, morbidity	Clinical database	Not stated	Departmental/hospital level	Yes	QA	Derived for monitoring
Borracci 2007[123]	Cardiac surgery	Adult	Mortality, morbidity	Clinical database	Not stated	Individual clinician level	Yes	QA	Published model
Brunelli 2011[124]	Thoracic surgery	Adult	Mortality, morbidity, other	Clinical database	Not stated	Departmental/hospital level	No	QA	Published model
Clarke 2010[125]	Cross specialty – hospital level indicators	Adult, paediatric	Mortality, morbidity, other	Administrative database	Not stated	Hospital level	Yes	QA	Derived for monitoring

Author and year	Specialty	Patient population	Outcome monitored	Data source	Methods used to disseminate data and embed VLADs in routine practice	Level of outcome reporting	Use of flags or alarm limits	QA or QI?	Published or risk-model derived for purpose of monitoring
Collett 2009[126]	Renal transplant	Adult	Mortality, morbidity	National transplant database	Three-monthly report, produced four months after last transplant included in dataset	Hospital level	Yes	QA	Derived for monitoring
Driessen 2016[127]	Gynaecology	Adult	Morbidity	Clinical database	Web-based dashboard	Individual clinician level	Yes	QA	Derived for monitoring
Duckett 2007[98]	Cross specialty – hospital level indicators	Adult, paediatric	Mortality, morbidity, other	Administrative database	Monthly dissemination in report format	Departmental/hospital level	Yes	QA	Derived for monitoring
Fusco 2012[128]	Cross specialty – hospital level indicators	Adult	Mortality, morbidity, other	Administrative database	Reports	Departmental/hospital level	Yes	QA	Derived for monitoring
Kuhan 2018[129]	Vascular surgery	Adult	Morbidity	Not specified	Results emailed to surgeons taking part every 3 months* * clarified via email with author	Individual clinician level	No	QA	Derived for monitoring
Lovegrove 1999[130]	Cardiac surgery	Adult	Mortality	Clinical database	Not stated	Individual clinician level	No	QA	Derived for monitoring
Lovegrove 1997[91]	Cardiac surgery	Adult	Mortality	Clinical database	Not stated	Individual clinician level	No	QA	Published model
Morton 2008[131]	Cross specialty	Adult	Morbidity	Clinical database	A deidentified hospital report was provided 6-monthly including state-wide aggregate and hospital level data.	Departmental/hospital level	Yes	QA	Published model

Author and year	Specialty	Patient population	Outcome monitored	Data source	Methods used to disseminate data and embed VLADs in routine practice	Level of outcome reporting	Use of flags or alarm limits	QA or QI?	Published or risk-model derived for purpose of monitoring
Pagel 2013[95]	Cardiac surgery	Paediatric	Mortality	National audit database	VLAD charts presented to multidisciplinary clinical teams within each centre. The first time VLAD charts were shown at each centre, analysts from University College London gave a brief introduction to the risk model and VLAD chart methodology. Charts were shown at participating sites approximately monthly. Prototype software was given to each centre at the end of the implementation period and is now being used independently.	Departmental/hospital level	No	QA/QI	Published/derived for monitoring
Patella 2016[132]	Thoracic surgery	Adult	Morbidity	Clinical database	VLAD charts presented to internal audit meetings	Departmental/hospital level	No	QA/QI	Derived for monitoring
Roberts 2013[133]	Burns	Adult	Mortality	Clinical database	Not stated	Departmental/hospital level	Yes	QA	Published model
Sketcher-Baker 2010[134]	Obstetrics, Gynaecology	Adult	Morbidity	Administrative database	Monthly disseminated VLAD monitoring tool report	Departmental/hospital level	Yes	QA	Not stated

Author and year	Specialty	Patient population	Outcome monitored	Data source	Methods used to disseminate data and embed VLADs in routine practice	Level of outcome reporting	Use of flags or alarm limits	QA or QI?	Published or risk-model derived for purpose of monitoring
Snyder 2014[135]	Transplant surgery	Adult, paediatric	Mortality, morbidity	National transplant database	Monthly updates to the monitoring charts for programs to review. Charts supplied to transplant programs via secure transplant registry website to be used.	Departmental/hospital level	Yes	QA/QI	Published model
Vasilakis 2011[136]	Orthopaedic surgery	Not specified	Morbidity	Clinical database	Access to software based on hospital servers. Previous use of non-continuous analyses in three-monthly reports	Individual clinician level	No	QA	Derived for monitoring
Williams 2015[137]	Orthopaedic surgery	Adult	Mortality	Clinical database	Not stated	Departmental/hospital level	No	QA	Published model

QA = quality assurance, QI = quality improvement, VLAD = variable life-adjusted display, O-E = observed minus expected, CUSUM = cumulative sum

Table 2-1: Summary of studies included in review (continued)

2.4.2 Quality assessment

The overall quality assessment of studies is included within Table 2-2. Details of the MMAT assessment are available in Appendix B-2. Twelve studies were single-centre, 10 were multi-centre. Data collection was prospective in all studies. Four out of the 22 studies reported baseline patient characteristics, comorbidity, and demographics;[123,124,127,132] selection bias was evident in four studies.[122,127,132,136] Potential for bias included evidence of either patient selection in the monitoring process[122,132,136] and selection or opt in of clinicians being monitored within a study.[127]

Nine studies provided satisfactory detail on the implementation of the outcomes monitoring and methods used to embed VLADs into routine clinical practice. Of these, only four studies presented new qualitative data about the implementation.[95,125,127,134] Driessen et al. used a structured qualitative evaluation involving surveys of clinicians using their monitoring tools.[127] Despite this study providing good detail on the implementation of monitoring and scoring highly from a quantitative aspect, the overall rating (*) is a reflection of the limited qualitative methodology. The remaining three studies presenting new qualitative data failed to provide information about how the data were gathered or analysed and received an overall rating of * or ** using the MMAT tool.[95,125,134] No study undertook a rigorous qualitative assessment of monitoring implementation. The 18 studies that did not present new qualitative data drew from previously published literature to enable discussion of the perceived issues in implementation.

Studies generally set out to describe the development and rationale for implementing continuous monitoring but failed to evaluate that implementation. Six studies provided information about large scale regional or national implementation of monitoring schemes.[98,125,128,131,134,135] However, only two of these provided satisfactory detail about how the monitoring tools were embedded into routine clinical practice and neither evaluated its impact.[98,135] There was a general lack of detail provided by studies on how monitoring was perceived by clinical teams. No

studies attempted to define the impact of monitoring either from a quantitative or qualitative perspective.

Overall 14 out of 18 quantitative descriptive studies received an MMAT summary score of ** or *** (out of a possible ***). Of the qualitative or mixed methods studies, which could have received a maximum summary score of ****, two received *[125,127] and two received **.[95,120] This demonstrates the poor qualitative methodology within included studies.

Table 2-2: Quality assessment of studies included in review

Author and year	Country and setting of study	Number of centres involved	Patients (n)	Description of patient demographics	Risk of selection bias	Detail provided about monitoring intervention	Detail about methods to embed VLADs	Original qualitative data presented in paper	MMAT assessment of quality
Albert 2004[119]	Germany, tertiary care	Single	14487	No	No	Satisfactory	Satisfactory	No	**
Albert 2004[120]	Germany, tertiary care	Single	13000	No	No	Satisfactory	Satisfactory	No	**
Arrowsmith 2006[121]	UK (England), tertiary care	Single	Not stated	No	No	Satisfactory	Satisfactory	No	*
Belliveau 2012[122]	Canada, tertiary care	Single	178	No	Yes	Satisfactory	Limited	No	**
Borracci 2007[123]	Argentina, tertiary care	Multicentre	502	Yes	No	Limited	Limited	No	**
Brunelli 2011[124]	Italy, tertiary care	Single	511	Yes	No	Limited	Limited	No	***
Clarke 2010[125]	Australia, state-wide	Multicentre	Not stated	No	No	Satisfactory	Limited	Yes	*
Collett 2009[126]	UK, national	Multicentre	2218	No	No	Satisfactory	Satisfactory	No	***
Driessen 2016[127]	Netherlands, international	Multicentre	2066	Yes	Yes	Satisfactory	Satisfactory	Yes	*
Duckett 2007[98]	Australia, state-wide	Multicentre	Not stated	No	No	Satisfactory	Satisfactory	No	***
Fusco 2012[128]	Italy, state-wide	Multicentre	Not stated	No	No	Satisfactory	Limited	No	***
Kuhan 2018[129]	UK, tertiary care	Single	941	No	No	Limited	Limited	No	**
Lovegrove 1999[130]	UK, tertiary care	Single	4318	No	No	Limited	Limited	No	*
Lovegrove 1997[91]	UK, tertiary care	Single	5000	No	No	Limited	Limited	No	*

Author and year	Country and setting of study	Number of centres involved	Patients (n)	Description of patient demographics	Risk of selection bias	Detail provided about monitoring intervention	Detail about methods to embed VLADs	Original qualitative data presented in paper	MMAT assessment of quality
Morton 2008[131]	Australia, state-wide	Multicentre	49804	No	No	Satisfactory	Limited	No	**
Pagel 2013[95]	UK, tertiary care	Multicentre	2649	No	No	Satisfactory	Satisfactory	Yes	**
Patella 2016[132]	UK, tertiary care	Single	348 + 50*	Yes	Yes	Limited	Limited	No	*
Roberts 2013[133]	UK, tertiary care	Single	439 + 78*	No	No	Satisfactory	Satisfactory	No	**
Sketcher-Baker 2010[134]	Australia, state-wide	Multicentre	Not stated	No	No	Limited	Limited	Yes	**
Snyder 2014[135]	USA, national	Multicentre	Not stated	No	No	Satisfactory	Satisfactory	No	**
Vasilakis 2011[136]	UK, tertiary care	Single	4820	No	Yes	Satisfactory	Limited	No	**
Williams 2015[137]	UK, tertiary care	Single	277	No	No	Limited	Limited	No	**

Table 2-2: Quality assessment of studies included in review (continued)

*Where two figures are given, first refers to the number of patients included in the first phase (risk-modelling or retrospective analysis) and the second to the number of patients involved in the prospective monitoring phase of the study. MMAT – Mixed Methods Appraisal Tool[113] (see Appendix B-2)

2.4.3 Outcome monitored and rationale for monitoring

Eight studies used mortality as a single outcome measure, six used morbidity, and eight reported more than one outcome, using a combination of mortality, morbidity and other outcomes. The timepoint at which mortality was recorded varied between in-hospital, 28-day and 30-day. Definitions of morbidity varied between studies, but all were in-hospital outcomes.

Two studies noted that a limitation of using mortality as an outcome measure is that any deficiency in care that does not result in death will not be identified and that mortality may not be an adequately sensitive measure of poor performance.[127,137] One study suggested the use of an alternative outcome, choosing multiple outcomes, or using a combined outcome measure may be preferable.[124] Several studies discussed that in order to provide a more comprehensive analysis of performance, outcome indicators (covering postoperative domains) and process indicators (covering preoperative and intraoperative domains) should be used together.[132,138–140]

VLADs were most commonly used in cardiac surgery, which appeared as a single specialty in eight out of 22 studies. Four studies used VLADs to monitor outcomes across multiple medical and surgical specialties (> two). Three of these studies reported implementation in Australia, where state-wide monitoring has been initiated in Queensland and Victoria.[98,125,131] One study reported region-wide monitoring in Lazio, Italy.[128]

Nineteen studies implemented monitoring from a quality assurance perspective, three studies had a quality assurance and quality improvement aim.[95,132,135] No studies stated the sole aim of implementation was to support quality improvement.

Twelve studies used alarm limits to identify variation in outcomes beyond prespecified limits. Methods used to calculate and display these limits varied between studies and included:

1. Statistical testing of the difference between expected and observed outcomes[128]
2. A prespecified change in the expected minus observed value over a set number of cases[122]
3. Predefined limits calculated using baseline data – the change in either absolute or relative risk[126,127,134]
4. Use of funnel plots and Bayesian shrinkage techniques calculated in the background with alarms mapped onto the VLAD[131]
5. A prespecified difference between expected and observed outcome, either empirically determined (for example two excess deaths compared to expected) or based on baseline data[121,133]
6. Use of Cumulative sum chart alongside VLAD with limits and/or alarms mapped onto the VLAD[98,135]

2.4.4 Choice of risk-adjustment model for use in monitoring process

Half of the studies (11/22) used a risk-model derived for the purpose of VLAD monitoring. The level of detail given about the risk modelling process varied between studies. Some provided a brief overview of model development, giving information about the variables included and their coefficients but did not follow the TRIPOD guidelines.[141] Other studies gave no information about the risk-adjustment process. Pagel et al.[95] published their risk-modelling work separately.[142] Of the 10 studies that used an existing risk model, little, if any data was provided to demonstrate its applicability for use in the study population. Only one of the 10 studies using an existing risk model presented demographic data on their study population.

Three studies (four articles) used the EUROscore risk model as the method of risk-adjustment when monitoring mortality after cardiac surgery.[119–121,123] Authors of the studies discussed that VLAD curves produced during monitoring are highly dependent upon the risk model used and that where a risk model underestimates the risk on an outcome the curve will be falsely reassuring. The EuroSCORE has been shown to overestimate mortality in lower risk groups and underestimate

mortality in higher risk groups.[143] Three studies acknowledged the importance of having a well-calibrated risk model with good discrimination for VLAD monitoring, but even with this adjustment is never perfect.[95,128,135] Other studies commented on the need to understand caveats to any risk model and that regular updates should be performed to ensure performance remains satisfactory.[120,123,129,133] One study stated that recalibration of published risk models to local settings may improve the quality of monitoring[120] although another argued that models produced from a large pool of national data may actually improve performance and support comparative audit between centres.[129]

2.4.5 Method of VLAD dissemination and strategies to embed into routine clinical practice

Eleven studies used static reports to disseminate results, produced with a frequency ranging from monthly to annually. Four studies developed a web-based or hospital network interface, allowing users to interact with their local data and enabling access to up to date information at any time.[119,120,127,136] One study made analysis software available to participating sites to allow them to produce their own VLAD monitoring charts.[95]

Departmental meetings, such as audit or morbidity and mortality meetings were a common method used to embed VLADs in routine clinical practice.[95,119,120,132] The frequency of meetings was either weekly or monthly. Eight studies made no comment on how results were disseminated or how VLADs were embedded locally.

2.4.6 Facilitators and barriers to implementation of prospective monitoring using VLADs

A range of facilitators and barriers were discussed in studies. The four main themes identified in the analysis of qualitative data are discussed in detail below.

2.4.6.1 Engaging clinical teams

Building engagement

Two studies commented that the use of clinical databases, where clinicians and other members of the hospital team are responsible for data collection, can promote engagement[119,129], whilst one study using an administrative dataset stated that a lack of involvement from clinicians may have been due to poor confidence in the data source.[125] Fourteen studies used a clinical database as the source of data, three used a national audit database and four studies used administrative data. One did not specify the data source.[129]

Another method used to promote clinician engagement with the monitoring process included involvement of clinicians from the initial stage of selecting an outcome measure to monitor.[95,125,127,134] Where multicentre monitoring was introduced, working groups including a range of key stakeholders were used to engage clinicians in deciding the indicators to be monitored and how to define them.[98,125,128,134] Two studies reported that gaining clinical perspective on indicators can help to identify problems with a definition or data collection process.[125,128] One article also reported that a close working relationship with clinicians leads to improved and more clinically relevant indicator measures, and involving local teams in working groups can improve participants' knowledge of VLAD methodology and indicator definitions.[134] They suggested that this involvement may facilitate the creation of a pool of experts in the medical community who act as champions for the monitoring process.[134] Educational sessions to improve clinicians' and data collectors' understanding of indicator definitions and VLAD methodology were used by two studies.[95,134] No information was included about the structure or delivery of the educational programme however.

Maintaining engagement

One study suggested that a formal review process aimed at addressing problems with definitions or clinical coding identified by hospital staff may help to maintain engagement.[134] Another study involved clinicians directly in the investigation of any apparent excellent or poor performance in order to develop local leadership skills.[125] The importance of unit culture viewing audit positively and that investigations do not seek to apportion blame was stated by one study.[133] Three studies stated that fear of censure and reprisals can act as a barrier to clinician engagement and the use of indicators developed for formative purposes as performance measures can further reduce engagement.[95,125,135]

One study sought to engage the multidisciplinary team in the validation of data (for example stroke physicians in the monitoring of carotid endarterectomy outcomes) as a method to increase the resilience and reliability of the monitoring process.[129]

Whilst VLADs are traditionally presented by case number[91], one study suggested that presenting them by date may support clinicians in identifying contextual factors more easily, such as the implementation of new care pathways, changes in staffing or seasonality of outcomes.[123] This in turn may help to maintain engagement of clinicians with the monitoring process.

2.4.6.2 The burden placed on clinical teams by the monitoring process

Burden of data collection

Three studies used trained data collectors to reduce variation in recording of outcomes[124,131,132] and one suggested clinician engagement may be reduced when monitoring places a significant strain on resources.[125] Spreading the workload of data collection between individuals may reduce this burden. Individuals responsible for data collection varied between studies and included junior doctors,[129] surgeons,[119,120,129,132] infection control practitioners[131] or other trained clinical staff[124]. Where monitoring was implemented on a large-scale basis such as that in Queensland, Australia, administrative datasets were used.[125,131,134] Suggested advantages of administrative datasets included the relatively low cost of data collection and the ability to capture information on a large population. One study

also commented that the use of administrative datasets avoids the need for voluntary participation of hospitals or clinicians.[128]

Burden of investigation of VLAD results

Four studies that used flag or alarm limits stated that the number of outcomes being monitored and the level at which flags are set contributes to the burden on resources.[125,133–135] One study suggested that monitoring programmes need to balance the cost of investigating alarms more frequently (where a change in outcome may only be due to statistical artifact) against the need to identify special cause variation[98], whilst another stated that setting trigger points too low can cause a significant burden on resources.[133]

2.4.6.3 Integration of technology within monitoring process

One study stated that a lack of automation and the ad-hoc production of reports in monitoring process was an obstacle to implementing continuous monitoring.[136] Another study stated bi-annual or annual analysis introduces a lag of months to years from data collection to feedback of results, leading to complaints that data presented are old.[135]

Web-based dashboards were used or discussed in five studies.[119,120,127,135,136] Suggested advantages of these dashboards were that they are easily accessible and allowed immediate feedback of results. One study stated that rapid and automated analysis promotes earlier identification of unfavourable trends.[119] Despite this potential for rapid data feedback some studies used static monthly or annual reports.[95,98,121,126,128,129,131,134,135]

Four studies stated that a significant advantage of VLAD methodology is that it can provide immediate feedback to clinical teams about their outcomes.[95,127–129] This contrasts with analyses performed every 6 or 12 months which one study stated can mask runs of adverse events if outcomes are otherwise satisfactory.[131] By automating data analyses and report production into the monitoring process the potential of VLAD methodology to provide immediate feedback can be realised. Automation also supports more frequent production of reports compared to manual

production. This is particularly relevant where large scale monitoring programmes are implemented.[98,125]

One study whose implementation of VLAD monitoring required software installation on hospital networks suggested data security issues related to this installation may be a barrier to implementation[136] but modern web-based tools used by three studies overcame this difficulty.[119,120,127] The use of an interactive online dashboard with immediate feedback of results was suggested by one study to facilitate engagement, registration behavior, and intrinsic motivation to improve performance.[127] Another study stated that software should be responsive to users' needs, for example allowing the addition of case-mix details or national averages.[95]

Two studies stated that using a variety of methods to show results may improve understanding and using online forms to gather feedback on any problems with the monitoring process can support communication between sites and the central team.[128,134]

Three studies discussed the benefits of building intrinsic error checking and data verification technology into software systems. They suggested these checks can help improve the accuracy of data and are critical to ensure accurate risk-adjustment.[95,119,128] Building software solutions that integrate data from multiple sources can reduce the data collection burden whilst improving data quality; allowing comparison of outcomes from a range of sources.[128]

2.4.6.4 VLADs as a formative monitoring tool

In the first paper to described VLAD, Lovegrove et al. stated it should be used with common sense and their results should not be overinterpreted.[91] Several studies describe their role as a basis for learning and reflection.[91,95,98,127,132,133,135,137] Fives studies stated that VLADs do not provide information about causes of a trend in outcomes and should not be used to draw firm conclusions about quality of care.[98,132,134–136] In ten studies, concerning trends in VLADs led to further scrutiny of data and processes of care through detailed multidisciplinary

audit.[91,95,98,127,128,132–135,137] Two studies stated that investigations should not seek to apportion blame however.[95,133]

Open publication of VLAD charts was only used in one study.[128] Results were published to create an incentive to improve results, however concern regarding confidentiality is a barrier to clinician and hospital engagement. One study stated that open publication of results may mean hospitals see a programme of routine monitoring as a risk rather than an opportunity.[95] Another stated that care should be taken when comparing institution performance, especially where outcome definitions are not robust.[136]

The context and intended use of the monitoring process should be discussed with teams.[95] Although outcomes at an individual clinician level were often reported,[91,119–121,123,127,129,130,136] the use of programme based (hospital or departmental level) outcomes may be more appropriate to reflect the multidisciplinary nature of modern care.[95,123] Reassuring clinicians that monitoring charts will not be shared with third parties may support engagement.[135]

Twelve out of the 22 studies used alarm limits to identify where trends in outcomes deviated beyond prespecified limits. Reasons for not including flag or alarm limits included: to not imply what is or is not acceptable performance[91]; to limit the scope for complacency[95]; and because the tool was intended to trigger further scrutiny of performance and not draw firm conclusions.[136] VLADs were generally viewed as a descriptive technique rather than statistical hypothesis testing.[130] They are helpful in providing a general picture of performance, but not to indicate when a centre should be concerned and take action.[135]

2.5 Discussion

There was a relative paucity of literature describing the prospective implementation of monitoring using VLADs. Despite state-wide implementation at a governmental level in Queensland, Australia, and a smaller roll out in Victoria, Australia there were few studies that described in detail the implementation, its impact on outcomes, or qualitative data about the monitoring process in these regions.

VLADs were introduced in Queensland in response to the Bundaberg Hospital crisis in 2005, which centred around the poor clinical performance of Dr. Jayant Patel, an Indian-born American surgeon who served as the Director of Surgery at Bundaberg Base Hospital in Queensland, Australia, from 2003 to 2005. His tenure was marked by a significant increase in patient mortality and morbidity and a clinical review subsequently found that Dr. Patel directly contributed to the deaths of at least eight patients and exhibited an unacceptable level of care in numerous other cases.[144,145] The crisis highlighted critical systemic failures within Queensland Health, including deficiencies in the credentialing and oversight of overseas-trained medical practitioners, a lack of robust internal complaints management and incident reporting systems, and a culture that tolerated clinical issues rather than addressing them proactively.[144–146]

The over-riding rationale for monitoring implementation in Queensland in response to these issues was to provide quality assurance, which is consistent with the reason for monitoring described in other studies included in this review. Only three studies appeared to apply a QI agenda alongside QA.

Table 2-3 summarises some of the key findings of the review.

Key findings
• VLADs were most frequently used to monitor mortality after cardiac surgery – but were also used to monitor outcomes (including mortality, morbidity or complications, length of stay, readmission) in other surgical and medical specialties • VLAD monitoring was generally implemented from a QA perspective but details of how they were embedded in routine clinical practice were limited • There was limited qualitative evaluation of the VLAD monitoring programmes and their impact on outcomes • Four main themes were identified regarding facilitators and barriers to monitoring implementation: - Engaging clinical teams - The burden placed on clinical teams by the monitoring process - Integration of technology within monitoring process - Use of VLADs as a formative monitoring tool

Table 2-3: Key findings of this review

Information provided by studies on the technology used to produce, disseminate and embed VLADs into clinical practice was very limited. Four studies provided information on the software solutions implemented, but the level of detail was not adequate for researchers to reproduce their methods.[95,120,127,136] The most comprehensive description was provided by Driessen et al.[127] Despite additional details being released about the software used in one study,[133,147] overall there was no clear consensus about what features the software should have and how it can be optimized to facilitate the monitoring process.

The use of computer or web-based programmes to collect data and automate the analysis was viewed as having a positive effect in a number of ways: increasing data quality through validation checks at the time of data entry;[95,127] allowing immediate, real-time feedback of results;[119,127,133] allowing users to interact with their results;[120,127,136] and allowing users to feedback any problems they have identified with the monitoring process.[125] Harnessing technology to reduce the burden of data collection, analysis and reduce time to the feedback of results was viewed as a method to promote engagement.[120,127] The use of web-based

dashboards enables clinicians and other stakeholders to interrogate their data at timepoints relevant to local context. This contrasts with traditional approaches of annual or bi-annual reporting where timing is determined by external individuals where reports are produced manually. By incorporating interactive elements clinicians can explore their local data more easily without central bodies having to undertake complex analyses, thus supporting engagement with local data.[127,136]

Only one study reported new qualitative data that was collected to formally evaluate the implementation of VLAD monitoring.[127] This study gathered feedback from users of their web-based portal in survey format, and the interactive clinician accessible dashboard with ability to monitor one's own performance was generally positively received.

Studies frequently relied on information from other published work to describe barriers and facilitators rather than producing new evaluative work. Formal mixed methods evaluations of monitoring programmes in both a QA and QI setting could help to quantify the impact of continuous monitoring on outcomes and improve understanding of contextual factors that support or inhibit their uptake.[148] Further research is needed to establish whether facilitators and barriers to implementation for QI differ to those of QA, and how adoption can be increased in either circumstance.

Limitations of this review

There are several limitations to this study that are important to highlight.

First, the focus was on studies that described prospective implementation of VLADs. A significant number of studies were excluded because they applied VLADs retrospectively to describe results or demonstrate a previous change (Figure 2-1: PRISMA flow diagram of systematic review). Whilst excluded studies would not have added knowledge related to prospective implementation, they may have included information on existing applications of VLADs in healthcare settings not captured in prospective studies. Examples include settings where VLADs have

been used retrospectively to assess the impact of pathway or service reconfiguration. An illustration of VLADs being used retrospectively is a study by Groven et al. who used them to demonstrate a change in patient survival after centralisation of trauma services in Norway.[149]

Studies using exponentially weighted moving average charts (EWMA), RA-CUSUM charts reporting log-likelihood ratios and other monitoring tools not closely related to VLADs were also excluded. Whilst studies using alternative methods to VLADs may have provided additional information about potential barriers and facilitators to prospective monitoring, such as whether involving clinicians in the design of the monitoring programme promotes engagement, other findings would have been specific to the method used. An example of this is statistical considerations around the use of confidence intervals where methods used to define them vary according to the chosen tool for monitoring. The decision to exclude methods not closely related to VLADs was aimed at ensuring barriers and facilitators were directly relevant to the prospective implementation of monitoring using VLAD methodology.

Second, the lack of formal qualitative evaluation within included studies meant that many of the barriers and facilitators to implementation discussed in the included studies were drawn from other published work. The relative lack of qualitative data about the implementation of monitoring schemes is a gap in the literature that needs to be addressed.

Third, studies may be preferentially submitted and accepted for publication where the implementation of monitoring was viewed as beneficial, leading to publication bias.

Finally, despite a literature search of multiple databases and citations, it is possible that some eligible studies were missed. Despite strategies to prevent this such as the citation and reference search of the sentinel VLAD publication[91] and all included studies, it is possible a small number of appropriate articles may have been omitted.

2.6 Grey literature search – Implementation of VLADs in Queensland, Australia

Peer-reviewed articles included in the systematic review identified the state-wide implementation of VLADs to monitor a variety of healthcare outcomes in Queensland, Australia and smaller scale adoption in Victoria and Western Australia. To capture additional information related to the implementation in Queensland a targeted grey literature search of Australian governmental websites was performed. The availability of information related to the Victorian and Western Australia implementations was limited compared to that in Queensland and so they are not discussed here.

The search identified 10 organisational reports which detailed the implementation and ongoing use of VLADs in Queensland. One report was excluded as it had been superseded by a more recent version, leaving nine that are discussed here.

2.6.1 Background and summary of the Queensland monitoring scheme

The Australian state of Queensland introduced widespread mandatory VLAD monitoring in 2007 following the Bundaberg Hospital Crisis.[150] The process was managed centrally with hospitals required to respond in a variety of ways depending on their local outcomes.

VLADs were introduced as a screening tool designed to identify potential areas of concern, or strengths to improve patient safety and quality of care. The implementation was designed to improve understanding of causation and to determine whether corrective action is necessary.[151]

A flagging process was included in the VLAD programme, which creates an alert or signal when a pre-specified level of variation is reached. A tiered approach based on three flagging levels, dependent upon the indicator being monitored was used (see: 2.6.2). A flag or alert suggests to clinicians and hospital teams that over time there had been more (or fewer) patients experiencing the outcome than expected.

Whilst some peer reviewed publications were produced from this process, the depth of information provided in these studies was limited.[98,131,134] Table 2-4 summarises the reports we retrieved on the VLAD monitoring programme produced by Queensland Governmental agencies.

Table 2-4: Summary of organisational reports produced by Queensland Government

Report title	Year	Specialty (Operation)	Outcome monitored	Definition of outcome
Report on the VLAD Laparoscopic Cholecystectomy Indicator Review: Summary of Activities[152]	2012	Gastrointestinal surgery (Cholecystectomy)	Readmission, long stays (previously monitored 'complications of surgery'	Readmission to hospital within 30-days of discharge with an emergency admission and for a condition that could be considered a consequence of the initial treatment received. Long stays: LOS >15 days with a principle diagnosis of heart failure
Report on the VLAD Heart Failure Indicator Review: Final Summary of Activities[153]	2012	Cardiology/General Medicine	Readmission, long stays (Previously monitored mortality up to 2011)	Readmission to hospital within 30-days of discharge Long stays – LOS >2 days for elective and LOS >6 days for emergency cholecystectomies
Paediatric Tonsillectomy and Adenoidectomy VLAD Indicator Working Group: Summary of Activities to-date: April 2010 - September 2010[154]	2011	Paediatric ENT surgery	Readmission, long stay,	Readmission: within 15 days with a principle diagnosis code considered to be related to original procedure Long stays: LOS >= 2 days
Guideline for Variable Life Adjusted Display and other National Patient Safety Indicators[151]	2017	Not stated – general report produced by Queensland Government: Health Service Directive	NA	
Report on the Stroke VLAD Indicator Review: Summary of activities[155]	2012	Stroke medicine	Inpatient mortality	All inpatient deaths considered, where patient had a LOS <=30 days. Excludes same day discharges/admissions <24 hours who did not die.

Report title	Year	Specialty (Operation)	Outcome monitored	Definition of outcome
Report on the Orthopaedic VLAD Indicator Review: Summary of activity[156]	2012	Orthopaedic surgery (Fractured neck of femur, Hip replacement, Knee replacement)	Mortality, morbidity, readmission, long stay	Mortality: in-hospital deaths Morbidity: Predefined list of clinical codes related to admission that were considered to be complications. Defined by working group. Readmission: patients readmitted to any Queensland Health hospital within a specified interval since discharge. Interval depends on both operation performed and complication readmission diagnoses codes. Long stay: hip replacement: LOS >=23 days. Knee replacement: LOS >= 17 days. Both selected as 90th centile for LOS data.
Report on the Obstetric VLAD Indicator Review: Summary of activity[157]	2014	Obstetric	Selected primiparae instrumental delivery, selected primiparae caesarean section, selected primiparae episiotomy and 3rd/4th degree perineal tears, selected primiparae induction of labour	Monitoring of number of given interventions/outcomes in selected population
Report on Acute Myocardial Infarction VLAD Indicator Review: Summary of activity[158]	2012	Cardiology	Mortality, readmission, long stay	Mortality: In-hospital deaths with a LOS <=30 days Readmission: unplanned/unexpected readmission with prespecified diagnosis code, readmission time frame either 7 or 30 days depending on diagnosis code Long stay: LOS >= 12 days (90th centile of LOS data)

LOS = length of stay,

2.6.2 Methods used to calculate flag limits

Across the Queensland implementation, flagging limits included both an upper and lower level. The upper flag indicates a hospital's outcome rate is lower than the state outcome rate (possible good performance). A lower level flag indicates the hospital rate is higher than the state outcome rate (possible poor performance). Indicators are assigned to groups A to D depending on the acceptable variation in outcomes. Lower variation is tolerated in group A compared to B and so on. Table 2-5 summarises the outcomes monitored, the groups assigned and the level of variation that results in flags.

2.6.3 Dissemination of VLADs

In the Queensland programme, VLADs are generated on a monthly basis using administrative data and published using a web-based clinical monitoring system. A monthly notification report listing any new flags is sent via email from the monitoring system to the following individuals: Hospital and Health Service Chief Executive; VLAD Authorising Officer; VLAD Hospital Coordinator; and other authorised users of the clinical monitoring system

Table 2-5: Indicator groups, flag levels and outcomes monitored in Queensland, Australia[151]

Flagging level	Variation from state average to trigger flag	Indicator and outcome
Indicator group A	Level 1: 10%, Level 2: 20%, Level 3: 30%	Selected primiparae assisted births, episiotomy or 3rd and 4th degree perineal tears (public facilities)
Indicator group B	Level 1: 30%, Level 2: 50%, Level 3: 75%	Acute myocardial infarction readmission Fractured neck of femur in-hospital mortality Fractured neck of femur complications of surgery Pneumonia in-hospital mortality Stroke in-hospital mortality Selected primiparae caesarean section (private mothers in public facilities) Selected primiparae caesarean section (public mothers in public facilities) Selected primiparae induction of labour (public facilities) Selected primiparae instrumental delivery (public facilities)
Indicator group C	Level 1: 50%, Level 2: 75%, Level 3: 100%	Abdominal Hysterectomy complications of surgery Acute Myocardial Infarction long stay and in-hospital mortality Colorectal carcinoma complications of Surgery Depression long stay and readmission Heart Failure long stay and readmission Hip Replacement (primary) complications of surgery Hip Replacement long stay and readmission Knee Replacement (primary) complications of surgery Knee Replacement long stay and readmission Laparoscopic Cholecystectomy long stay and readmissions Prostatectomy complications of surgery Schizophrenia long stay and readmission Selected primiparae (unassisted births) Episiotomy or 3rd and 4th degree Vaginal Hysterectomy complications of surgery
Indicator group D	Level 1: 100%, Level 2: 125%, Level 3: 150%	Paediatric Tonsillectomy and Adenoidectomy long stay and readmission

2.6.4 Response to flags during VLAD monitoring

The Queensland reports state that a flag should not immediately be interpreted as indicating good or bad performance, because a range of explanations as to why the VLAD has flagged are possible, one being chance. In response to a VLAD notification report the hospital is required to conduct a review and submit a report that considers details of cases reviewed, any issues identified and an action plan to correct an unfavourable result or to maintain a positive result.

For lower level two and lower level 3 flags, hospitals are mandated to conduct a review and submit a response within 30 days. It is recommended that lower level 1 flags (possible poor performance at level one) and upper level flags (flags due to positive performance) do not require review, but they are included in the notification report. These flags may provide an opportunity to learn from improvement.

All responses to level three flags are reviewed by a VLAD committee that meets monthly. This committee, which sits within the Queensland Department of Health, clinically reviews the adequacy and action plans or responses. If concerns are raised then further information may be requested, following which there is an escalating response from the VLAD committee. Issues unable to be resolved by the committee are escalated within the Department of Health. The VLAD committee feed information back to the VLAD system team about indicators, their definitions or any problems with the monitoring process.

Where a review is initiated, the Queensland Health Service Directive recommend a pyramid model of investigation (see:

Figure 2-2). The pyramid model of investigation is a hierarchical approach to identify causation.[104] Under this model, factors at the base of the pyramid are more likely to be causes than factors at the apex.

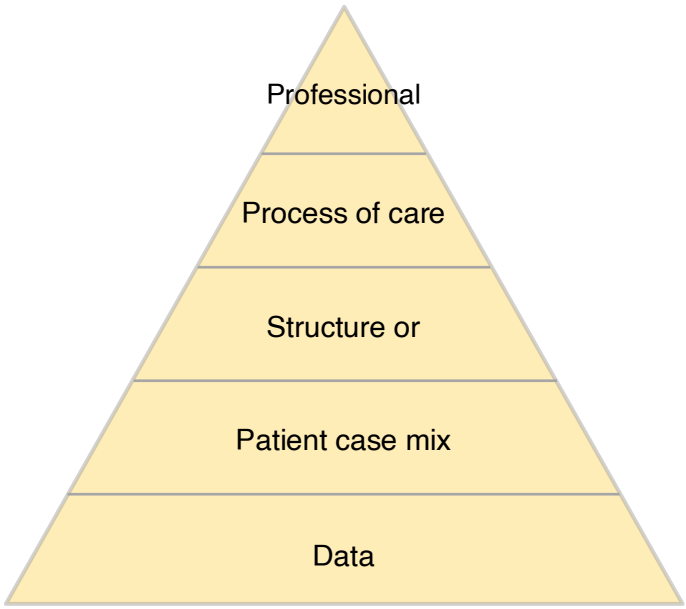


Figure 2-2: Pyramid model for investigating outcomes[104]

This model is used to support clinical teams in investigating apparent variation shown in the VLADs. The structured approach seeks to identify common causes of variation and avoid assigning blame to individuals. The model was first developed by Mohammed et al. linked to an investigation following on from the conviction in 2000 of GP, Harold Shipman, of mass murder in the UK.[104]

2.7 Conclusions

VLADs have been implemented to prospectively monitor outcomes in a wide variety of healthcare settings. Whilst predominantly used to monitor mortality after surgery, they have been used to monitor morbidity or other binary outcomes in both medical and surgical specialties. The largest implementation of VLADs to date was in Queensland, Australia where they have been implemented on a state-wide basis in response to a well publicised hospital scandal. Grey literature relating to the Queensland implementation (Section 2.6) offers useful information on what

outcomes were monitored, how VLADs were implemented and what the expected response to them was. However, despite the state-wide implementation there are very limited peer-reviewed studies detailing the experience and learning that has occurred from it.

VLADs were most commonly implemented from a QA perspective, although some studies included a QI aim. Continuous outcome reporting incorporating risk-adjustment has been suggested as a method to improve the use of national audit data for quality improvement.[9] VLADs offer a potential method to deliver this, with or without the use of alarm limits. Their use as a QI tool to date is limited, however.

The impact of VLAD monitoring programmes have not been formally assessed and many included studies report no new quantitative or qualitative data about their effect on patient level outcomes. Whilst there has been separate quantitative analysis of the impact of some of the studies included in this review, these evaluations did not assess the individual role of VLADs within the monitoring programme.[159–162] Future research using VLADs or other continuous monitoring techniques should include qualitative evaluation to improve understanding of how monitoring programmes change outcomes and influence safety and improvement culture. Techniques to implement continuous monitoring that maximise positive effects such as early identification of problems and learning from good performance, whilst avoiding negative associations of censorship and blame, need to be evaluated.

Chapter 3 Development and temporal validation of a multivariable risk-adjustment model for morbidity on postoperative day 7 in adult patients undergoing major surgery

3.1 Abstract

Aims

To develop and internally validate, then temporally validate a risk-adjustment model suitable for use in reporting risk-adjusted postoperative day 7 morbidity outcomes in near real time across a range of surgical specialties recruiting to the Perioperative Quality Improvement Programme.

Methods

Data from a prospectively collected multi-centre clinical dataset from 63 NHS hospitals in England were used to develop the model. Postoperative morbidity was defined using the validated and reliable Postoperative Morbidity Survey at postoperative day 7.[65,66,76] Twenty-eight variables were included in the initial 'full model'. Backwards step-wise selection was used to fit logistic regression models in 1000 bootstrap re-samples. Variables selected into >80% of bootstrap models were included in the 'final' model which was fitted using penalised maximum likelihood estimation.

Internal validation in the derivation cohort was performed through Harrell's bootstrap correction of optimism.[163] A temporal validation of the final model was also performed using a dataset that was not available at the time of initial model development. This cohort included data from 89 NHS hospitals. Model discrimination in the validation cohort was assessed using the C-statistic, calibration was assessed visually and with the Hosmer-Lemeshow test,[164] and overall model accuracy was assessed with the Brier score.[165]

Results

A total of 4466 cases contributed to model derivation. Ten variables were selected into the final model: surgical specialty, surgery severity category, gender, ASA grade, body mass index, preoperative heart rate, systolic blood pressure, age (years), number of operations in last 30 days, and respiratory history findings.

The optimism corrected C-statistic in the derivation cohort was 0.68 (95% CI 0.66-0.69). The Hosmer-Lemeshow test showed no evidence of a lack of fit (p=0.41).

There were 8251 cases included in the temporal validation cohort. The C-statistic of the PQIP model was 0.68 (95% CI 0.66-0.69) in the validation cohort, and the Brier score was 0.135. The Hosmer-Lemeshow test suggested a possible lack of fit ($p<0.05$) in the temporal validation dataset. Further analysis suggested this lack of fit may be due to overestimation of risk in patients undergoing urological procedures and poor calibration in the head and neck category. Calibration varied significantly between surgical specialties. Model performance was favourable compared to published morbidity risk models.[67,166]

Conclusion

A risk-adjustment model for day 7 postoperative morbidity in patients undergoing major surgery in the UK was developed and temporally validated. Discrimination was favourable compared to published risk models and calibration was acceptable.

3.2 Introduction

The Perioperative Quality Improvement Programme (PQIP) is a national study running in over 100 NHS hospitals across the UK. One of its primary aims is to increase the collection and use of perioperative data to support local level quality improvement.[167,168] In order to implement monitoring of morbidity outcomes using a variable life-adjusted display (VLAD)[91] it is necessary to have a well calibrated risk-prediction model with adequate discrimination. The aim was to develop a model for implementation in a near real-time dashboard that will report expected minus observed postoperative morbidity as a continuous display to be piloted in 10 NHS hospitals.

Models to predict postoperative morbidity are less common in the literature than those predicting postoperative mortality. This may in part be related to the frequent use of mortality as a primary outcome measure in surgical studies. One commonly used morbidity risk-prediction model is the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM), which was first published in 1991.[166] Since its publication there have been several further iterations of the postoperative mortality element designed to improve predictive performance among different surgical specialties.[169–171] These modifications focussed solely on the mortality predictions however, and did not update the morbidity tool. As a consequence, one of the most frequently used morbidity risk prediction models is now more than 25 years old[172] and has been demonstrated to over-estimate morbidity in a recent surgical population.[173] The SORT morbidity tool is a more contemporary morbidity prediction model that is a modification of the validated SORT mortality tool.[67,174] The development dataset for the SORT morbidity model varies in terms of case-mix and complexity of surgery from the population eligible for recruitment to PQIP however and therefore its applicability in the planned setting be questioned (see Table 3-1).

One further tool of note is the American College of Surgeons National Surgical Quality Improvement Programme (ACS-NSQIP) risk prediction calculator.[79] This calculator provides predictions for a range of postoperative complications and has

been demonstrated to have good discrimination.[79] A potential drawback of this model is the large number of variables needed, many of which are not routinely collected within the PQIP dataset. The ACS-NSQIP model was derived and is utilised in the American healthcare system and it has not been validated for use in a UK setting.

Given these limitations of existing risk models it was considered important to develop, internally and then temporally validate a bespoke risk-adjustment model for a UK population of patients undergoing major elective surgery for the purpose of enabling continuous reporting of postoperative morbidity outcomes.

Table 3-1: Comparison of cohort characteristics for model derivation and internal validation of the PQIP morbidity and the SORT morbidity[67] models

	PQIP model derivation cohort - n (%)	SORT morbidity development cohort - n (%)
n	4466 (100)	1934 (100)
Age [Median, IQR]	[67, 57-73]	[62.6, 48.7-71.7]
Female	1744 (39.1)	1121 (58.0)
ASA Physical status		
I	496 (11.1)	313 (16.2)
II	2738 (61.3)	1161 (60)
III	1181 (26.4)	432 (22.3)
IV	51 (1.1)	26 (1.3)
V		1 (0.1)
Surgical specialty		
Ortho	0 (0)	873 (45.1)
Colorectal	2314 (51.8)	652 (33.7)
UpperGI	425 (9.5)	108 (5.6)
Vascular	0 (0)	122 (6.3)
Bariatric	0 (0)	125 (6.5)
Other	1727 (38.7)	53 (2.7)
Severity of surgery		
Minor	595 (13.3)	108 (5.6)
Intermediate		108 (5.6)
Major		211 (10.9)
Xmajor	1785 (40.0)	1507 (77.9)
Complex	2086 (46.7)	

3.3 Model derivation and internal validation

3.3.1 Methods

3.3.1.1 Data collection

Data were collected prospectively within a web-based database as part of the national PQIP study.[167,168] Patient recruitment started in December 2016. At the time of analysis PQIP was recruiting consenting adult patients undergoing non-cardiac major elective surgery in NHS hospitals in England and Wales. Model derivation data was exported from the main study database on 1st February 2018. All case records locked at the time of data export were eligible for inclusion, and included surgery performed between the 1st December 2016 and 31st January 2018. Appendix A-1 shows the procedure list eligible for recruitment to PQIP at the time of model derivation and internal validation. Procedures that were not eligible at the time of model derivation but subsequently added are shown in italic.

3.3.1.2 Defining postoperative morbidity

Morbidity was defined using the Postoperative Morbidity Survey (POMS) at postoperative day 7.[76] This survey has been validated for use after major surgery and has been demonstrated to have good inter-observer reliability.[65,66] Due to concerns that POMS may be too sensitive in the setting of major surgery,[67] a previously defined subset of the nine-domain POMS criteria was used, termed 'POMSmajor' (previously referred to as 'high-grade').[67] POMSmajor includes those POMS criteria equivalent to a Clavien-Dindo grade II or above complication. The Clavien-Dindo grading of surgical complications is a frequently used classification, which divides complications into five grades ranging from a minor deviation from the normal course (grade I) to death (grade V).[77] Appendix C-1 shows the POMS domains and their equivalent Clavien-Dindo classification as previously defined by Wong et al.[67] Patients discharged before postoperative day 7 were assumed to have no POMS-defined morbidity.

3.3.1.3 Data cleaning

Outcome data

Cases who withdrew from the study before postoperative day 7 did not have outcome data collected and were therefore excluded. Any cases with a calculated length of stay >7 days but with missing outcome data were also excluded from further analysis. Patients with a postoperative length of stay <7 days were assumed to be morbidity free at postoperative day 7 and those patients who died prior to postoperative day 7 were assigned to have morbidity on postoperative day 7. This decision was made on the interpretation of mortality being the most severe form of morbidity.

Predictor variable data

Any variable with a missing data percentage above 20% was excluded from the full model. Attempts were made to contact sites directly to obtain missing data. Where this was possible values were entered manually into the dataset. Implausible values were removed and treated as missing. For the predictor variable body mass index (BMI) cases with a value of <12kg/m2 were removed. For cases where it was evident height and weight data had been entered into the incorrect field, values were manually corrected. The pattern of white cell count values >40^10^{-9}/L suggested that values may have been entered incorrectly by a factor of 10 (the normal range for a white cell count is 4.0-12.0^10^{-9}/L). Sites were contacted and asked to correct any erroneous values. Any values that remained >40^10^{-9}/L were removed and treated as missing.

Serum urea and serum creatinine variables were explored in combination due to the clinical association between them. Figure 3-1 shows a scatter plot of urea and creatinine values. The x- and y-axes scales have been log transformed in this figure. The serum urea values for the group of 11 cases that can be seen at the extreme right-hand side of the Figure 3-1 were removed and treated as missing data. This decision was based on the implausibility of having serum urea values in this range paired with the relatively normal serum creatinine values. It was felt likely

that these serum urea values may have been entered incorrectly by a factor of 10. Serum creatinine values were subsequently winsorised at the 99.5th centile.

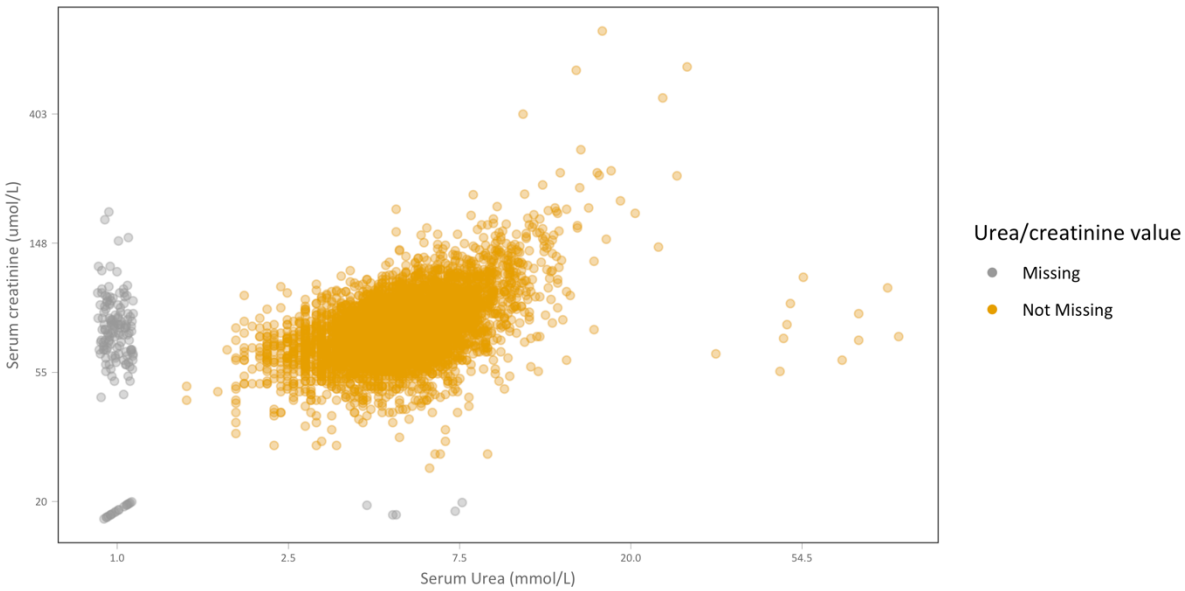


Figure 3-1: Scatter plot of serum urea and serum creatinine values prior to data cleaning. The group of 11 cases that can be seen to the right-hand side of the plot were subsequently removed and treated as missing data.

Additional data validation checks have been built into the PQIP database for the height, weight, white cell count and serum urea variables in response to this analysis to improve data quality for future analyses.

Table 3-2 shows how other predictor variables were managed. Continuous variables were winsorised at the 0.5th and 99.5th centiles where unusual outlier patterns were present. Implausible values were removed and treated as missing values. To reduce over-fitting of the model categorical variable groups that contained <1% of cases (<45 cases) were combined where clinically appropriate or excluded if not, in keeping with previously published work.[175,176] Cases were analysed on a complete case basis for predictor variables with a missingness of less than 20%.

Table 3-2: Candidate variables considered for inclusion in the PQIP model, data types and ranges within the derivation and internal validation cohort

Candidate variable	Type	Range of continuous variable (winsorised range)	Transformations
Surgical specialty	categorical		
Urgency of surgery	categorical		
Severity of surgery	categorical		Minor, intermediate and major combined
Age (years)	continuous	18-95	
Sex	categorical		
Body mass index (kg/m2)	continuous	14.4-61	
ASA grade	ordinal		ASA IV and V combined
Serum sodium (mmol/L)	continuous	113-150 (129-150)	
Serum potassium (mmol/L)	continuous	2.5-7.2 (3.3-5.6)	
Serum urea (mmol/L)	continuous	1.5-27.9	log transformed
Serum creatinine (umol/L)	continuous	26-767 (26-188)	log transformed
White cell count (x10^9/L)	continuous	1.7-27.1	
Haemoglobin (g/L)	continuous	62-200	
Preoperative heart rate (beats per minute)	continuous	38-162 (38-122)	
Systolic blood pressure (mmHg)	continuous	55-211	
Preoperative oxygen saturations (%)	continuous	84-100	
ECG findings	categorical		
Respiratory history findings	ordinal		Dyspnoea on light exertion and dyspnoea at rest categories combined
Cardiac history findings	ordinal		Peripheral oedema and raised JVP categories combined
New York Heart Association functional classification	ordinal		Category III and IV combined
History of diabetes	categorical		Type I and Type II (treated with insulin) categories combined
History of cerebrovascular disease	binary		Yes, without hemiparesis and Yes, with hemiparesis categories combined to give binary variable
Smoking history	categorical		
Alcohol consumption	categorical		
History of liver disease	categorical		
Respiratory infection in last 30-days	binary		

Candidate variable	Type	Range of continuous variable (winsorised range)	Transformations
Diagnosis of cancer in last 5 years	binary		
Operations in last 30 days	categorical		n = 2 and n = >2 categories combined into '2 or more' category
Albumin	continuous	NA	excluded due to missing data >20%
Glasgow coma scale	integer	NA	excluded

3.3.1.4 Model estimation

Twenty-eight candidate predictors were considered for inclusion in the initial 'full model' (see Table 3-2). All 28 candidate predictors have been used previously in published risk-prediction models,[166,174,177,178] and are routinely collected as part of the PQIP dataset. Only risk factors which were deemed non-modifiable at the point of admission for surgery were considered as predictor variables. Intra- and postoperative variables that may be influenced by the quality of care patients receive were excluded.

Five predefined interaction terms were considered for inclusion in the full model: ASA x respiratory history findings, ASA x age, age x systolic blood pressure, age x heart rate, and ASA x haemoglobin. For each interaction term with an initial strength of $p<0.10$ in the derivation cohort, the stability of across 100 bootstrap re-samples was tested.[179] Pre-defined criteria for inclusion in the full model was $p<0.05$ in at least 80% of bootstrap re-samples based on previously published research.[176]

In order to develop a parsimonious model that could be implemented in a software architecture, the most significant predictors of major postoperative morbidity were identified by fitting backwards-stepwise logistic regression models, selected on Akaike information criterion (AIC) across 1000 bootstrap re-samples.[180,181] Variables selected into at least 80% of final bootstrap models were included in the 'final model'. Penalised maximum likelihood estimation (PMLE)[182] was used to fit a logistic regression model using the variables selected in the backwards-stepwise procedure. PMLE was used to improve the predictive accuracy of the model without

sacrificing discriminative ability.[183,184] Instead of applying a linear shrinkage factor to correct for over-optimism, PMLE shrinks each regression coefficient individually[163,184–187] and maximises the penalised log likelihood rather than the log likelihood that is maximised in standard logistic regression modelling procedures (maximum likelihood estimation - MLE).

The penalised log likelihood is maximised by adjusting the maximum likelihood by the penalty factor:

$$logL - 0.5\lambda \sum_i (s_i\beta_i)^2,$$

where L is the maximum likelihood of the fitted model, λ the penalty factor, β the estimated regression coefficient for each predictor i in the model, and s_i is a scaling factor for each β_i to make $s_i\beta_i$ unitless.[163] The optimum penalty was calculated by fitting a series of logistic regression models with varying penalties. For each model the penalty factor, AIC and AIC corrected[188] are recorded. The penalty factor that gives the highest corrected AIC was chosen and applied when the final model was fit.

3.3.1.5 Internal validation of model in derivation cohort

Internal validation was performed using Harrell's bootstrap correction of optimism, across 1000 bootstrap re-samples.[163,180] As the model was fit using the entire cohort, its performance in future cohorts will be overestimated.[163] The optimism corrected C-statistic gives a more accurate estimation of future model performance than the apparent C-statistic when a model is fit using the entire dataset.[163] In order to calculate the optimism corrected C-statistic logistic regression models are fitted to bootstrap datasets produced from the original derivation cohort. Model performance in this bootstrap dataset is then compared to that when the same bootstrap model is applied to the original dataset. The difference between the two is termed the 'optimism'. By repeating the process '*b'* number of times (in the case of the PQIP model, *b=1000)* and then finding the mean of the optimism calculations across the *b* bootstrap samples it is possible to calculate the optimism corrected C-

statistic (equivalent to the area under the receiver operating characteristic curve - AUROC) using the equation:

$$\text{Optimism Corrected C-statistic} = \text{Apparent C-statistic}_{(\text{original sample})} - \text{mean(optimism)}_{(\textit{bootstrap samples})}$$

3.3.1.6 Comparison with published morbidity risk models

The performance of the model was compared to that of two previously published morbidity risk models (POSSUM[166] and SORT[67]). To ensure a fair comparison both models were recalibrated to the event rate within the derivation dataset. To do this the approach of Eugene et al.[12] was followed, firstly calculating the predicted log odds of postoperative morbidity for each patient in the derivation cohort using the published model formulae. A logistic regression model was then fitted to the predicted log odds, together with an intercept term. The log odds predicted by the original published formulae were then multiplied by the coefficient derived from the calibration logistic regression model. The estimated intercept was finally added to these predicted log odds to obtain a re-calibrated value.[12] Discrimination was assessed using the C-statistic and calibration was assessed both visually and with the Hosmer-Lemeshow test.

3.3.1.7 Statistical analysis

Statistical analyses were performed using R version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria), with the following external packages enabled: rms, tidyverse, DescTools, PredictABEL, pROC, bootStepAIC, plotROC and nanair. Means and standard deviations are reported for normally distributed data. For non-normally distributed data, medians and inter-quartile ranges (IQRs) are reported. Backwards stepwise model selection was performed using the bootStepAIC package. Logistic regression models were fit using the lrm function, and shrinkage factors were estimated using the val.prob function from the rms package.

3.3.2 Results

A total of 4466 were included in the model derivation and internal validation. Figure 3-2 summarises the patients included and excluded at each stage of the analysis. Table 3-3 shows summary statistics of the dataset used at each stage, including those patients excluded from the derivation dataset due to missing data. Figure 3-3 shows the data completeness of predictor variables after data cleaning and prior to removal of cases with missing data. Overall data completeness was excellent, after excluding serum albumin as a candidate predictor (proportion missing >20%), the remaining dataset had a missingness of 0.24% overall.

There were 838 (18.8%) patients with POMSmajor-defined morbidity at postoperative day 7 in the derivation cohort (see Table 3-4).

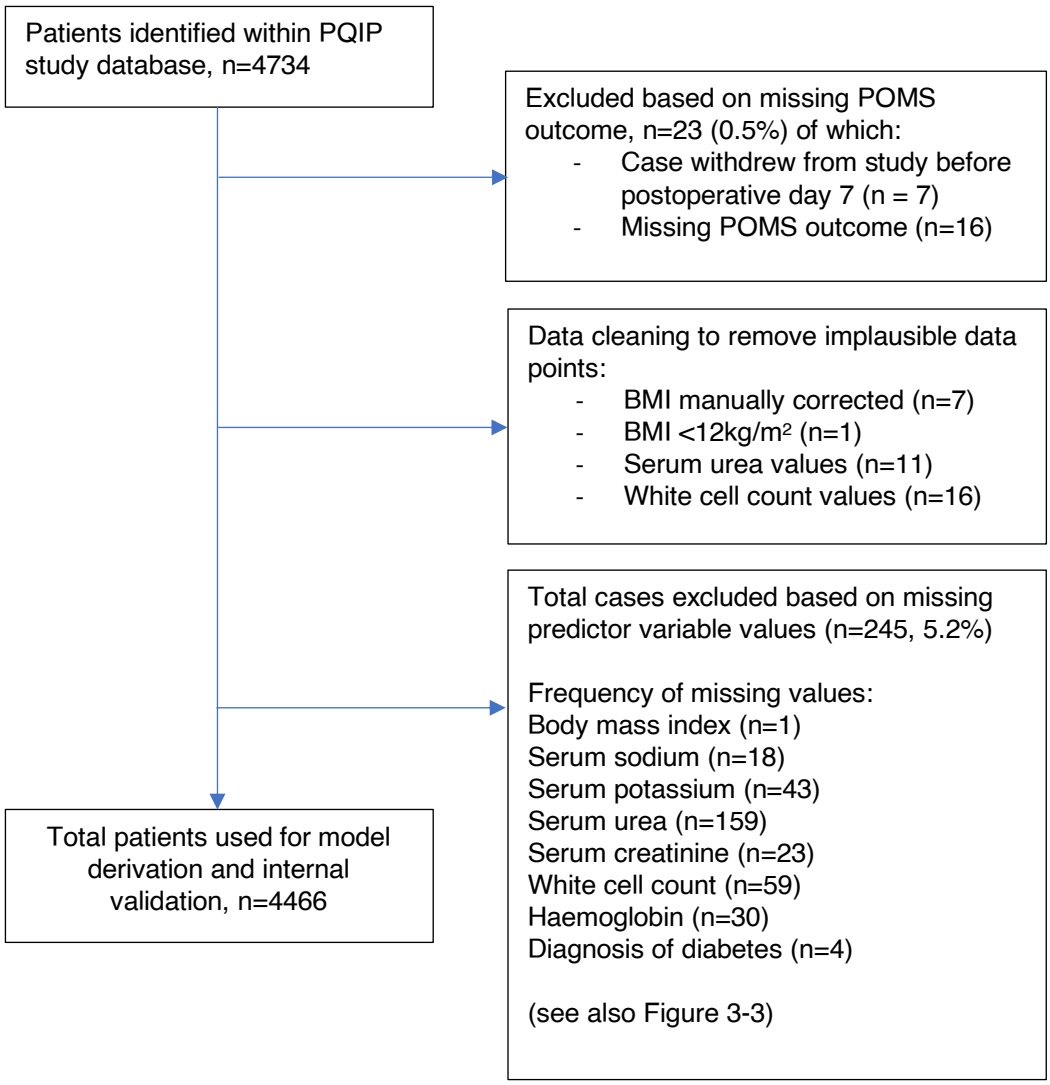


Figure 3-2: Flow diagram summarising cases included and excluded from the model derivation and internal validation dataset

Of the 28 predictor variables considered, ten were selected into the final model: surgical specialty, surgery severity category, gender, ASA grade, body mass index, preoperative heart rate, systolic blood pressure, age (years), number of operations in last 30 days, and respiratory history findings. The frequency with which each predictor variable was selected into the 1000 bootstrap models is shown in Appendix C-2.

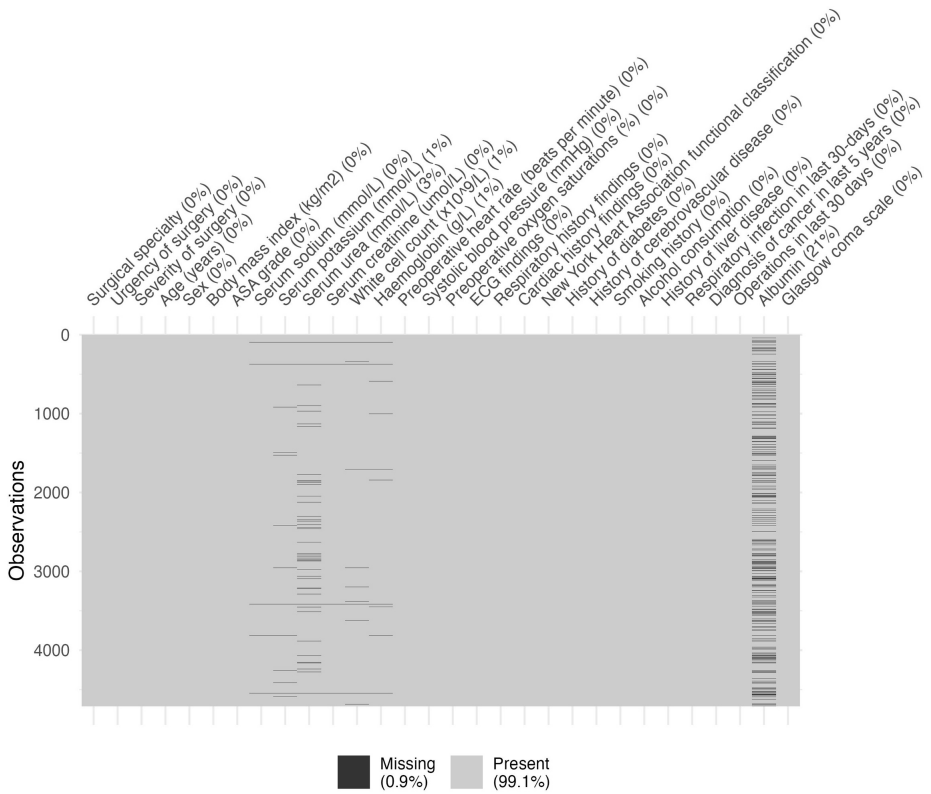


Figure 3-3: Missing data present in the derivation dataset after removal of implausible values

Figure 3-4 shows the AIC and AIC corrected for each penalty factor. The optimum penalty for PMLE model estimation was 8, which reduced the effective degrees of freedom used to fit the model from 19 to 16.9. The coefficients of the final logistic regression model fitted by PMLE are shown in Table 3-5.

The bootstrap optimism corrected C-statistic in the derivation cohort was:

Optimism Corrected C-statistic = Apparent C-statistic$_{(original\ sample)}$ - mean(optimism)$_{(bootstrap\ samples)}$

$$= 0.685 - 0.009$$

$$= 0.68 \text{ (95\% CI 0.66-0.69)}$$

Table 3-3: Descriptive data for derivation cohort and cases excluded owing to missing data

	Derivation and internal validation cohort [n=(%)]	Cases excluded due to missing data [n=(%)]
Demographic data		
Age (years) [median, IQR]	[67, 57-73]	[66, 58-73]
Female sex	1744 (39.1%)	88 (35.9%)
Body mass index (kg/m2) [median, IQR]	[27, 23.9-30.4]	[27.5, 23.8-30.4]
Surgical specialties		
Lower gastrointestinal	2314 (51.8%)	122 (49.8%)
Urology	585 (13.1%)	29 (11.8%)
Upper gastrointestinal	425 (9.5%)	17 (6.9%)
Hepatobiliary	447 (10%)	21 (8.6%)
Thoracics	422 (9.4%)	8 (3.3%)
Head and neck	127 (2.8%)	43 (17.6%)
Abdominal - other	146 (3.3%)	5 (2%)
Severity of surgery		
Minor/Intermediate/Major	595 (13.3%)	21 (8.6%)
Xmajor	1785 (40%)	98 (40%)
Complex	2086 (46.7%)	126 (51.4%)
NCEPOD classification		
Elective	3991 (89.4%)	204 (83.3%)
Expedited	475 (10.6%)	41 (16.7%)
ASA physical status		
I	496 (11.1%)	33 (13.5%)
II	2738 (61.3%)	149 (60.8%)
III	1181 (26.4%)	62 (25.3%)
IV/V	51 (1.1%)	1 (0.4%)
Comorbidities		
Diagnosis of cancer in last 5 years	3387 (75.8%)	192 (78.4%)
NYHA classification		
I	3741 (83.8%)	210 (85.7%)
II	609 (13.6%)	30 (12.2%)
III/IV	116 (2.6%)	5 (2%)
Respiratory history findings		
No dyspnoea	3787 (84.8%)	206 (84.1%)
Dyspnoea on exertion or chest x-ray: mild COPD	533 (11.9%)	32 (13.1%)
Dyspnoea limiting exertion to <1 flight of stairs or chest x-ray: moderate COPD or Dyspnoea at rest/respiratory rate > 30 at rest or chest x-ray: fibrosis or consolidation	146 (3.3%)	7 (2.9%)
Cardiac history findings		
No failure	3414 (76.4%)	191 (78%)
Diuretic, digoxin, antianginal or antihypertensive therapy	963 (21.6%)	51 (20.8%)
Peripheral oedema, warfarin therapy or borderline cardiomegaly, raised jugular	89 (2%)	3 (1.2%)

	Derivation and internal validation cohort [n=(%)]	Cases excluded due to missing data [n=(%)]
venous pressure or cardiomegaly		
Cerebrovascular disease with or without hemiparesis	179 (4%)	9 (3.7%)

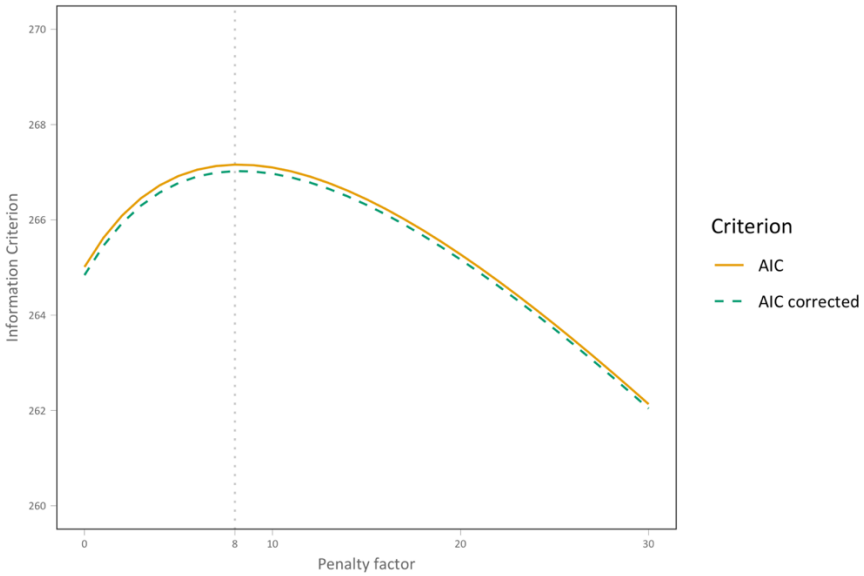


Figure 3-4: Optimal penalty factor to optimise the corrected AIC when fitting the PQIP model using penalised maximum likelihood estimation

Table 3-4: POMS outcomes at postoperative day 7 in the derivation cohort

	Derivation and internal validation cohort	Excluded based on missing predictor variable data (derivation)
Total number of patients [n (%)]	4466 (100)	245 (100)
Patients still in hospital at postoperative day 7 [n (%)]	2112 (47.3)	108 (44.1)
POMS-defined outcomes		
Patients with POMS-defined morbidity [n (%)]	1299 (29.1)	65 (26.5)
Patients with POMSmajor-defined morbidity [n (%)]	838 (18.8)	46 (18.8)

Figures in parentheses represent the number of patients suffering POMS- or POMSmajor-defined morbidity as a percentage of the total number of patients in either cohort

The area under the receiver operating characteristic curve (AUROC) and calibration plot for PQIP model in the derivation cohort is shown in Figure 3-5. Calibration of the PQIP model for each surgical specialty in the derivation and internal validation cohort is shown in Figure 3-6.

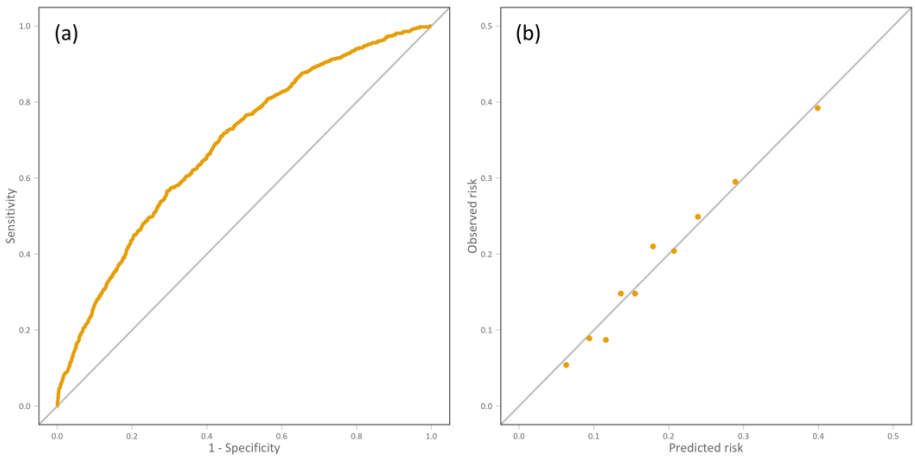


Figure 3-5: (a) AUROC curve for the PQIP morbidity model and (b) calibration plot comparing the observed day-7 POMSmajor morbidity against predicted in deciles of predicted risk

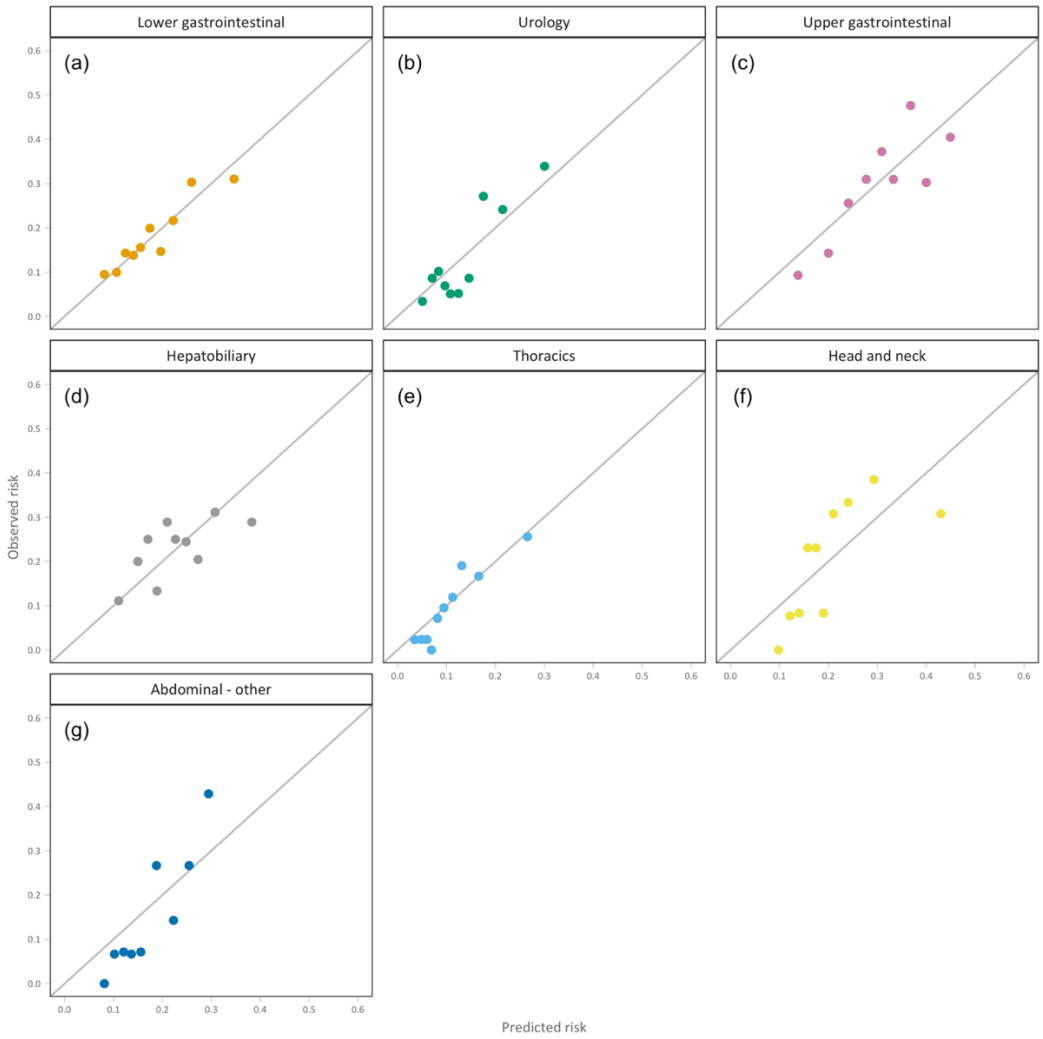


Figure 3-6: Calibration plot comparing the observed day-7 POMSmajor morbidity against that predicted from the PQIP model in deciles of predicted risk for each surgical specialty in the derivation and internal validation dataset

Table 3-5: Coefficients for the PQIP-morbidity adjustment model. PQIP, Perioperative Quality Improvement Programme

	Coefficient	Standard error	z-value	p-value
Intercept	-3.8525	0.4716	-8.17	<0.0001
Surgical specialty				
Lower gastrointestinal		Reference		
Urology	-0.2462	0.1348	-1.83	0.0677
Upper gastrointestinal	0.5464	0.1212	4.51	<0.0001
Hepatobiliary	0.1108	0.1286	0.86	0.3887
Thoracics	-0.4661	0.1788	-2.61	0.0091
Head and neck	0.0179	0.2037	0.09	0.9301
Abdominal - other	0.2941	0.2003	1.47	0.1420
Severity of surgery				
Minor/Intermediate/Major		Reference		
Xmajor	0.4519	0.1714	2.64	0.0084
Complex	0.9666	0.1634	5.92	<0.0001
Sex				
Male		Reference		
Female	-0.4087	0.0840	-4.86	<0.0001
American Society of Anaesthesiologists Physical Status Classification System (ASA-PS)				
I		Reference		
II	0.2927	0.1446	2.02	0.0429
III	0.6069	0.1573	3.86	0.0001
IV/V	0.6517	0.2895	2.25	0.0244
Body mass index				
Per 1-unit increase (kg/m2)	0.0253	0.0075	3.39	0.0007
Heart rate				
Per 1-unit increase (beats per minute)	0.0108	0.0029	3.76	0.0002
Systolic blood pressure				
Per 1-unit increase (mmHg)	-0.0055	0.0022	-2.53	0.0115
Age (years)				
Per 1-unit increase (years)	0.0088	0.0034	2.61	0.0091
Number of operations in last 30 days (including planned operation)				
One		Reference		
Two or more	0.4509	0.1520	2.97	0.0030
Respiratory history findings (POSSUM variable)				
Normal findings		Reference		
Dyspnoea on exertion	0.4542	0.1141	3.98	<0.0001
Dyspnoea on light exertion/Dyspnoea at rest	0.5910	0.1846	3.20	0.0014

3.3.3 Model performance compared to existing morbidity models

The discrimination of the PQIP model was superior to those of the POSSUM and SORT morbidity despite recalibration to the event rate in the derivation cohort (C-statistic of PQIP model: 0.68 (95% CI 0.66-0.69) vs. POSSUM morbidity: 0.62 (95% CI 0.60-0.64) and SORT morbidity: 0.52 (95% CI 0.50-0.54) - Table 3-6). Calibration of the PQIP model was also superior to POSSUM and SORT morbidity models across the range of risk predictions in the derivation dataset (Figure 3-7).

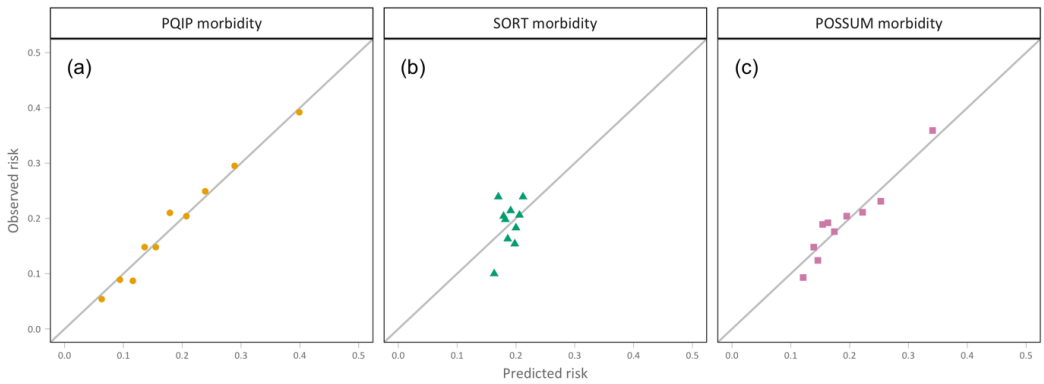


Figure 3-7: Calibration plot comparing the observed day-7 POMSmajor morbidity against predicted in deciles of predicted risk for: (a) PQIP model (Perioperative Quality Improvement Programme); (b) SORT morbidity (Surgical Outcome Risk Tool morbidity); and (c) POSSUM morbidity model

Table 3-6: Discrimination of the PQIP risk model compared to published models. POSSUM, Portsmouth Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity; SORT morbidity, Surgical Outcome Risk Tool; PQIP, Perioperative Quality Improvement Programme

Model	C-statistic in development study/other external validation	Sample size of original study	C-statistic within PQIP derivation dataset (95% CI) - n=4466
PQIP morbidity	0.68		0.68 (0.66-0.69)
POSSUM			
Copeland et al.[166]	Not stated	1372	0.62 (0.6-0.64)
Jones et al.[189]	0.82	117	
Brooks et al.[190]	0.92	949	
Gonzalez-Martinez et al.[191]	0.77	721	
Bagnall et al.[173]	0.51	1380	
SORT morbidity			
Wong et al.[67]	0.72	1583	0.52 (0.5-0.54)

3.3.4 Sensitivity analyses

Given the poor calibration when predicting outcomes of patients undergoing head and neck surgery (Figure 3-6), a sensitivity analysis was performed to assess the difference in coefficients estimated after excluding these patients from the derivation cohort. Table 3-7 shows the minimal effect of excluding these patients on the coefficients of the model. The apparent C-statistic of this adjusted model in the derivation cohort was unchanged (0.68, 95% CI 0.66-0.70)

Table 3-7: Change in PQIP risk-model coefficients for sensitivity analysis excluding patients undergoing head and neck surgery from derivation cohort

	Coefficient estimated in PQIP model	Coefficient estimated in sensitivity analysis (excluding head and neck patients)
Intercept	-3.8525	-3.9195
Surgical specialty		
Lower gastrointestinal	Reference	
Urology	-0.2462	-0.2468
Upper gastrointestinal	0.5464	0.5598
Hepatobiliary	0.1108	0.1238
Thoracics	-0.4661	-0.4658
Head and neck	0.0179	NA - excluded
Abdominal - other	0.2941	0.3107
Severity of surgery		
Minor/Intermediate/Major	Reference	
Xmajor	0.4519	0.4718
Complex	0.9666	0.9677
Sex		
Male	Reference	
Female	-0.4087	-0.4137
American Society of Anaesthesiologists Physical Status Classification System (ASA-PS)		
I	Reference	
II	0.2927	0.2777
III	0.6069	0.5670
IV/V	0.6517	0.6428
Body mass index		
Per 1-unit increase (kg/m2)	0.0253	0.0281
Heart rate		
Per 1-unit increase (beats per minute)	0.0108	0.0108
Systolic blood pressure		
Per 1-unit increase (mmHg)	-0.0055	-0.0057
Age (years)		
Per 1-unit increase (years)	0.0088	0.0092
Number of operations in last 30 days (including planned operation)		
One	Reference	
Two or more	0.4509	0.4542
Respiratory history findings (POSSUM variable)		
Normal findings	Reference	
Dyspnoea on exertion	0.4542	0.4599
Dyspnoea on light exertion/Dyspnoea at rest	0.5910	0.6146

3.4 Temporal validation

3.4.1 Methods

3.4.1.1 Data collection

A temporal validation was performed using a cohort not available at the time of initial model derivation. Data were collected prospectively in the same national PQIP database as model derivation [167,168]

Temporal validation data were exported on the 1st April 2019, and included patients undergoing surgery between the 1st February 2018 and 31st March 2019. Appendix A-1 shows the procedure list eligible for recruitment to PQIP at the time of model derivation and validation. Procedures that were not present in the model derivation dataset but were eligible for inclusion at the time of temporal validation data export are shown in italic.

3.4.1.2 Data cleaning

Outcome data

The same outcome data cleaning procedure applied to the derivation data was applied to the temporal validation data. Cases who withdrew from the study before postoperative day 7 did not have outcome data collected and were therefore excluded. Cases with a calculated length of stay >7 days but with missing outcome data were also excluded. Patients with a postoperative length of stay <7 days were assumed to be morbidity free at postoperative day 7. Patients who died before postoperative day 7 were assigned to have morbidity on postoperative day 7, a decision based on the interpretation of mortality being a severe form of morbidity in keeping with the decision made at the model derivation stage.

Predictor variable data

The same predictor variable data cleaning procedures performed at the model derivation stage were applied to temporal validation data. Implausible values were removed and treated as missing. For the predictor variable body mass index (BMI)

any cases with a value of <12kg/m^2 or >65kg/m^2 were removed and treated as missing. As during model derivation, BMI values were corrected where it was clear height and weight data had been entered into the incorrect field. White cell count values >40^10-9/L were removed and treated as missing.

Serum urea values were removed and treated as missing where the serum urea was recorded as >27mmol/L with a serum creatinine of <134umol/L; the same rule applied to remove implausible outliers in the derivation dataset owing to concerns that urea values may have been entered incorrectly by a factor of 10. Any continuous variables winsorised in the model derivation analysis were again winsorised at the same values used during model derivation (see Table 3-2).

The grouping of categorical variables used in the model derivation stage were applied to temporal validation data. Table 3-2 shows which categories were grouped.

Model discrimination in the temporal validation cohort was assessed using the C-statistic, calibration was assessed visually and with the Hosmer-Lemeshow goodness of fit test.[164] Overall model performance was measured with the Brier score.[165] As the model is used to report risk-adjusted outcome at the surgical specialty level a sensitivity analysis assessing model performance within specialties was performed. A second sensitivity analysis was performed to assess model performance after exclusion of patients undergoing robotic prostatectomies, owing to a decision to exclude these from recruitment to PQIP between model derivation and the temporal validation stages.

3.4.1.3 Comparison of PQIP model performance to that of published morbidity models

Model performance of the PQIP model in the temporal validation cohort was compared to the two previously published morbidity risk models (POSSUM[166] and SORT[67]). Previously published models were recalibrated to the event rate within the temporal validation dataset using the same method as that used at the derivation stage.[12] The PQIP model was not calibrated to the event rate in the

temporal validation dataset, instead the equation calculated at the model derivation stage was used.

3.4.2 Results

As variables selected into the final PQIP model at the derivation stage were mandatory within the PQIP dataset, the impact of missing data in the temporal validation cohort was lower than that in the derivation cohort (see Figure 3-2 and Figure 3-8).

Table 3-8 shows the case mix of the model derivation/internal validation cohort compared that of the temporal validation and Table 3-9 shows postoperative morbidity outcomes for the derivation/internal validation and temporal validation cohorts.

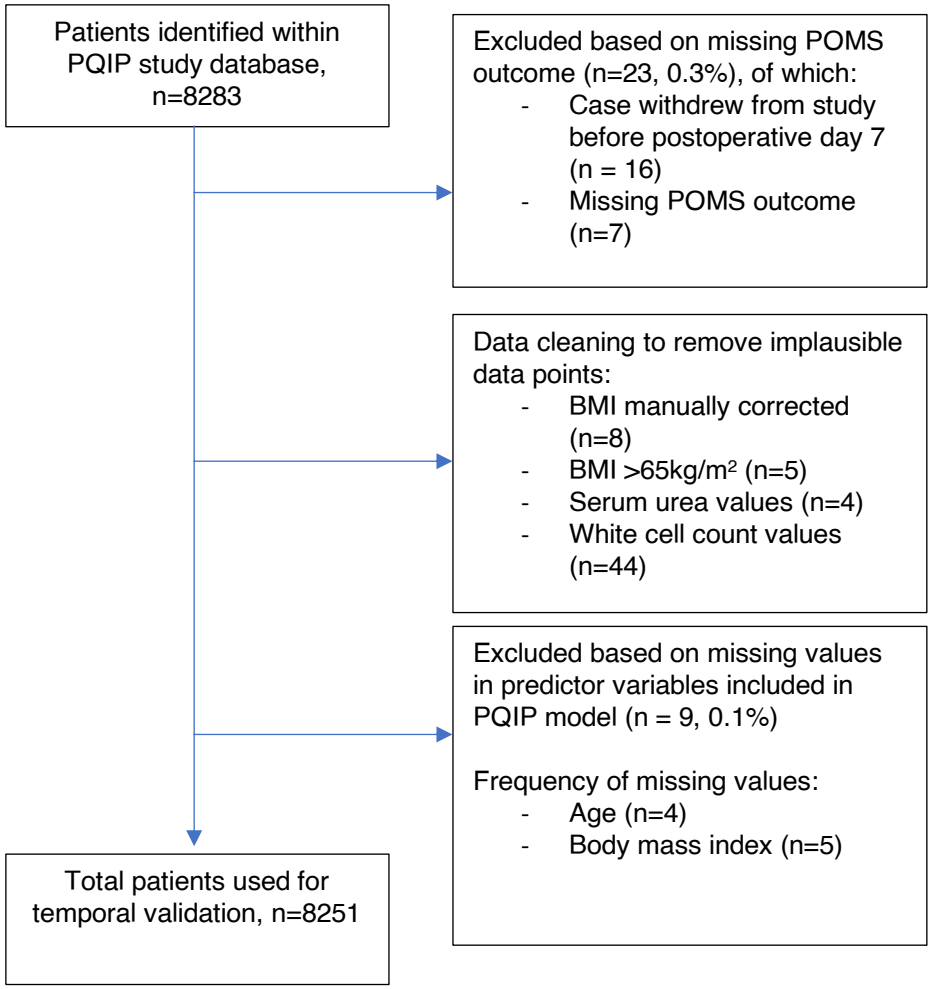


Figure 3-8: Cases included and excluded from temporal validation dataset

Table 3-8: Comparison of derivation/internal validation and temporal validation cohorts

	Derivation and internal validation cohort [n=(%)]	Temporal validation cohort [n=(%)]
Demographic data		
Age (years) [median, IQR]	[67, 57-73]	[66, 56-73]
Female sex	1744 (39.1%)	3236 (39.2%)
Body mass index (kg/m2) [median, IQR]	[27, 23.9-30.4]	[27, 23.9-30.6]
Surgical specialties		
Lower gastrointestinal	2314 (51.8%)	4047 (49%)
Urology	585 (13.1%)	1348 (16.3%)
Upper gastrointestinal	425 (9.5%)	723 (8.8%)
Hepatobiliary	447 (10%)	828 (10%)
Thoracics	422 (9.4%)	753 (9.1%)
Head and neck	127 (2.8%)	228 (2.8%)
Abdominal - other	146 (3.3%)	324 (3.9%)
Severity of surgery		
Minor/Intermediate/Major	595 (13.3%)	1042 (12.6%)
Xmajor	1785 (40%)	2743 (33.2%)
Complex	2086 (46.7%)	4466 (54.1%)
NCEPOD classification		
Elective	3991 (89.4%)	7397 (89.6%)
Expedited	475 (10.6%)	854 (10.4%)
ASA physical status		
I	496 (11.1%)	827 (10%)
II	2738 (61.3%)	5134 (62.2%)
III	1181 (26.4%)	2199 (26.7%)
IV/V	51 (1.1%)	91 (1.1%)
Comorbidities		
Diagnosis of cancer in last 5 years	3387 (75.8%)	6227 (75.5%)
NYHA classification		
I	3741 (83.8%)	6847 (83%)
II	609 (13.6%)	1202 (14.6%)
III/IV	116 (2.6%)	202 (2.4%)
Respiratory history findings		
No dyspnoea	3787 (84.8%)	7068 (85.7%)
Dyspnoea on exertion or chest x-ray: mild COPD	533 (11.9%)	934 (11.3%)
Dyspnoea limiting exertion to <1 flight of stairs or chest x-ray: moderate COPD or Dyspnoea at rest/respiratory rate > 30 at rest or chest x-ray: fibrosis or consolidation	146 (3.3%)	249 (3%)
Cardiac history findings		
No failure	3414 (76.4%)	6230 (75.5%)
Diuretic, digoxin, antianginal or antihypertensive therapy	963 (21.6%)	1860 (22.5%)

	Derivation and internal validation cohort [n=(%)]	Temporal validation cohort [n=(%)]
Peripheral oedema, warfarin therapy or borderline cardiomegaly, raised jugular venous pressure or cardiomegaly	89 (2%)	161 (2%)
Cerebrovascular disease with or without hemiparesis	179 (4%)	323 (3.9%)

Table 3-9: POMS outcomes at postoperative day 7 in the derivation and temporal validation cohorts

	Derivation and internal validation cohort	Temporal validation cohort
Total number of patients [n (%)]	4466 (100)	8251 (100)
Patients still in hospital at postoperative day 7 [n (%)]	2112 (47.3)	3648 (44.2)
POMS-defined outcomes		
Patients with POMS-defined morbidity [n (%)]	1299 (29.1)	2087 (25.3)
Patients with POMSmajor-defined morbidity [n (%)]	838 (18.8)	1445 (17.5)

Figures in parentheses represent the number of patients suffering POMS- or POMSmajor-defined morbidity as a percentage of the total number of patients in either cohort

The C-statistic of the PQIP model was maintained between the derivation cohort and the temporal validation cohort (0.68, 95% CI 0.66-0.68 – see Table 3-10). The Brier score was 0.14, indicating good overall performance of the model. Discrimination of the POSSUM and SORT morbidity models in the temporal validation cohort was similar to that in the derivation cohort (Figure 3-9).

The Hosmer-Lemeshow test suggested a possible lack of fit ($p<0.05$) and the PQIP model was shown to overestimate risk in the 17.2-22.6% predicted risk band (see Figure 3-10a). This overestimation may be attributable to the poor calibration of predicted and observed morbidity risk in those patients undergoing urology and head and neck surgery in the temporal validation cohort (see Figure 3-11b).

Table 3-10: Discrimination of the PQIP model compared to published models within the temporal validation cohort. POSSUM, Portsmouth Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity; SORT morbidity, Surgical Outcome Risk Tool; PQIP, Perioperative Quality Improvement Programme

Model	C-statistic in development study/other external validation	Sample size of original study	C-statistic within PQIP temporal validation dataset (95% CI) - n=8251
PQIP morbidity	0.68		0.68 (0.66-0.69)
POSSUM			
Copeland et al.[166]	Not stated	1372	0.61 (0.59-0.63)
Jones et al.[189]	0.82	117	
Brooks et al. [190]	0.92	949	
Gonzalez-Martinez et al.[191]	0.77	721	
Bagnall et al. [173]	0.51	1380	
SORT morbidity			
Wong et al.[67]	0.72	1583	0.53 (0.51-0.55)

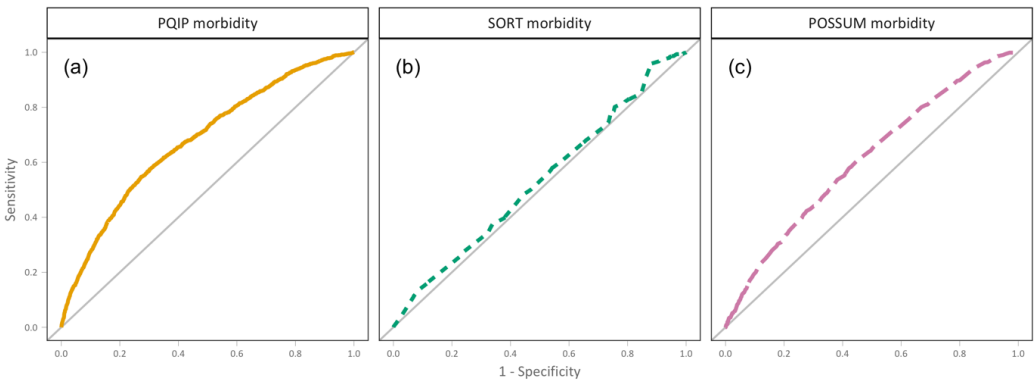


Figure 3-9: Area under the receiver operating characteristic curve for: (a) PQIP morbidity (Perioperative Quality Improvement Programme); (b) SORT morbidity (Surgical Outcome Risk Tool morbidity); and (c) POSSUM (Portsmouth Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity) risk models in the temporal validation dataset. SORT morbidity and POSSUM have been calibrated to the event rate in the temporal validation dataset.

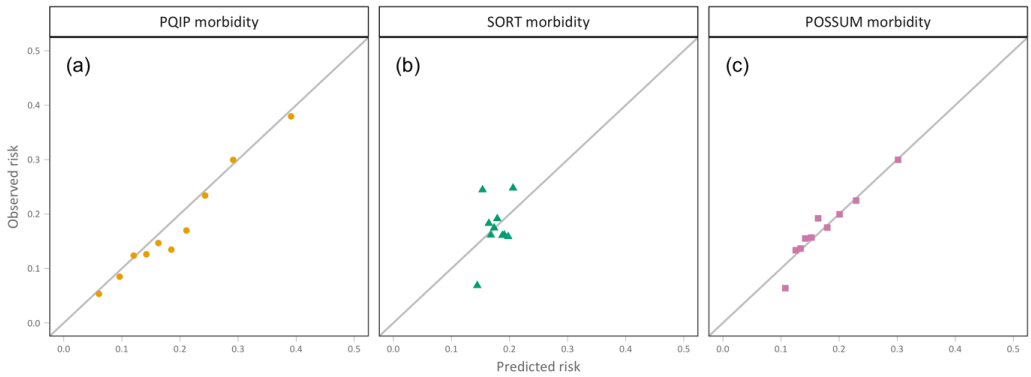


Figure 3-10: Calibration plot comparing the observed day-7 POMSmajor morbidity against that predicted by: (a) PQIP morbidity model; (b) SORT morbidity model; and (c) POSSUM morbidity model, in the temporal validation dataset.

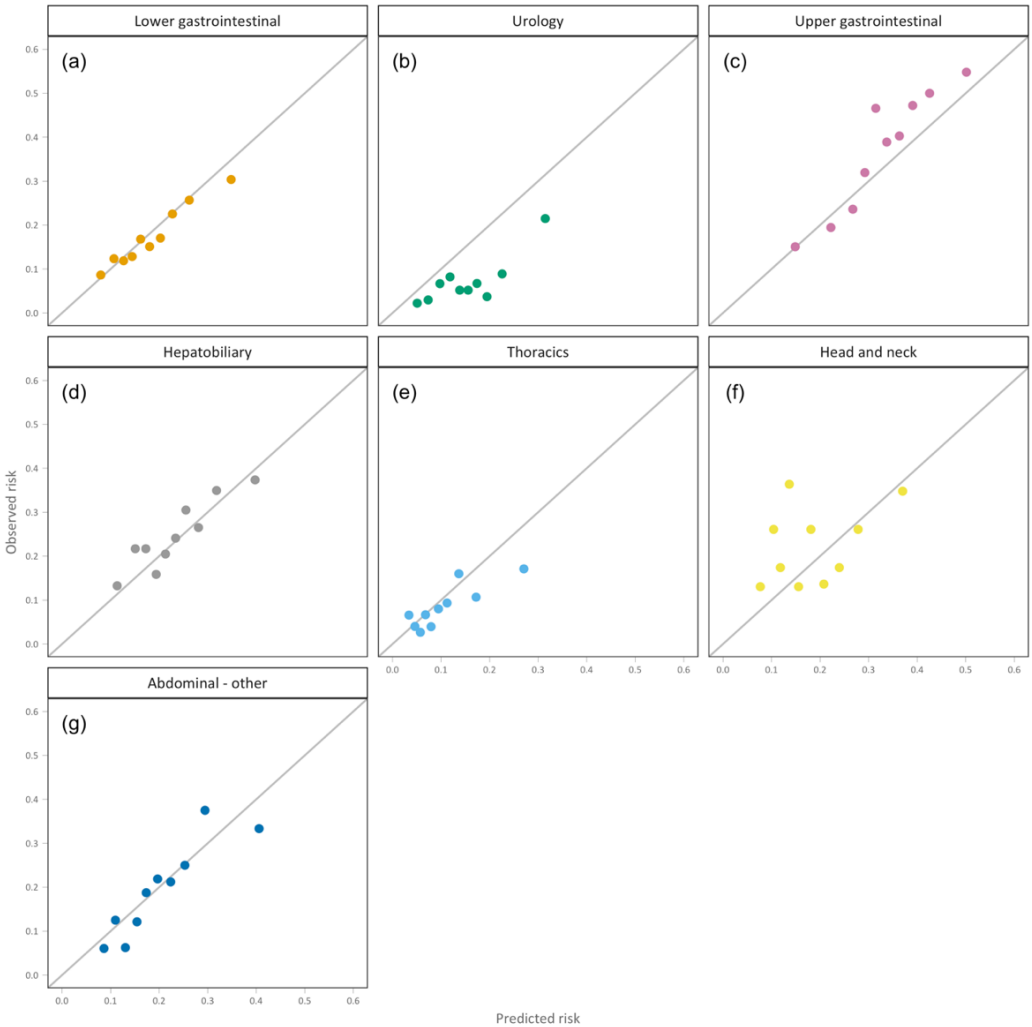


Figure 3-11: Calibration plot comparing the observed day-7 POMSmajor morbidity against that predicted by the PQIP morbidity model in deciles of predicted risk for each surgical specialty in the temporal validation cohort

3.4.3 Further exploration of model calibration

Whilst overall model calibration was acceptable in the temporal validation cohort (Figure 3-10), the over-estimation of risk in the 17.2-22.6% band was further investigated. This over-estimation of risk was particularly evident in patients undergoing Urology surgery (Figure 3-11b). During recruitment of patients into the two datasets changes were made to PQIP eligibility criteria. Patients undergoing prostatectomies with the use of robotic assistance were excluded from recruitment. These cases had lower morbidity rates compared to patients undergoing other Urology procedures (see Table 3-11). The performance and calibration of the PQIP model was therefore also assessed in the temporal validation cohort after exclusion of patients undergoing robotic prostatectomies. Figure 3-12 shows the calibration plot for the PQIP model in the temporal validation cohort for Urology surgery including robotic prostatectomies (Figure 3-12a) and after excluding them (Figure 3-12b).

Appendix C-4 explores how the application of linear shrinkage factors can be used improve model calibration across the different surgical specialties.

Table 3-11: Comparison of morbidity outcomes in robotic prostatectomy procedures and all other urological procedures in the temporal validation cohort

	Urology procedures excluding robotic prostatectomies	Robotic prostatectomies	
Total cases (n)	992	356	
Cases with POMSmajor morbidity (n)	93	3	
Predicted risk from PQIP model - mean, [+/- SD]	0.145 [+/- 0.083]	0.180 [+/- 0.050]	
Observed morbidity rate	0.094	0.008	(p<0.001)

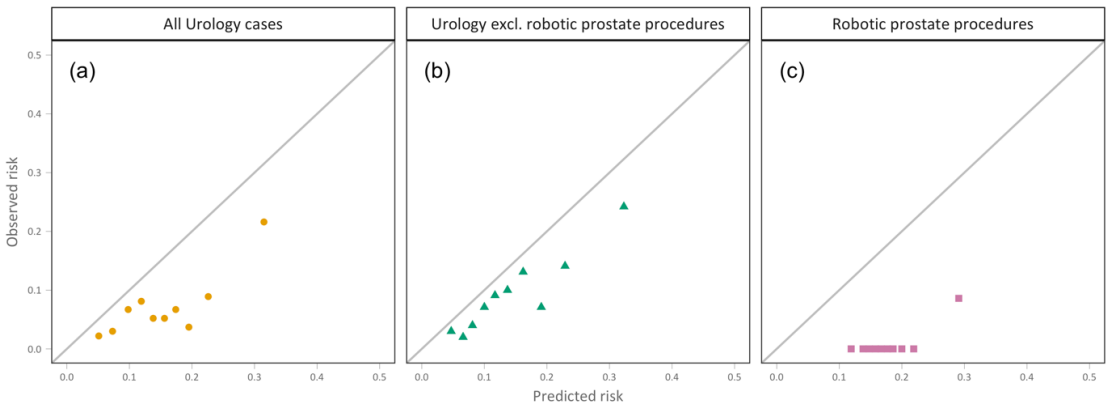


Figure 3-12: Calibration plot comparing the observed day-7 POMSmajor morbidity against that predicted by the PQIP morbidity model in deciles of predicted risk for cases undergoing Urology surgery in the temporal validation cohort with: (a) all Urology cases; (b) Urology cases excluding robotic prostatectomies; and (c) only robotic prostatectomies

3.5 Discussion

3.5.1 Main findings

A logistic regression model has been developed and temporally validated to risk-adjust postoperative day 7 morbidity outcomes in the setting of major surgery. Model discrimination was acceptable 0.68 (95% CI 0.66-0.69). Model discrimination was maintained between derivation and temporal validation cohorts. The model out performed published morbidity risk models[67,166] in both derivation and temporal validation cohorts which included patients undergoing major surgery across a heterogeneous group of surgical specialties.

A major strength of this study is the data quality with low overall missingness. The multi-centre dataset, with 63 hospitals contributing to the development cohort and 89 to the validation cohort hopefully supports model performance being maintained in future similar populations.

3.5.2 Performance of model vs comparators

The poor calibration of the SORT morbidity model is likely to be a reflection of the low discrimination of the model in both the derivation and temporal validation datasets (C-statistic 0.52 and 0.53 in the derivation and temporal validation cohorts respectively). This performance may be due to the difference in case-mix and complexity of surgery between the cohort used to develop the SORT morbidity development and that used in this study. The SORT morbidity model was derived from a single centre cohort of patients undergoing generally less severe, predominantly orthopaedic procedures (Table 3-1).

After calibration to both derivation and temporal validation cohorts the POSSUM model still overestimated risk in the lowest decile of risk prediction (Figure 3-10). The smaller range of predicted risk deciles compared to the PQIP model also demonstrates a poorer ability to differentiate between higher and low risk groups.

3.5.3 Limitations

3.5.3.1 Dataset and predictor variables

The derivation and validation cohort included patients undergoing major elective and expedited surgery in England and Wales. Very few patients underwent a procedure classified as minor or intermediate in severity. For this reason, minor, intermediate and major categories were grouped together. It is likely that the model will over-estimate risk if applied to a cohort of patients undergoing minor and intermediate procedures and its use in this group of patients is therefore cautioned.

Whilst seven different surgical specialties are represented within the model, the majority of patients underwent a general surgical procedure (lower gastrointestinal, upper gastrointestinal, hepatobiliary, abdominal - other). The heterogeneous nature of the sample may be beneficial in terms of the wide-ranging applicability of the model, but its ability to discriminate the risk of morbidity may have been reduced by the heterogeneity.

Intra- and postoperative variables that may be influenced by the care patients receive were excluded and not considered as possible candidate variables. The rationale for this was that the intended application of the risk model is to monitor postoperative outcomes and support hospital teams monitor and improve their outcomes. The inclusion of intra- and postoperative variables would potentially result in the outcome being adjusted for the quality of care being delivered, not just risks directly related to the patient. A potential downside of this decision to exclude intra- and postoperative variables is that the performance (particularly discriminative power) of the model may have been reduced.

3.5.3.2 Choice of outcome

Although the discrimination of the model developed was acceptable following recalibration, confidence intervals spanned the 0.70 cut-off that is considered to represent moderate discrimination.[87,163] It is important to acknowledge that the discriminative ability of morbidity models is frequently lower than that of mortality models in the published literature.[67,79,166,172,174] This poorer discrimination may be due to a greater range of factors impacting on morbidity risk compared to those of mortality. Mortality is frequently used as an outcome measure in clinical studies, and it is also therefore possible that clinical datasets designed to capture information relating to mortality do not adequately capture variables that impact on morbidity risk.

The method of defining the morbidity outcome may also affect overall model performance, with a stricter definition potentially leading to improved discriminative ability. The choice of POMSmajor as the outcome for this work was based on recent published morbidity modelling work and the concern that POMS may be too sensitive in the setting of major surgery.[67] When comparing the performance of the model derived here with the SORT morbidity model we used the 'high-grade' (POMSmajor) shrinkage factor presented by Wong et al. to predict morbidity.[67] This ensured a fair comparison of performance. The difference in the way morbidity was defined in the POSSUM derivation study may account for some of the difference in model performance. Recalibration of the POSSUM model to the event rate in the

validation cohort improved calibration but did not change the discrimination of the POSSUM model.

Whilst other morbidity or complication outcomes such as the original POMS criteria or the Clavien-Dindo grading of surgical complications were available as potential outcome measures, the intended application of the model for monitoring and use in a quality improvement setting meant that POMSmajor was viewed as a more appropriate outcome. As POMS can be divided into its nine constituent organ domains (Appendix C-1), the choice of POMSmajor as an outcome measure presents the potential for future work to develop domain or organ specific morbidity predictions and monitoring. These more detailed predictions and monitoring would allow clinical teams to develop more precise interventions to reduce risk and improve patient and hospital outcomes.

3.5.3.3 Statistical analysis and modelling technique

Model accuracy varied between individual surgical specialties. This variation may reflect the heterogeneity of patients presenting for each type of surgery. Calibration of the PQIP model appears best for the general surgical population (lower gastrointestinal, upper gastrointestinal, hepatobiliary and the abdominal - other group – see Figure 3-11). The calibration in the gastrointestinal specialties is contrasted by the relatively poor calibration in the other specialties (urology, thoracics, and head and neck surgery). This was particularly true for patients undergoing head and neck surgery (Figure 3-11f). The appropriateness of the model to monitor morbidity outcomes in these specialties may therefore be questioned. Poorly calibrated models may result in risk-adjusted outcomes appearing better than they are (if the model systematically over-estimates risk) or worse than they are (if the model underestimates risk), resulting in the potential for inappropriate investigation and incorrect conclusions being drawn from the monitoring process.

The risk-factors that contribute significantly to risk of postoperative morbidity are likely to differ between specialties. The method of variable selection and application of simple logistic regression may not be optimal for such a wide range of surgical

specialties. The model was developed as a single level logistic regression model. The variation in model performance across the range of specialties raises the possible need for more complex risk-modelling approaches, such as multi-level (also referred to as hierarchical) modelling. Alternatives to multilevel modelling include developing a model for each specialty if adequate data are available or to estimate linear shrinkage factors for each surgical specialty, thereby allowing the intercept and slope of the log odds of morbidity to vary between specialties. Multilevel modelling using surgical specialty as a higher order term may have improved performance across the range of specialties. The ACS-NSQIP risk adjustment process has recently adopted multilevel modelling to improve their predictions.[192]

The use of backwards variable selection techniques has been suggested to cause inflation of regression coefficients, as variables included in the model are not penalised appropriately to take into account variables removed during the selection procedure.[163] This is true of any variable selection procedure, including when univariate analysis is performed to assess the relationship between predictor variables and outcome prior to model fitting. By uncoupling variable selection from model fitting the result is a model that often performs poorly in future cohorts. The model is effectively over-fitted to the derivation cohort. In an attempt to counteract the variable selection procedure and prevent this over-fitting the model was fitted using PMLE. This provides a penalty factor, which effectively shrinks each regression coefficient, therefore improving predictive performance in external, unknown cohorts.[184] Despite the reduction in over-fitting by applying PMLE, it is likely that the penalty factor applied during the derivation stage may still not have been severe enough. Variables removed during the variable selection procedure were not considered in the penalty calculation applied during the model fitting process.

Machine learning techniques such as Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic-Net are two regularised regression techniques that combine variable selection and model fitting into one procedure.[193,194] Because these methods combine the variable selection with model fitting, the regression

coefficients are penalised appropriately which can help to reduce over-fitting and produce more stable predictive performance when applied to new data.[195] Whilst these methods have been used in the wider literature for several years, they have not yet been applied widely in the field of perioperative risk prediction.

3.6 Conclusions

A bespoke risk-adjustment model for major morbidity at postoperative day-7 in a high-risk UK surgical population was developed and validated. Model discrimination was maintained between derivation and temporal validation and calibration was acceptable. Restricting risk-adjustment to variables that are independent of the care patients receive such as demographic characteristics, baseline clinical status, or the planned operation they undergo, enables clinicians to evaluate outcomes over time in a manner that accounts for patient risk without inadvertently adjusting for the quality or nature of intraoperative and postoperative care. For example, the inclusion of ASA grade, a standardised measure of preoperative physical status, allows for appropriate case-mix adjustment while preserving the ability to detect variation in care delivery and outcomes.

The parsimonious model (10 variables) derived is simple to implement in a software solution and will be used to report risk-adjusted morbidity outcomes in near real-time through a web-based dashboard.

The model developed had higher discrimination and better calibration than existing risk models.[67,166]Model performance varied across specialties however, and future work should consider the use of multilevel modelling techniques or the use of specialty specific models as further data becomes available within the PQIP study.

The poor performance of existing risk models within the high-risk surgical population of the temporal validation cohort highlights the importance of understanding the case-mix and setting in which risk-adjustment models are developed. The application of a model to an external cohort that varies significantly from that used to derive it may result in inadequate performance, even after

recalibration to the event rate in the new cohort. Models applied to a case-mix and setting similar to those used to derive them they may still require frequent updating as event rates, in this case morbidity outcomes, change over time. The increased availability of large perioperative datasets through projects such as PQIP and ACS-NSQIP will enable more frequent updating of risk-adjustment models than has been done historically, ensuring estimates of risk remain accurate. This is particularly important where the risk-adjustment process is used to monitor outcomes over time within and between institutions, where failure to update models may result in falsely reassuring or alarming results.

Whilst the performance of the risk-adjustment model developed in this study was acceptable, it important to note that no risk-adjustment process is perfect. There will always be additional factors that impact upon patients' risk that are not accounted for in the model estimates. It is important to understand the limitations of a risk-adjustment model and any context in which it may perform poorly. Understanding these limitations will support its application into clinical settings.

Chapter 4 Intervention aims, design, and methods

4.1 Background

Improvement work is described as having a 'hard core' and 'soft core'.[196] In Donabedian's framework of structure, process, and outcome the 'hard core' might be considered to be specific processes that are targeted for improvement.[6] A process may be improving the frequency with which a drug is given to patients who it is indicated in, or in the setting of the ePOCH study a process targeted for improvement was the use of lung protective ventilation strategies.[73] The ePOCH (Enhanced Peri-Operative Care for High-risk patients) study was a large-scale, stepped-wedge cluster-randomised trial conducted across 93 UK NHS hospitals, that evaluated whether a national quality improvement program implementing a comprehensive care pathway could reduce 90-day mortality in patients undergoing emergency laparotomy.[197,198] Despite good clinician engagement and modest improvements in some process measures, the study found no significant survival benefit from the intervention, highlighting the complexities and resource challenges in implementing large-scale quality improvement initiatives in emergency surgical care.[197,198]

The 'soft core' is a "gammut of complimentary arrangements involved in delivering the benefit that may take a variety of different forms".[196] The soft core could be contextual elements that play a pivotal role in the success or failure of improvement strategies. Local context such as resource and time availability, multidisciplinary engagement, and team culture all play an important role. Despite an understanding that local context plays a key role in successful implementation, data about how and why strategies succeed, or fail are frequently not collected. This lack of contextual qualitative data reduces the learning available from quality improvement work.[64]

A rapid-feedback evaluation (RFE) is a study design that aims to 'provide focused, timely evaluation conclusions'.[199,200] The RFE model[199] includes five steps:

1. analysis of existing data on programme performance
2. collection of new data on programme performance
3. preliminary evaluation
4. development and analysis of alternative designs for full-scale evaluation
5. assisting policy and management decisions

Rapid-feedback evaluations do not always prelude more complex evaluations. The information gathered by them may be adequate to answer the question posed with no further evaluation needed.[201] Although the RFE of the pomVLAD project is a stand-alone evaluation, its results will complement a more complex process evaluation of the PQIP study.[105]

The interrupted time series (ITS) study design is used for the evaluation of public health interventions and is particularly suited to interventions over a clearly defined period and that target population-level health outcomes.[202–204] The controlled interrupted time series (CITS) is an extension of the ITS design and includes a comparison group which do not receive the assigned intervention. It is arguably a stronger quasi-experimental design as two controls are present (baseline trend in the intervention group and the presence of the control group). The presence of the two controls allows for both within- group and between-group comparisons.[205] In order to perform a CITS analysis time series data are needed, with a minimum of three data points before and after the intervention.[206]

4.2 Objectives and research questions

4.2.1 Aims and objectives

1. Develop and implement an online dashboard reporting risk-adjusted postoperative morbidity outcomes and compliance with associated enhanced recovery recommendations to sites participating in the PQIP study

2. Evaluate whether access to near real time risk-adjusted data supports engagement of clinical teams with quality improvement
3. Document the views of PQIP collaborators of the dashboard and any perceived link to quality improvement activity
4. Explore PQIP collaborators' use and experience of using the dashboard and PQIP quarterly reports for the purposes of quality improvement
5. Identify factors that may be acting as barriers and facilitators to collaborator engagement with the dashboard, the data it displays, and subsequent quality improvement initiatives
6. Develop a series of recommendations for improving site engagement with their data displays and/or modify the dashboard
7. Develop a series of recommendations to increase the use of the dashboard and site reports for local quality improvement

4.2.2 Research questions

1. Does the implementation of an online dashboard reporting risk-adjusted postoperative morbidity outcomes and perioperative care recommendations in near real time result in a reduction in the incidence of postoperative morbidity in NHS hospitals recruiting to the PQIP study?
2. Does reporting of risk-adjusted morbidity outcomes and performance in delivery of a range of perioperative care processes result in greater compliance with those care processes?
3. Does the availability of near real time risk-adjusted morbidity data increase engagement of clinicians with their local data?
4. What is the programme theory supporting the pomVLAD intervention?
5. What are staff members' perceptions of the pomVLAD intervention?
6. Do local investigators use the pomVLAD dashboard and PQIP quarterly reports? If so, how?
7. How do staff members describe their experiences of using the pomVLAD dashboard?

8. What factors act as barriers to site collaborators engaging with their local data and pomVLAD dashboard and using them in quality improvement initiatives?
9. Are there factors that act as facilitators to increase engagement and use of the dashboard/reports for quality improvement?
10. What are PQIP collaborators' recommendations for improving the dashboard?
11. What are PQIP collaborators' recommendations to increase the use of the dashboard to drive quality improvement?

4.3 Methods

4.3.1 Study design

A mixed-methods evaluation of the pomVLAD intervention was performed. This included a controlled interrupted time-series evaluation of quantitative outcome measures alongside a qualitative rapid feedback evaluation[207] capturing staff experiences and implementation processes.

4.3.2 Study population

4.3.2.1 Quantitative analysis inclusion criteria

Patient level inclusion criteria

Patients recruited to the PQIP study and undergoing colorectal surgery between 1st July 2017 and 30th June 2019 at one of the 20 identified NHS hospital sites were eligible for inclusion in the analysis. Eligible surgical procedures are shown in Table 4-1. Patients undergoing concurrent, non-colorectal procedures were included in the analysis if the primary operation recorded was a colorectal procedure. Data were exported from the main study database on 10th September 2019. Records locked by local investigators, and therefore considered to be complete, were included in the analysis.

Hospital level inclusion criteria and group allocation

The top 20 recruiting hospitals of patients undergoing colorectal surgery in the PQIP study, as of 14th March 2018 were included in the study. All records (complete and incomplete) were considered when calculating patient recruitment. Twenty sites were chosen to allow equal allocation to intervention and control groups. The decision to include the top 20 recruiting sites aimed to ensure adequate patient recruitment through the 12-month implementation period.

Following identification of the top recruiting 20 sites, 10 sites were randomly allocated to the intervention group using computer software.[208] Sites not allocated to the intervention group became the control group. Sites were allocated pseudonyms to anonymise outcome and interview data.

Table 4-1: Colorectal procedures eligible for recruitment to the PQIP study

Surgical procedure
Abdominal operation for Hirschprung's disease
Abdominal revision of restorative proctocolectomy
Abdominoperineal resection with anastomosis (+/- pouch)
Abdominoperineal resection with end colostomy
Abdominoperineal pull-through resection with colo-anal anastomosis +/- colonic pouch and associated stoma
Anterior resection
Colectomy (total and ileorectal anastomosis)
Colectomy and colostomy and preservation of rectum
Completion proctectomy
Excision of retroperitoneal tumour (+/- ureterolysis)
Exploratory laparotomy
Hartmann's procedure
Ileoanal anastomosis and creation of pouch
Ileo-caecal resection (with anastomosis or ileostomy formation)
Laparotomy for enterocutaneous fistula
Left hemicolectomy (with anastomosis/colostomy)
Pan proctocolectomy and ileostomy
Partial excision of rectum and sigmoid colon for prolapse
Redo operations on ileum and colon
Resection of small bowel (+/- tumour)
Reversal of Hartmann's procedure
Right hemicolectomy (with anastomosis/colostomy)
Sigmoid colectomy
Total mesorectal excision (TME) including trans-anal/TATME
Transanal endoscopic microsurgery

4.3.2.2 Qualitative evaluation inclusion criteria

Five control and five intervention sites were identified for inclusion in the qualitative study. To obtain interview data from a range of local contexts, sites were stratified into one of three groups:

1. Higher engagement and higher compliance: These sites actively interacted with the central PQIP study team regarding the pomVLAD intervention and consistently adhered to the secondary outcome measures defined in 4.3.3.2 below, using PQIP data up to 14th March 2018.
2. Lower engagement and higher compliance: These sites demonstrated good adherence to the secondary outcome measures in their PQIP data but had less subjective interaction with the central PQIP study team regarding the pomVLAD intervention.
3. Lower engagement and lower compliance: These sites had less interaction with the central PQIP study team and showed poorer adherence to the secondary outcome measures based on their PQIP data.

These categories were chosen to ensure that the qualitative study obtained interview data from a range of local contexts with the aim of capturing perspectives from sites that varied in their:

Engagement with the central PQIP study team and the pomVLAD intervention: This was subjectively assessed based on interactions after notification of allocation to the intervention group potentially reflecting their responsiveness to discussions about pomVLAD and general proactivity in engaging with the study.

Compliance with secondary outcome measures: This was objectively measured using PQIP data up to March 14, 2018. This reflected a site's adherence to the key process measures prior to the implementation of pomVLAD. This compliance was considered relevant as high levels of compliance may have represented pre-existing improvement activity and potentially also have reduced a site's ability to increase adherence during the study period. Conversely, sites with lower

compliance prior to the intervention may have had a greater scope for improvement during the study period.

By stratifying based on both "engagement" (which includes responsiveness to pomVLAD discussions and overall interaction) and "compliance" with process measures, the qualitative study sought to understand how different levels of site participation and adherence might influence the implementation and outcomes of the intervention. This approach aimed to ensure that the qualitative findings were representative of the diverse experiences across the participating sites.

Site allocation to the specified groups is shown in Table 4-2. Sites were then selected at random from each of the three strata, maintaining proportionate representation from each level. Five control sites for interview were selected at random from the 10 potential sites. Random selection was performed using computer software.[208] Sites were pseudo-anonymised for the analysis by changing NHS hospital site names to famous climbs from the Tour de France.

The five control sites randomly selected to participate in the rapid feedback evaluation were: Alpe d'Huez, Col de l'Iseran, Col du Télégraphe, Col de la Croix de Fer, and Puy de Dome.

Table 4-2: Rapid feedback evaluation site stratification and allocation

Hospital site	Sites included in qualitative analysis
Higher engagement, higher compliance	
Col du Tourmalet	Col du Tourmalet
Col d'Aubisque	Col de Vars
Col de Vars	Col des Aravis
Col de Portet d'Aspet	
Col des Aravis	
Higher engagement, lower compliance	
Col d'Aspin	Col d'Allos
Col d'Allos	
Col de Peyresourde	
Lower engagement, lower compliance	
Col du Galibier	Col du Galibier
Col d'Izoard	

4.3.3 Outcomes

4.3.3.1 Definition of primary outcome measure

Morbidity was defined using a subclassification of the Postoperative Morbidity Survey (POMS), 'POMSmajor' previously termed 'high-grade' morbidity, at postoperative day 7.[67,76] Risk-adjustment of POMSmajor morbidity was performed using the risk-adjustment model developed in Chapter 3.

4.3.3.2 Definition of secondary outcome measures

Secondary outcomes were calculated as percentages aggregated to calendar month at the hospital level. Definitions and the denominator used for each secondary outcome calculation are shown in Table 4-3.

Table 4-3: Secondary outcome definition and the denominator used in their calculation

Secondary outcome measure	Definition	Denominator
Preoperative carbohydrate loading	Percentage of patients documented as receiving a clear carbohydrate drink approximately 2 hours prior to surgery	All patients not diagnosed with diabetes*
Intraoperative warming	Percentage of patients who received forced air warming AND intravenous fluid warming in theatre	All patients*
Avoidance of nasogastric tubes	Percentage of patients arriving in the postoperative recovery area WITHOUT a nasogastric tube in situ	All patients*
Avoidance of abdominal drains	Percentage of patients arriving in postoperative recovery WITHOUT an abdominal drain in situ	All patients*
DrEaMing (Drinking, Eating, and Mobilising) within 24 hours of surgery ending	Drinking – patients tolerating free fluids within 24 hours of surgery ending Eating – patients tolerating at least soft diet within 24 hours of surgery ending Mobilising – patients mobilising from bed to chair with maximal assistance of one person within 24 hours of surgery ending	All patients*

Secondary outcome measure	Definition	Denominator
	DrEaMing – percentage of patients meeting ALL definitions of drinking, eating, and mobilising above.	
Patients receiving all process measure recommendations	Percentage of patients meeting all eligible secondary outcome measures	All patients who meet denominator criteria for secondary outcomes * above
Records locked within 14 days of hospital discharge	Percentage of records locked in 14 days or less from hospital discharge	All patients who survived to hospital discharge

4.3.4 Study time-period

The PQIP study started recruiting patients in December 2016 and was originally planned to continue until December 2022.[26] Patient level data used for the CITS included patients undergoing surgery between 1st July 2017 and 30th June 2019, allowing one year of data pre- and post-implementation to be analysed. Data available were restricted to these dates.

The preintervention period was defined as the 1st July 2017 to 30th June 2018. The intervention point was defined as 1st July 2018, a timepoint when the pomVLAD dashboard was available to sites and local investigators had either attended the breakout session at a PQIP collaborative event or been offered and individual site initiation call. There was no transition period defined in the analysis. The postintervention period ran from 1st July 2018 to 30th June 2019.

Due to varying recruitment numbers, with some hospitals recruiting only 1-2 patients per week, data were aggregated to calendar month at the hospital level for the CITS analysis. Aggregation to a longer period, such as quarterly, would have reduced the number of time points available for trend analysis. A shorter period, such as weekly, would have increased the statistical noise in the analysis owing to small recruitment numbers.

Interviews with site investigators were conducted in two rounds. Baseline interviews took place between August and September 2018. Second round interviews were conducted between April and June 2019.

4.3.5 Intervention

4.3.5.1 Definition of intervention and theory of change

The 'pomVLAD intervention' consisted of the following three elements:

1. pomVLAD dashboard – an online data display incorporating the VLAD showing risk-adjusted morbidity outcomes and performance against the five secondary outcomes
2. Site initiation briefing – this included either attendance of local investigators at a breakout session at one of the national PQIP collaborative events or an individual site setup phone call with the local principal investigator for the PQIP study
3. Introductory booklet and supporting information (Appendix D-1 and Appendix D-2) and website hosting further information and resources about the pomVLAD project[209]

Figure 4-1 shows the theory of change for the pomVLAD intervention and the modalities used to analyse its outcomes. It was hypothesised that site initiation and provision of supporting materials would encourage local data collection and engagement with the pomVLAD dashboard. This engagement would lead to quality improvement activity based on the enhanced recovery process measures shown in the dashboard. Improvement in compliance with these measures and other associated improvement activity would then lead to reduced postoperative morbidity outcomes.

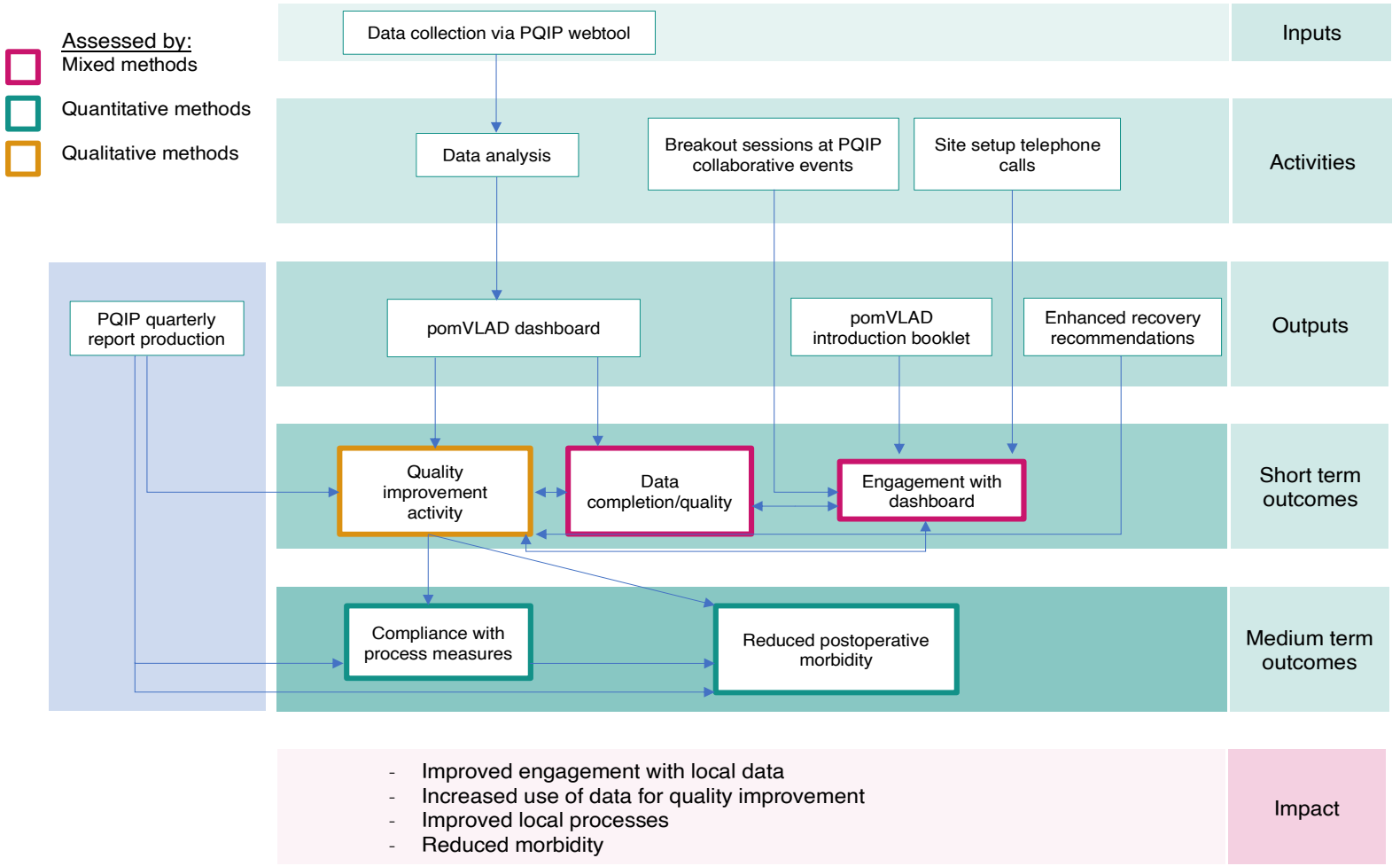


Figure 4-1: Theory of change for the pomVLAD intervention

4.3.5.2 Development of the dashboard

The PQIP data collection webtool and existing dashboards were developed and maintained by Netsolving software developers.[210] The pomVLAD dashboard was developed by JB/SC/SRM with clinical input from the PQIP national project team in conjunction with Netsolving. It was deployed within the existing PQIP website, accessible to local investigators after log in.

Users were able to interact with the display which allowed selection of date ranges displayed and subsetting by surgical specialty. Process measures were shown in dial format above the main VLAD display. Dials were colour coded according to performance; pink/red < 60% compliance, orange 60-79% compliance, and green 80% and above compliance (see Figure 4-2).

4.3.5.3 Selection of process measures reported on the dashboard

Enhanced recovery recommendations made by the Enhanced Recovery After Surgery (ERAS) Society for patients undergoing colorectal surgery were used as a basis for the process measures reported in the pomVLAD dashboard.[32,211,212] Recommendations made by the ERAS Society were compared to the PQIP case record form to assess whether adequate information was collected to allow compliance reporting.

ERAS Society recommendations that were captured in the PQIP case record form were considered for inclusion in the pomVLAD dashboard. Due to the one-year implementation period of the pomVLAD intervention a decision was made to focus on recommendations that would not require the wholesale establishment of new services or care pathways. To develop a refined enhanced recovery care 'bundle' the association between documented enrolment on an enhanced recovery pathway and delivery of processes recommended by the ERAS Society was analysed, using data from the PQIP study (Table 4-5). Patients undergoing colorectal surgery on or before the 30th January 2018 were included in this enhanced recovery analysis which aimed to identify the processes of care associated with what local sites identified as an 'enhanced recovery pathway'.

Five enhanced recovery process measures were incorporated into the pomVLAD dashboard: Preoperative carbohydrate loading; intraoperative warming; avoidance of nasogastric tubes; avoidance of surgical site drains; delivery of DrEaMing (drinking, eating, and mobilising) within 24 hours of surgery ending. Table 4-3 provides information on how these processes of care were defined. Five processes were selected to provide a concise enhanced recovery bundle of care, avoid over burdening local investigators with data, and provide specific targets for improvement.[213]

Table 4-4: Data collection of ERAS Society recommendations within the PQIP study

ERAS Society recommendation[32,211,212]	Data collected
Preadmission	
Cessation of smoking and excessive intake of alcohol	Yes
Preoperative nutrition screening and, as needed, assessment and nutritional support	No
Medical optimisation of chronic disease	No
Preoperative	
Structured preoperative information and engagement of the patient and relatives or caretakers	No
Preoperative carbohydrate treatment	Yes
Preoperative prophylaxis against thrombosis	No
Preoperative prophylaxis against infection	Yes
Prophylaxis against nausea and vomiting	No
Intraoperative	
Minimal invasive surgical techniques	Yes
Standardized anaesthesia, avoiding long-acting opioids	Partially
Maintaining fluid balance to avoid over- or underhydration, administer vasopressors to support blood pressure control	No
Epidural anaesthesia for open surgery	Yes
Restrictive use of surgical site drains	Yes
Removal of nasogastric tubes before reversal of anaesthesia	Yes
Control of body temperature using warm air flow blankets and warmed intravenous infusions	Yes
Postoperative	
Early mobilization (day of surgery)	Yes (day 1)
Early intake of oral fluids and solids (offered the day of surgery)	Yes (day 1)
Early removal of urinary catheters and intravenous fluids (morning after surgery)	Yes (day 1)
Use of chewing gums and laxatives and peripheral opioid-blocking agents (when using opioids)	No
Intake of protein and energy-rich nutritional supplements	No
Multimodal approach to opioid-sparing pain control	No
Multimodal approach to control of nausea and vomiting	No
Prepare for early discharge	No
Audit of outcomes and process in a multiprofessional, multidisciplinary team on a regular basis	No

Table 4-5: Care processes delivered by enrolment on enhanced recovery pathway

ERAS Society recommendation	Enrolled on pathway (n=1809)	Not enrolled on pathway (n=327)	Pathway status unknown (n=313)	p-value
Current smokers referred to smoking cessation clinic	39/193 (20.2%)	6/34 (17.7%)	4/29 (13.7%)	p=0.050
Antibiotic prophylaxis within 60 minutes of knife to skin	1772 (98.0%)	317 (96.9%)	302 (96.5%)	p=0.200
Preoperative carbohydrate loading provided*	1372 (75.8%)	134 (41.0%)	128 (40.9%)	p<0.001
Neuroaxial anaesthesia for planned open procedures	396/567 (69.8%)	97/162 (59.9%)	83/122 (68.0%)	p=0.060
Avoidance of surgical site drains*	1066 (58.9%)	169 (51.7%)	154 (47.1%)	p<0.001
Nasogastric tubes not present on arrival in recovery*	1692 (93.5%)	253 (77.4%)	285 (91.1%)	p<0.001
Use of forced air warming and warmed IV fluids* intraoperatively	1245 (68.8%)	148 (45.3%)	206 (65.8%)	p<0.001
IV fluids discontinued on postoperative day 1	1098 (60.7%)	126 (38.5%)	174 (55.6%)	p<0.001
Drinking on postoperative day 1	1556 (86.0%)	227 (69.4%)	251 (76.8%)	p<0.001
Eating on postoperative day 1	1224 (67.7%)	145 (44.3%)	184 (58.8%)	p<0.001
Mobilising on postoperative day 1	1483 (82.0%)	206 (63.0%)	244 (78.0%)	p<0.001
DrEaMing (Drinking, Eating and Mobilising) on postoperative day 1*	1077 (59.5%)	111 (33.9%)	166 (53.0%)	p<0.001

*Notes: Patients were recorded as either being on an enhanced recovery pathway, not on a pathway or pathway status unknown. Pearson's Chi-squared test was used to test for difference between process measure delivery and enhanced recovery pathway enrolment category. * denotes processes or care incorporated into pomVLAD dashboard.*

4.3.5.4 Development of the variable life-adjusted display

Risk-adjustment of postoperative morbidity was incorporated into the VLAD display, using the standard approach of displaying observed minus expected outcomes over time.[91] Expected outcomes were calculated using the risk-adjustment model derived in Chapter 3. The x-axis for the VLAD was calendar date, rather than case number which is often reported in the literature. This decision was made so that results had appropriate time context for local site investigators, allowing improved monitoring of outcomes in relation to timed interventions. Figure 4-2 shows an example of the pomVLAD dashboard display. During the first iteration of the pomVLAD dashboard local investigators could not see the number of patients that were recruited in the selected time frame. There were other dashboards available to them on the PQIP website that did show this information. Following feedback during the first round of interviews recruitment numbers were added to the pomVLAD dashboard.

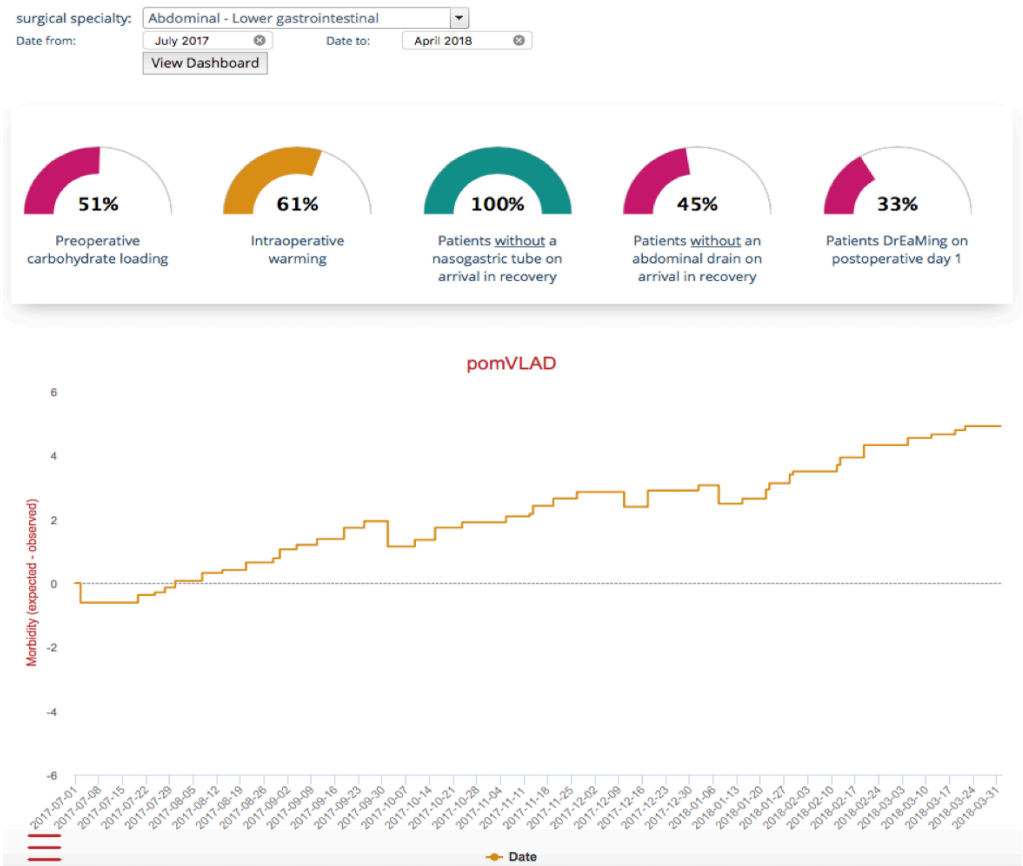


Figure 4-2: Example pomVLAD dashboard display

4.3.5.5 Implementation of the pomVLAD intervention

Local investigators were contacted and notified their site was allocated to the intervention group prior to the 16th April 2018. Two PQIP events for local collaborators were held in London and Manchester on the 16th April 2018 and 12th June 2018, respectively. These events were timed to coincide with the release of the first national PQIP annual report highlighting national performance against a range of perioperative process and outcome measures.[47]

Local investigators from intervention sites attending either of these meetings were invited to a group breakout session where the pomVLAD project, its aims, and objectives were explained. Background information about the variable life-adjusted display was also provided as well as information on how to access and interpret the pomVLAD dashboard (Appendix D-3). Questions raised by local clinicians about the project were answered at this time. Additional email contact was made with registered local investigators at each of the intervention sites on 3rd May 2018, providing further information about the study. Sites with no representatives present at either of the national events were offered a site initiation phone call to provide the information given at these events. Breakout sessions, telephone calls, and email contact were all delivered by a single investigator (JB).

The pomVLAD dashboard was live and viewable to site investigators as of the 17th May 2018.

4.3.5.6 Tailoring and modification of intervention

Following the first round of interviews changes were made to the dashboard based on feedback from site investigators. The aim of these changes was to support quality improvement activity and increase engagement with the dashboard. Changes made included:

1. Addition of run charts below the VLAD display showing percentage compliance with each of the five process measures by calendar month (see Figure 4-3 and Figure 4-4 below)
2. Identification of cases with postoperative morbidity in the VLAD display. The VLAD was modified to display the ID of individual cases when the cursor was placed over a relevant change point of the VLAD line.

The first change aimed to support improvement work related to the five process measures by adding time series data plots. The second change allowed local investigators to easily identify cases that caused either an upward or downward shift in the VLAD. During the first round of interviews with intervention sites local investigators felt easier identification of cases would support investigation and improved understanding of local morbidity outcomes. Updates to the dashboard were deployed in February 2019.

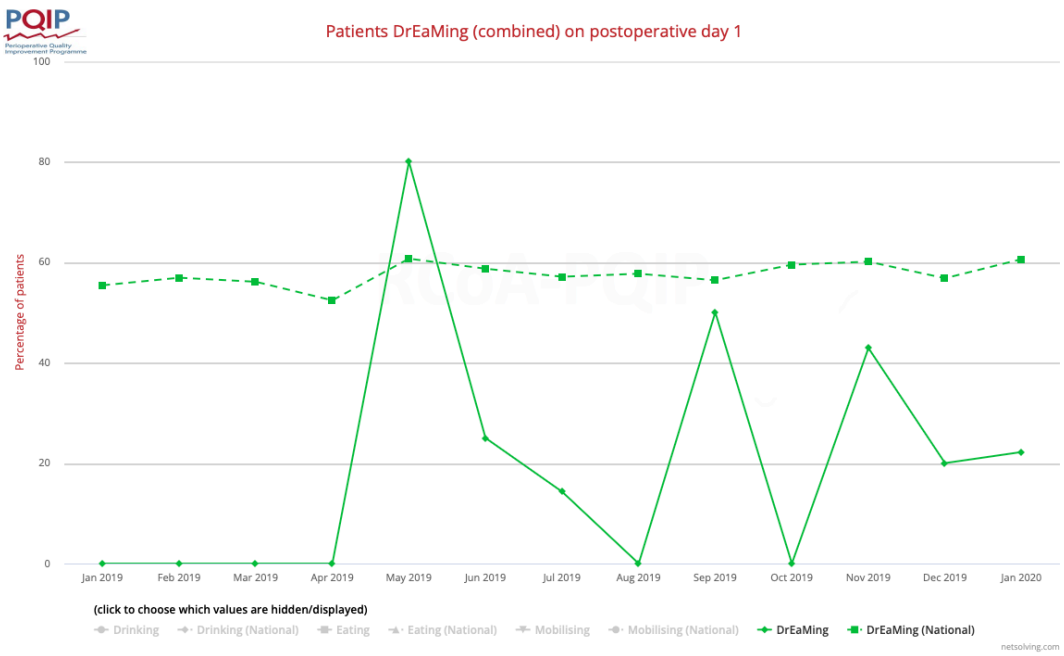


Figure 4-3: Example DrEaMing run chart added to pomVLAD dashboard in February 2019

Figure 4-4: Example run charts added to pomVLAD dashboard in February 2019 showing process measure adherence

4.3.5.7 Adherence to intervention

After launch, the pomVLAD dashboard was available to local investigators at all intervention sites throughout the study period. Access to the PQIP webtool was password protected with login details provided at site registration to the PQIP study. Webtool access was required to view the pomVLAD dashboard.

Attendance at one of the breakout sessions or a site initiation telephone call was suggested but not mandatory. At least one representative from all intervention sites received a form of site initiation (either attendance at a breakout session or individual telephone setup call). Engagement with the dashboard and any associated quality improvement activity following implementation was voluntary.

Investigators at sites allocated to the qualitative analysis interview group received email contact prior to interviews to enable arrangements to be made to conduct them. This contact may have served as a prompt to engage with the dashboard and local data although no suggestion to this effect was made. Intervention sites not allocated to the interview group did not receive any additional prompts to engage during the intervention period.

4.3.6 Data sources

4.3.6.1 Data collection and processing

Patient data were collected prospectively as part of the national PQIP study by local site investigators, typically research nurses and clinicians.[26] Non-identifiable patient level data were exported from the main study database on 10th September 2019. Records locked by local investigators, and therefore considered to be complete, were included in the analysis. Amendments were made to the main PQIP study protocol during the pomVLAD study period, but these had no material impact on the conduct of this study.[214] The PQIP patient study was approved by the Health Research Authority (London-Surrey Research Ethics Committee REC reference number: 16/LO/1827) and this analysis was approved by the PQIP National Project Team.

For the rapid feedback evaluation, the named lead research nurse(s), consultant anaesthetist and consultant surgeon for the PQIP study at each interview site were contacted by email and invited to take part in the qualitative evaluation of the study. If no response was received to the initial email, a follow-up email was sent within 4 weeks. The email provided background on the participant information sheet and consent form (Appendix D-4 and Appendix D-5). Written consent was obtained from participants prior to interview. The qualitative evaluation was classified as a service evaluation by the Joint Research Office, University College London and was therefore registered with the Department of Anaesthesia and Perioperative Medicine, University College London Hospitals NHS Trust.

Semi-structured interviews, following a prespecified topic guide, were conducted by a single interviewer (JB) (Appendix D-6). Interviews were conducted remotely by telephone and digitally recorded with participant consent. Files were transferred for transcription by Essential Secretary using an encrypted File Transfer Service with 256-bit SSL encryption. Recordings were stored securely on a password protected folder in a secured UCL shared drive for analysis. Participant identifier codes were stored in a password-protected file on a secured UCL shared drive which only named team members had access to via password-protected computers at the

Centre for Perioperative Medicine, Division of Surgery & Interventional Science, UCL. Data were deleted from the digital audio recording device once transferred.

Identifiable electronic data were held on UCL file servers, in shared or in personal folders. Anonymised interview transcripts, data for the documentary analysis, and non-identifiable quantitative data were stored for analysis on password-protected encrypted hard drives.

Anonymised interview transcripts were imported into qualitative analysis software, NVivo.[215] Transcripts were coded using deductive and inductive approaches. Deductive coding was based on the programme theory underpinning the pomVLAD intervention (Figure 4-1), the research questions identified, and barriers and facilitators to quality improvement activity identified in Chapter 2. Framework analysis was then used to identify patterns in the data and develop themes.[216]

Table 4-6: Coding categories used for transcript analysis

Parent code	Child code
Dashboard feedback	Experience of using the dashboard
	Participants perception of data feedback through dashboards compared to reports
	Recommendations for improving the dashboard
	Statistical considerations related to dashboard
Local engagement activity	Strategies to maintain local engagement with data collection and/or QI activity
	Triggers for local sites to engage with their data and/or dashboard
Perception of open compared to closed reporting of results	
Participants perception of their role in quality improvement	
Programme theory	Building engagement (central to local level)
	Identification of QI opportunities through the dashboard
	Monitoring of QI implementation
	Participant perception of the pomVLAD project
	Participant perception of the process measure recommendations
	QI activity related to the dashboard
	Understanding of data quality and completion
Barriers to local QI activity	
Facilitators to local QI activity	

4.3.6.2 Data quality and completeness

Qualitative analysis

Twenty-seven local investigators were contacted across the 10 NHS sites included in the rapid feedback evaluation. Three sites did not have a named surgical lead or contact details were not available for them. First round interviews were conducted with at least one local investigator at nine out of the 10 identified sites (five control sites, four intervention sites). Local investigators at one intervention site did not participate despite follow-up email contact (site pseudonym Col du Galibier).

Second round interviews included investigators from six out of the nine sites that took part in the first round (three control sites, three intervention sites). Seven participants who took part in the baseline interviews did not participate in the second round (two participants based at control sites, five participants based at intervention sites). Figure 4-5 and Figure 4-6 show the local investigators contacted at each interview site and those taking part in baseline and follow-up the interviews at each site.

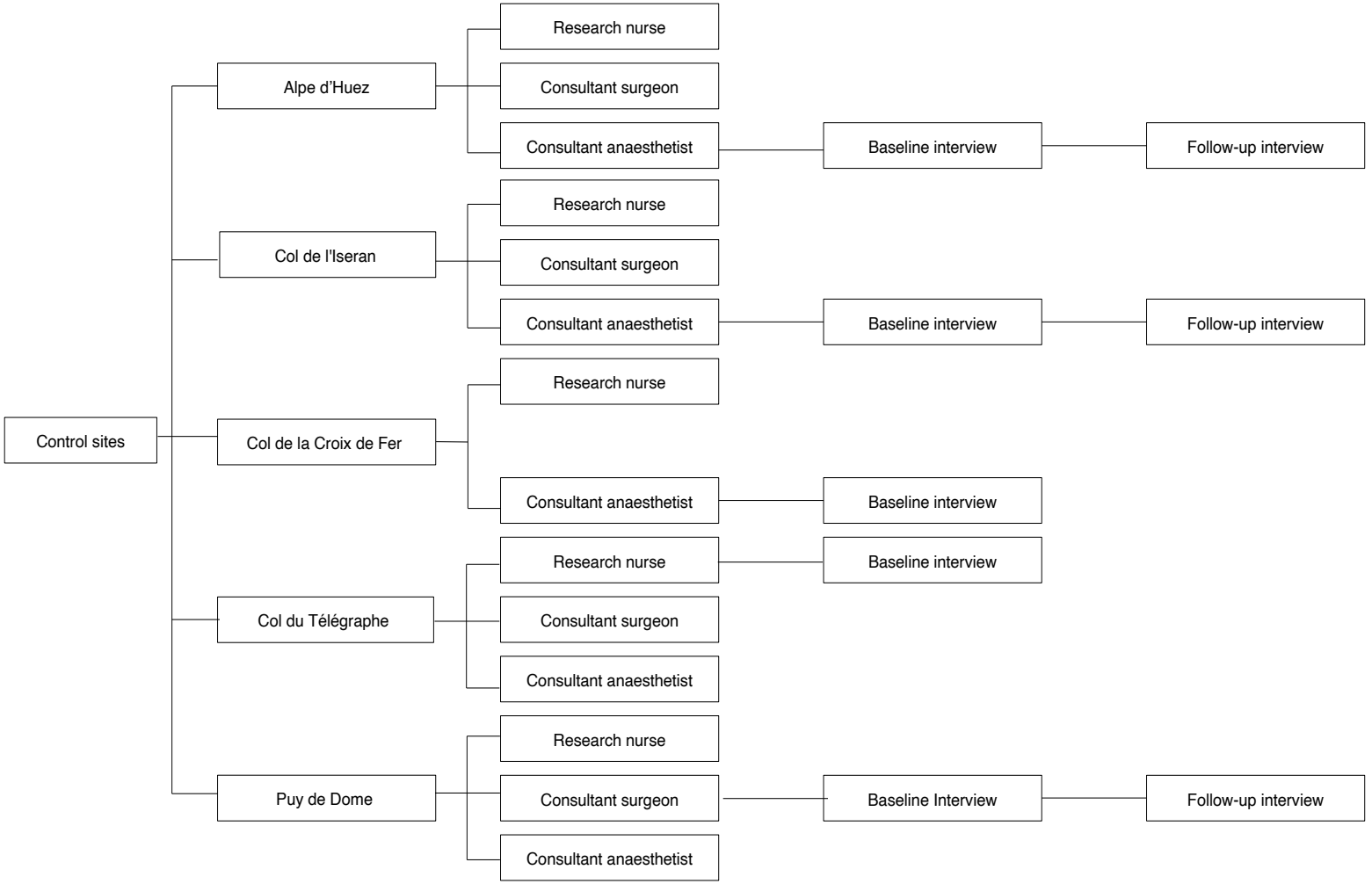


Figure 4-5: Control site interview sample construction showing individuals contacted at each site and those taking part

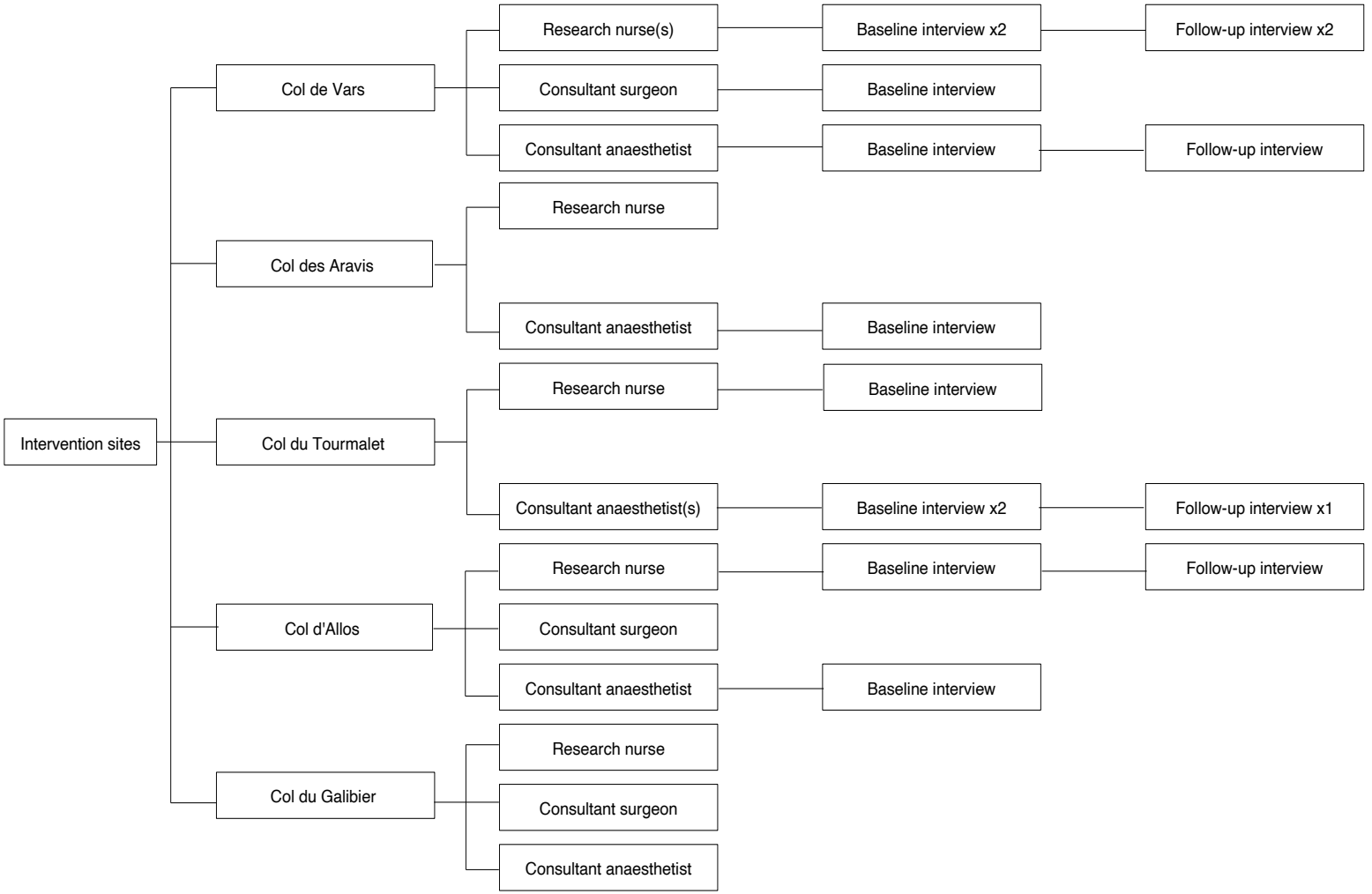


Figure 4-6: Intervention site interview sample construction showing participants contacted at each site and those taking part

4.3.7 Unit of analysis

Intervention effects in the CITS analysis were analysed at the hospital level with data aggregated to calendar month.

Qualitative interview data were coded individually and analysed at the participant level in the framework analysis.

4.3.8 Statistical methods

4.3.8.1 Impact regression model

A multilevel segmented linear regression model was used to estimate the impact of the intervention on primary and secondary outcomes.[203,204] Fixed effects were estimated using the equation below.[217,218] The model also incorporated a random intercept term for hospital site.

$$y = \alpha + \beta_1 T + \beta_2 X + \beta_3 XT + \beta_4 Z + \beta_5 ZT + \beta_6 ZX + \beta_7 ZXT + \varepsilon$$

Where:

T = Time from start of study period (months)
X = Study phase (pre- or post-implementation)
XT = Time after implementation (months)
Z = Intervention or control group
ZT = time from implementation for intervention group (months) and 0 for control group
ZX = study phase for intervention group (pre- or post-implementation) and 0 for control
ZXT = time after implementation for intervention group (months) and 0 for control

Interpretation of coefficients:

β_1 = Control group pre-implementation trend
β_2 = Control group post-implementation level change
β_3 = Control group post-implementation trend change
β_4 = Intervention/Control pre-implementation level difference
β_5 = Intervention/Control pre-implementation trend difference
β_6 = Intervention/Control post-implementation level difference
β_7 = Intervention/Control post-implementation change in slope difference pre- to post-implementation

The impact model allowed for a step change at the specified time of implementation (defined as 1st July 2018) in the intervention group compared to the control group (β_6) as well as estimating the difference in slope change following implementation between the two groups (β_7). The main impact model did not allow for a time lag or transition phase. The presence of a time-lag is investigated in a sensitivity analysis, allowing a three-month lag following implementation, setting the break point on 1st October 2018.

A multilevel segmented linear regression model is appropriate to assess the continuous primary and secondary outcomes at the hospital level. Risk-adjusted morbidity ratio was aggregated to month. Secondary outcomes measures, which are processes of care, were converted to percentage and aggregated to month.

4.3.8.2 Sensitivity and secondary analyses

Subgroup analyses were performed to assess primary and secondary outcome measures in hospital sites identified in the qualitative analysis as 'higher engagement, higher compliance', 'lower engagement, higher compliance', and 'lower engagement, lower compliance' compared to control sites.

A sensitivity analysis was performed to explore the effect of a lag period of 3-months following implementation. The intervention time-point was defined as 1st October in this analysis.

Recruitment at two control sites (Alpe d'Huez and Col du Soulor) reduced significantly during the implementation period. A second sensitivity analysis explored the effect of excluding these two sites from the analysis.

4.3.8.3 Handling of missing data

Cases with missing primary outcome data were excluded, incorporating both those cases who withdrew from the study before postoperative day 7 and those with a calculated length of stay >7 days but no POMS outcomes recorded. In keeping with model development, patients discharged before postoperative day 7 were assumed to have no POMS-defined morbidity. Those who died before postoperative day 7 were assigned to have morbidity at postoperative day 7. Cases with missing predictor data required for risk-adjustment of postoperative morbidity were also excluded from the analysis. Secondary outcome measures were mandatory fields within the study database.

4.3.8.4 Statistical software

Statistical analyses were performed using R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria), with the following external packages enabled: plyr, zoo, extrafont, scales, DescTools, nmle, lme4, tidyverse. Means and standard deviations are reported for normally distributed data. For non-normally distributed data, medians and inter-quartile ranges (IQRs) are reported.

Qualitative analysis was performed using NVivo 20 qualitative data analysis software.[215]

Chapter 5 Results

5.1 Quantitative analysis

5.1.1 Study population

5.1.1.1 Patient level study population

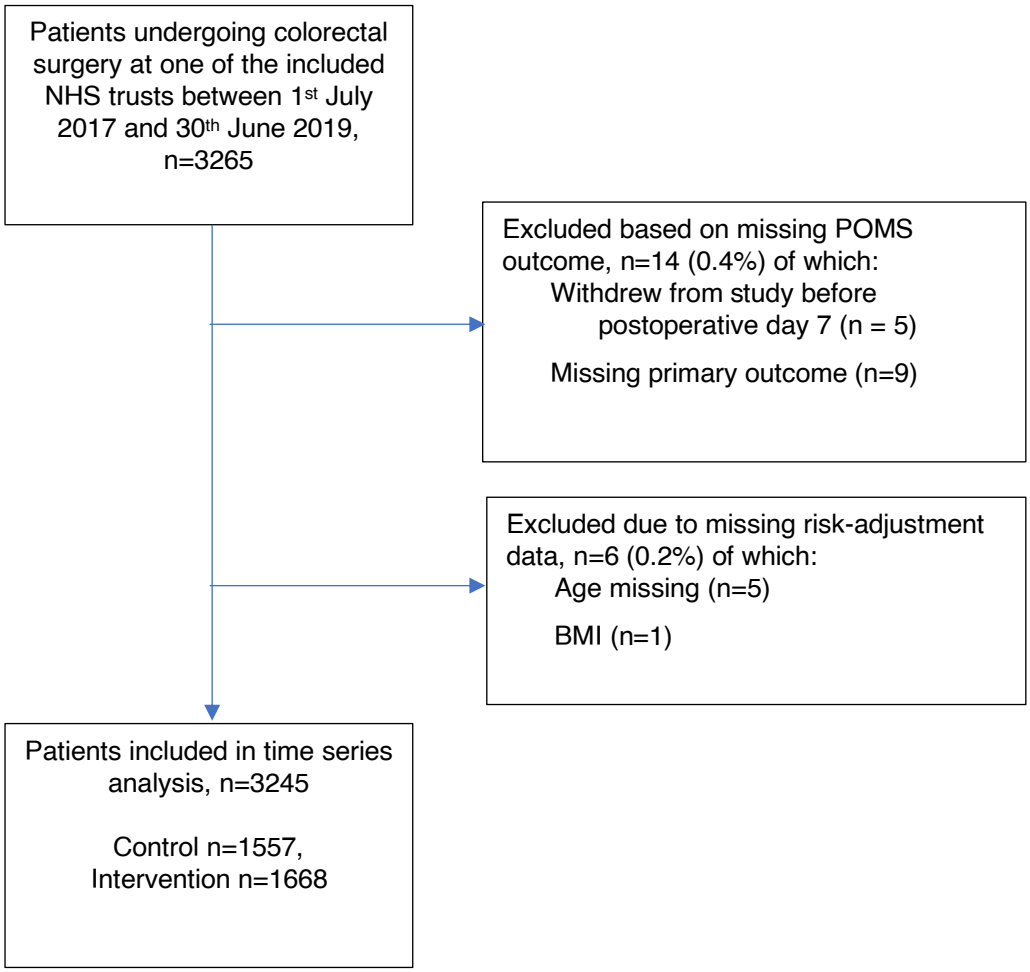


Figure 5-1: Time series analysis sample construction

Patient level baseline characteristics

A total of 3245 patients were included in the time series analysis, of which 1948 were recruited in the 12-months prior to implementation, and a further 1297 patients in the post-implementation period. Table 5-1 shows baseline patient characteristics prior to implementation in the two groups including variables used for risk-adjustment of postoperative morbidity. All patients were included in assessment of the primary and secondary outcome except for the provision of carbohydrate drinks prior to surgery. Patients diagnosed with diabetes (425/3245, 13.1%) were excluded from the analysis of preoperative carbohydrate delivery due to the limited evidence to support them in this group and concerns about their effect on blood glucose.[219] Cases where the delivery of carbohydrate drinks to non-diabetic patients was documented as unknown were classified as not receiving them (347/2820, 12.3%).

There was no significant difference in the baseline incidence of unadjusted postoperative morbidity defined by the POMSmajor between the two groups (Table 5-2). The incidence of Clavien-Dindo grade II and above complications was higher in the intervention group (28.6% vs. 23.6%, p=0.018)

Table 5-1: Baseline patient characteristics by group assignment

	Control	Intervention	p-value
Age	66 [54-74]	66 [56-74]	0.765
Sex			
Female	401	448	0.757
Male	528	571	
ASA grade			
I	83	137	<0.001
II	555	634	
III	273	240	
IV/V	18	8	
Body Mass Index ($kg.m^{2\ -1}$)	27.75 (+/- 5.73)	27.57 (+/- 5.41)	0.478
Heart rate (beats per minute)	79 (+/- 14)	77 (+/- 13)	0.045
Systolic blood pressure (mmHg)	135 (+/- 20)	134 (+/- 19)	0.282
Respiratory history			
No respiratory history	805	868	0.166
Dyspnoea on exertion	86	118	
Dyspnoea on light exertion/at rest	38	33	
Number of operations in last 30 days			
One	897	996	0.149
Two or more	32	23	
AXA surgical severity			
Minor/Intermediate/Major	16	13	0.520
Xmajor	533	569	
Complex	380	437	
Estimated risk of postoperative major morbidity*	0.171 [0.128-0.230]	0.160 [0.120-0.223]	0.068

*Notes: ASA – American Society of Anesthesiologists Physiological Score; AXA, AXA health classification; values are median [IQR] and mean (+/- standard deviation); significance testing for difference in means – paired t-test; significance testing for difference in medians – Wilcoxon rank sum test; significance testing for count data - Pearson's Chi-squared test; *risk-adjustment model derived in Chapter 3 used for risk estimation*

Table 5-2: Baseline unadjusted postoperative morbidity by group assignment

	Control n (%)	Intervention n (%)	p-value
Patients recruited	929 (47.7)	1019 (52.3)	0.025
Patients remaining in hospital at postoperative day 7	407 (43.8)	459 (45.0)	0.616
Patients with POMS-defined morbidity (including death on or before postoperative day 7)	241 (25.9)	290 (28.5)	0.232
Patients with POMSmajor-defined morbidity (including death on or before postoperative day 7)	158 (17.0)	176 (17.3)	0.925
Patients who died within 7 days of surgery	4 (0.4)	4 (0.4)	1.000
Patients with Clavien-Dindo grade II or above complication during hospital admission	220 (23.7)	294 (28.9)	0.011

Notes: Significance testing performed using two-proportion z-test.

5.1.1.2 Hospital level characteristics

Table 5-3 shows hospital characteristics by group assignment. There were no significant differences in hospital bed capacity, the number of surgical wards, the number of surgical ward beds, or critical care facilities available between intervention and control groups. The was a trend for intervention hospitals to have a greater number of general surgical wards and general surgical beds, but these did not reach statistical significance.

Table 5-3: Hospital characteristics by group assignment

	Control (n=10)	Intervention (n=10)	p value
Trust wide hospital beds[220]	742 [678-843]	851 [591-1022]	0.631
Hospital beds	625 [550-785]	675 [510-815]	0.970
General surgical beds	145 [123-192]	226 [125-258]	0.130
General surgical wards	5 [5-7]	8 [6-10]	0.075
Critical Care Facilities present in hospital	10/10	9/10	1.000
Critical care beds	18 [14-21]	22 [16-29]	0.520
Number of ventilated critical care beds	13 [8-17]	11 [10-14]	0.849
Enhanced care ward present in hospital	5/10	5/10	1.000
Enhanced care beds	4 [4-8]	4 [2-8]	0.915
Hospital provides tertiary level services	7/10	6/10	1.000
Emergency Department in hospital	9/10	10/10	1.000

Notes: Trust wide beds calculated from published data.[220] These figures may include additional hospital sites within a trust not included in this study. All other hospital level data shown in this table was collected by the SNAP-2 EPICCS study.[221,222] Figures are either median [IQR] or count. P-values for differences between groups are calculated using either the Wilcoxon rank sum test or the two-proportion z-test. Services considered as tertiary level included: bariatric surgery, bone marrow transplant, burns services, cardiothoracic, complex colorectal, complex cardiology, extracorporeal membrane oxygenation (ECMO), hepatobiliary, hyperacute stroke services, major trauma, maxillo-facial surgery, neurosurgery, transplant services, upper gastrointestinal surgery, vascular, and complex orthopaedic surgery.

5.1.1.3 Site recruitment throughout study period

All sites recruited patients during the pre- and post-implementation period. Seven sites (four control, three intervention) had a ≥50% fall in patient recruitment in the post-implementation period compared to pre-, namely, Col du Tourmalet, Col d'Izoard, Col d'Allos, Alpe d'Huez, Col du Soulor, Col du Glandon, and Puy de Dome. Only one site, Col de l'Iseran, increased recruitment in the post-implementation period (Figure 5-2).

The general reduction in recruitment may have been the result of both local challenges with resource available to support PQIP patient recruitment and the addition of new surgical specialties to the PQIP study in 2018 (namely the inclusion of burns/plastics, gynaecology, orthopaedics, vascular, and spinal surgery). The addition of these specialties resulted in a reduction in the proportion of colorectal patients recruited to PQIP at the national level (see Figure 5-3, reproduced from the Second national PQIP report[223]), which in turn may have impacted on the number of colorectal patients recruited during the implementation phase of the pomVLAD study. This is likely to have had a greater impact in pomVLAD sites given they were selected for inclusion based on their recruitment of colorectal patients prior to this change.

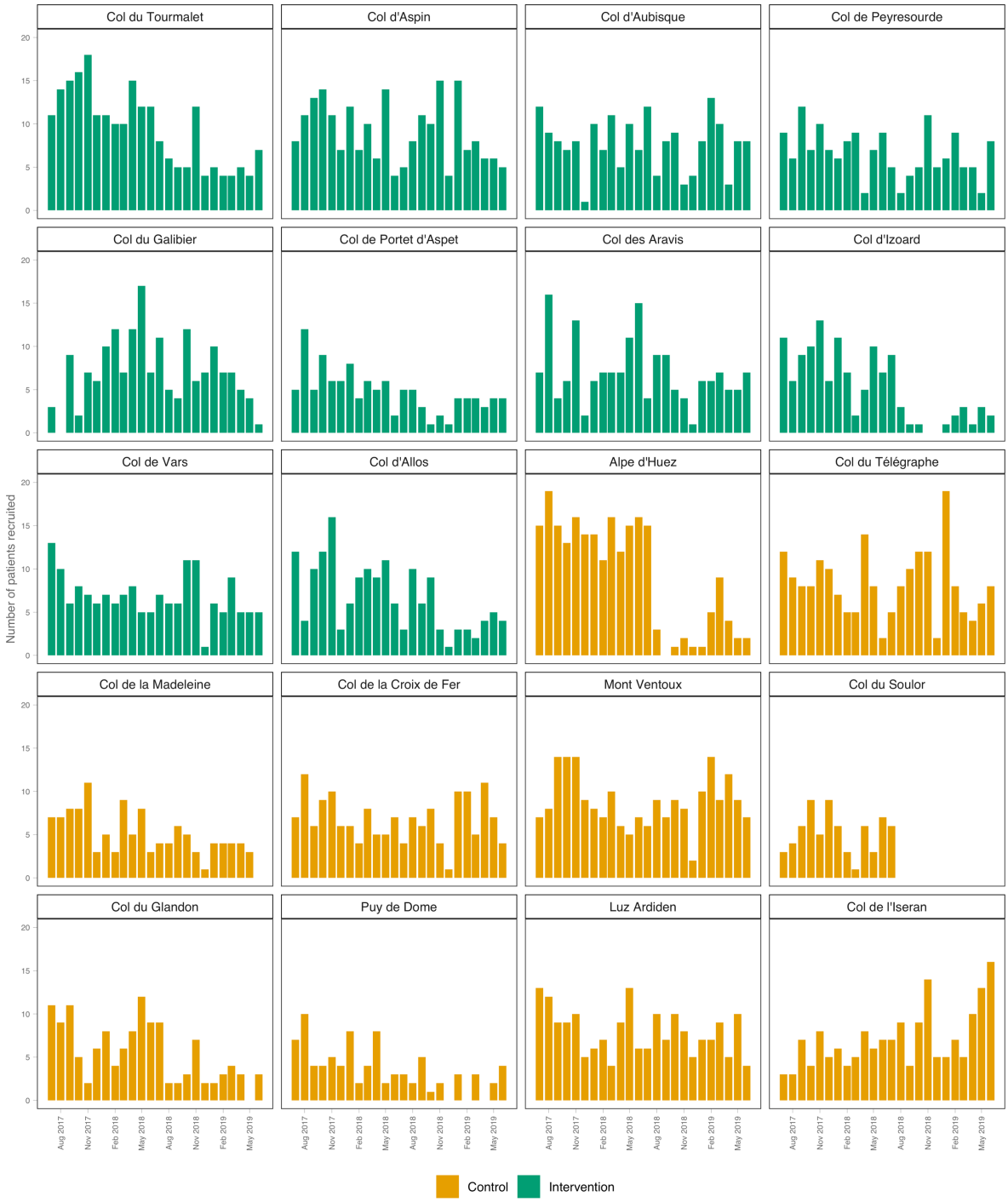


Figure 5-2: Site recruitment by month throughout entire study period

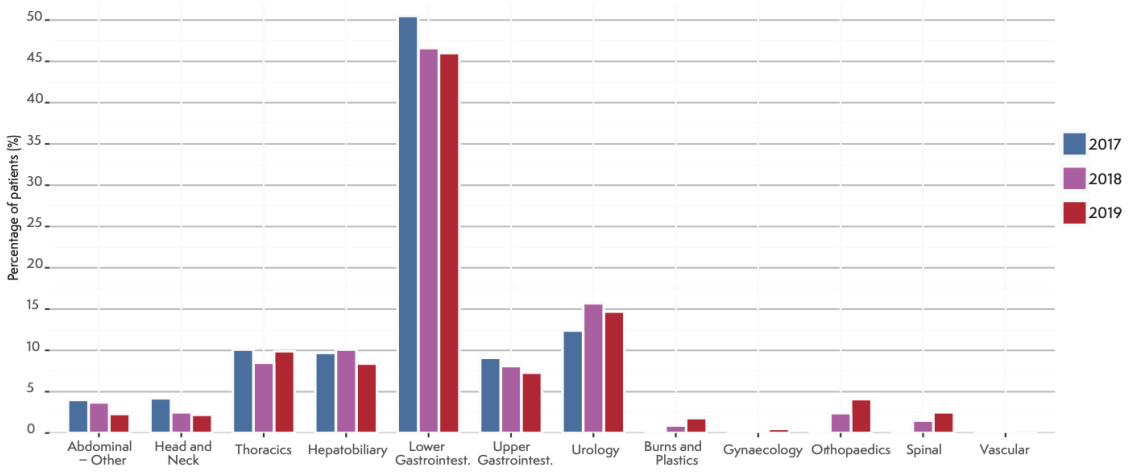


Figure 5-3: National patient recruitment to the PQIP study by year[223]

5.1.2 Outcomes and estimation

Primary outcome

The risk-adjusted postoperative morbidity ratio (observed/expected morbidity ratio, PMR) defined by POMSmajor was 0.915 and 0.953 in the control and intervention groups respectively pre-implementation. Following implementation, the PMR was 0.872 and 0.997 in the control and intervention groups respectively. Figure 5-4 shows the risk-adjusted time series morbidity data with pre- and post-implementation regression lines for the control and intervention groups.

Following implementation there was a non-significant reduction in PMR of 0.391 (95% CI -1.092-0.306). There was an increase in the slope of 0.065 (95% CI -0.036-0.166) following implementation, representing a non-significant trend back towards the baseline PMR (Table 5-4).

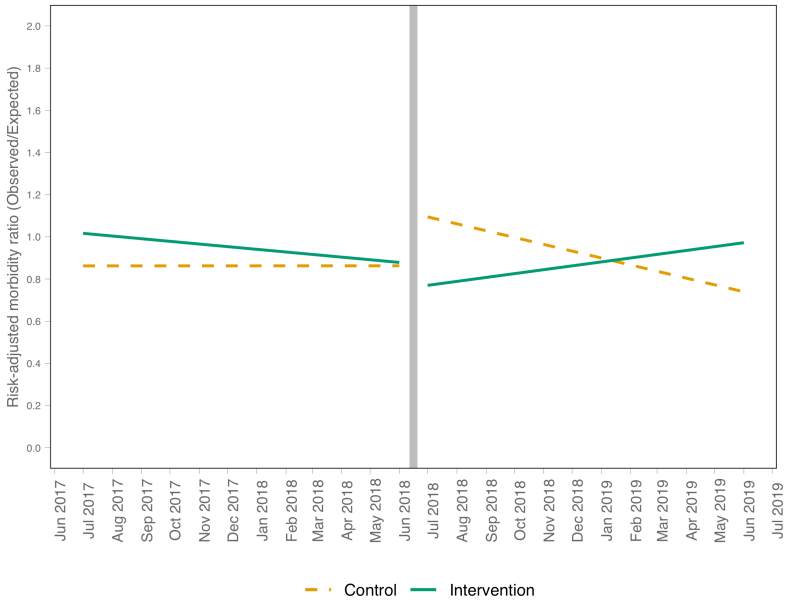


Figure 5-4: Regression trends in PMR by site assignment throughout the study period

Notes: PMR, risk-adjusted postoperative morbidity ratio; Linear regression lines show the control and intervention groups pre- and post-implementation. The vertical grey line represents the intervention time-point. Table 5-4 shows the estimated level and trend change in risk-adjusted postoperative morbidity from the multilevel CITS model.

Table 5-4: Multilevel controlled interrupted time series regression model estimates of level and slope change in PMR following implementation

	Parameter estimated	Estimate (95% CI)	p-value
Risk-adjusted postoperative morbidity ratio	Intervention/Control post-implementation level difference	-0.391 (-1.092-0.306)	0.276
	Intervention/Control change in slope difference Pre- to Post-implementation	0.065 (-0.036-0.166)	0.213

Notes: PMR, risk-adjusted postoperative morbidity ratio

Secondary outcomes

Compliance with process measures recommended in pomVLAD intervention

After correction for multiple testing there were no statistically significant level or trend changes in the postimplementation period. Figure 5-5 shows the regression lines of compliance in the intervention group compared to the control group

following implementation. Table 5-5 shows the estimated change in level and slope of compliance in the intervention group compared to the control group following implementation.

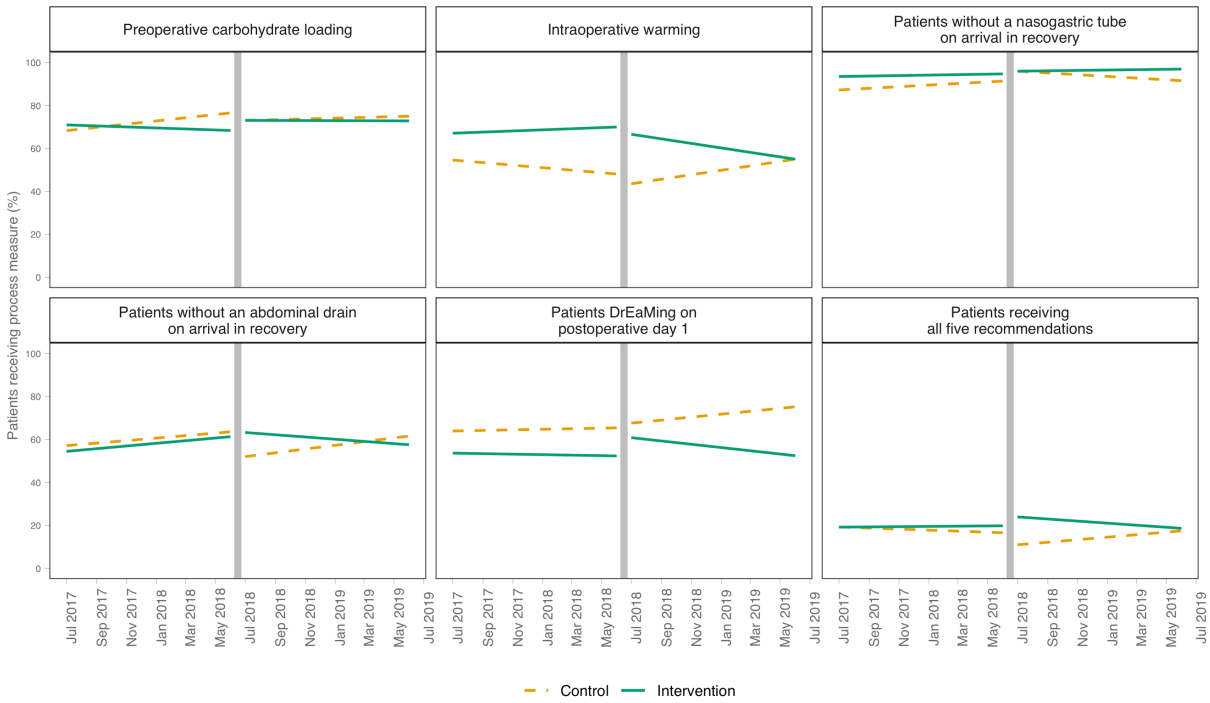


Figure 5-5: Regression trends in monthly adherence to process measures by site assignment throughout the study period

Notes: Linear regression lines show the control and intervention groups pre- and post-implementation. The vertical grey line represents the intervention time-point.

Table 5-5: Multilevel controlled interrupted time series regression model estimates of level and slope change of secondary outcomes following implementation

Variable	Parameter estimated	Estimate (95% CI)	p-value (Bonferroni correction)
Carbohydrate loading	Intervention/Control post-implementation level difference	10.684 (-6.984-28.325)	0.238 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	0.720 (-1.825-3.264)	0.582 (1.000)
Intraoperative warming	Intervention/Control post-implementation level difference	4.470 (-11.037 – 19.969)	0.574 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-2.913 (-5.153 - -0.672)	0.012 (0.081)
Avoidance of nasogastric tubes	Intervention/Control post-implementation level difference	-2.566 (-10.394 – 5.206)	0.521 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	0.850 (-0.278 – 1.976)	0.116 (0.992)
Avoidance of abdominal drains	Intervention/Control post-implementation level difference	15.265 (-1.000 – 31.520)	0.068 (0.474)
	Intervention/Control change in slope difference Pre- to Post-implementation	-1.613 (-3.960 – 0.740)	0.181 (1.000)
DrEaMing within 24 hours of surgery ending	Intervention/Control post-implementation level difference	11.200 (-5.325 – 27.643)	0.186 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-0.753 (-3.184 – 1.581)	0.514 (1.000)
Patients receiving all recommendations	Intervention/Control post-implementation level difference	12.425 (-0.851 – 25.672)	0.068 (0.478)
	Intervention/Control change in slope difference Pre- to Post-implementation	-1.350 (-3.267 – 0.566)	0.170 (1.000)

Records locked with 14 days of hospital discharge

There was a non- significant reduction of 9.3% (95% CI -28.887-10.282) in the percentage of case records locked within 14-days of hospital discharge in the intervention group compared to the control group following implementation. The increase in the slope trend of 1.1% (95% CI -1.720-3.940) was also not statistically significant (Table 5-6). Figure 5-6 shows the regression lines for the percentage of

records locked within 14-days of hospital discharge for the control and intervention groups in the pre- and post-implementation periods.

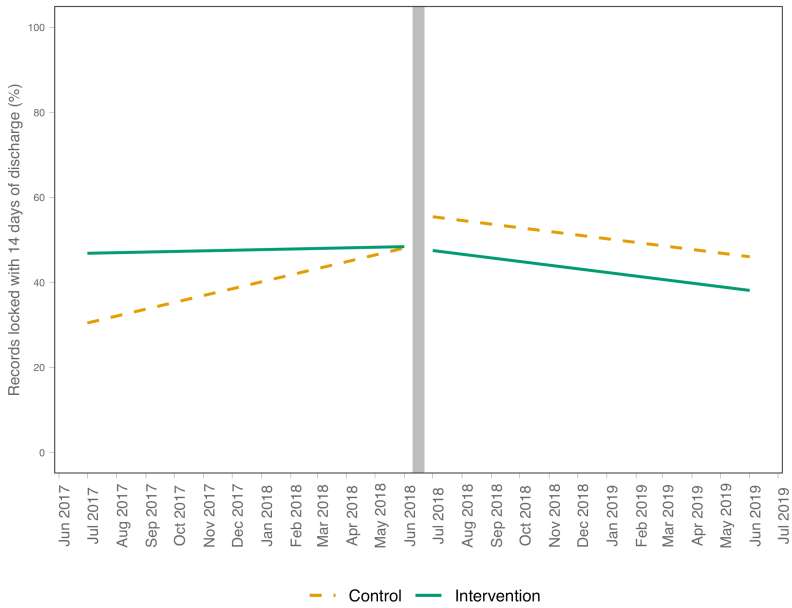


Figure 5-6: Regression trends in records locked within 14 days of hospital discharge by site assignment throughout the study period

Notes: Linear regression lines show the control and intervention groups pre- and post-implementation. The vertical grey line represents the intervention time-point. Table 5-6 shows the estimated level and trend change in the percentage of records locked within 14-days of hospital discharge from the multilevel CITS model.

Table 5-6: Multilevel controlled interrupted time series regression model estimates of level and slope change of records locked within 14-days of hospital discharge following implementation

Variable	Parameter estimated	Estimate (95% CI)	p-value (Bonferroni correction)
Records locked within 14 days	Intervention/Control post-implementation level difference	-9.318 (-28.887-10.282)	0.354 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	1.109 (-1.720-3.940)	0.445 (1.000)

5.1.3 Supplementary and sensitivity analyses

Sensitivity analysis – inclusion of 3-month time lag prior to break point

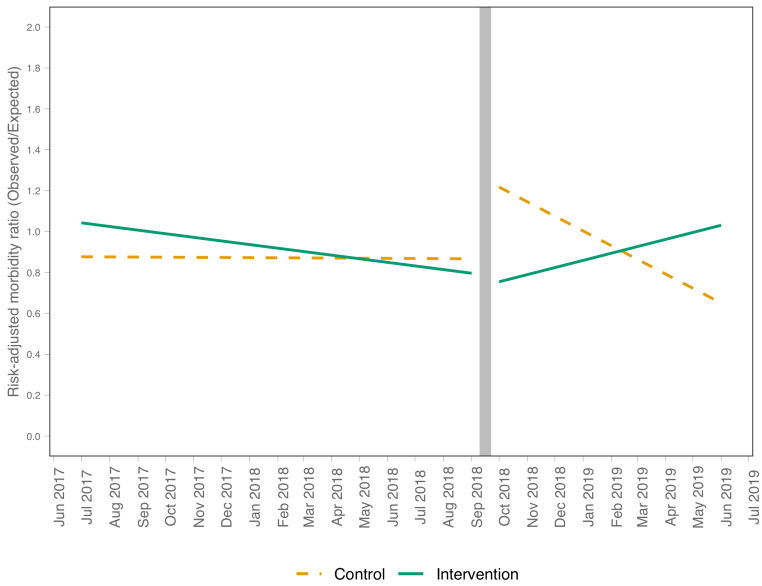


Figure 5-7: Trend in risk-adjusted morbidity ratio by site assignment pre- and post-pomVLAD implementation after allowing a 3-month lag period from implementation

Table 5-7: Estimated change in level and slope of risk-adjusted postoperative morbidity in intervention group compared to control group following implementation

	Parameter estimated	Estimate (95% CI)	p value
Risk-adjusted postoperative morbidity ratio	Intervention/Control post-implementation level difference	-0.248 (-1.040 – 0.540)	0.540
	Intervention/Control change in slope difference Pre- to Post-implementation	0.114 (-0.016 – 0.245)	0.088

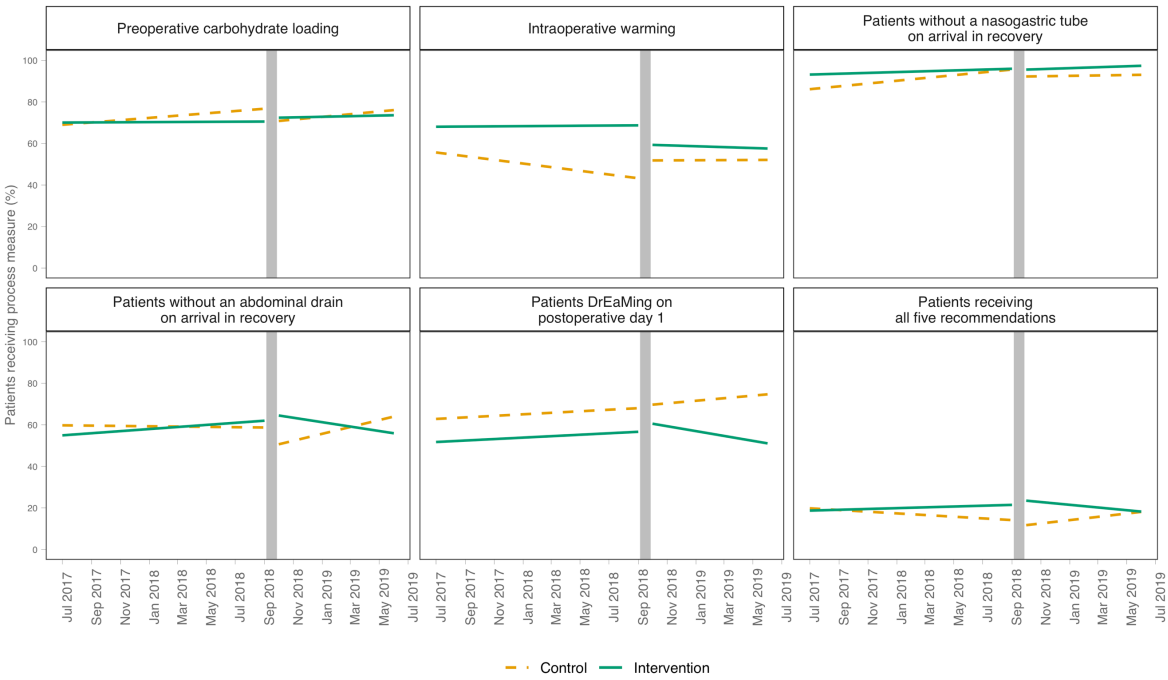


Figure 5-8: Trends in monthly adherence to process measures by site assignment pre- and post- pomVLAD implementation after allowing a 3-month lag period from implementation

Table 5-8: Estimated change in level and slope of secondary outcome measure delivery in intervention group compared to control group alowing a 3-month lag time from implementation

Variable	Parameter estimated	Estimate (95% CI)	p-value (Bonferroni correction)
Carbohydrate loading	Intervention/Control post-implementation level difference	9.808 (-9.685 – 29.250)	0.800 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-0.388 (-3.584 – 2.815)	0.502 (0.711)
Intraoperative warming	Intervention/Control post-implementation level difference	-8.656 (-25.570 – 8.244)	0.318 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-1.941 (-4.725 – 0.844)	0.175 (1.000)
Avoidance of nasogastric tubes	Intervention/Control post-implementation level difference	4.345 (-4.137 – 12.788)	0.317 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	0.445 (-0.946 – 1.841)	0.533 (0.449)
Avoidance of abdominal drains	Intervention/Control post-implementation level difference	7.775 (-10.096 – 25.659)	0.397 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-3.197 (-6.138 - -0.248)	0.035 (0.244)
DrEaMing within 24 hours of surgery ending	Intervention/Control post-implementation level difference	5.865 (-12.346 – 24.017)	0.529 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	-1.851 (-4.845 – 1.144)	0.228 (1.000)
Patients receiving all recommendations	Intervention/Control post-implementation level difference	11.877 (-2.767 – 26.490)	0.114 (0.797)
	Intervention/Control change in slope difference Pre- to Post-implementation	-2.663 (-5.070 - -0.251)	0.032 (0.221)

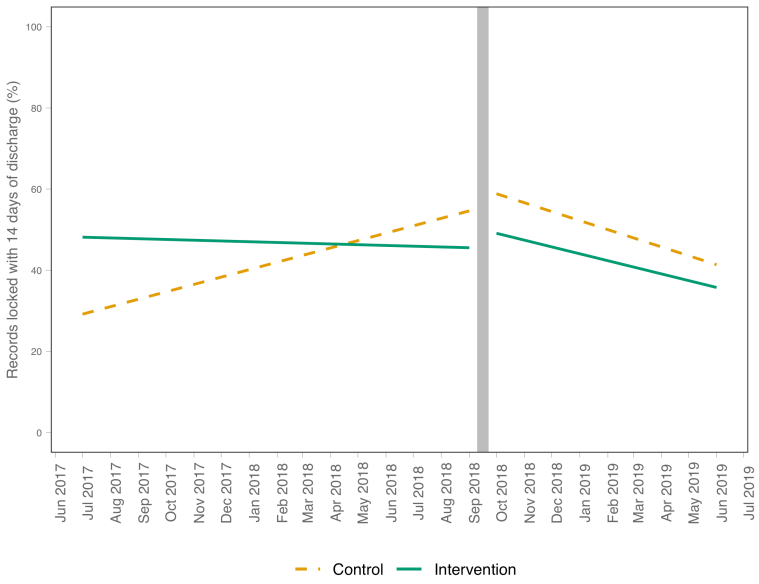


Figure 5-9: Records locked within 14 days of hospital discharge by site assignment pre- and post- pomVLAD implementation after allowing a 3-month lag period after implementation

Table 5-9: Estimated change in level and slope of records locked within 14 days of hospital discharge in intervention group compared to control group following implementation

	Parameter estimated	Estimate (95% CI)	p-value (Bonferroni correction)
Records locked within 14 days	Intervention/Control post-implementation level difference	-6.145 (-28.887 – 10.282)	0.572 (1.000)
	Intervention/Control change in slope difference Pre- to Post-implementation	2.938 (-1.720 – 3.940)	0.102 (0.711)

5.2 Qualitative analysis

5.2.1 Interview participant characteristics

Twenty-three interviews were conducted in total, including 16 individuals from nine hospital sites. One site, Col du Galibier, did not take part in interviews despite contact inviting their local investigators. Table 5-10 shows which local investigators participated in the qualitative analysis. Only one consultant surgeon took part in each of the control and intervention groups. Four research nurses took part in the interview process in the intervention group compared to only one from the control group.

Table 5-10: Interview participant characteristics

	Control group	Intervention group	p value
Research nurse	1	4	0.223
Consultant anaesthetist	3	5	
Consultant surgeon	1	1	
Length of interview in minutes; Median [IQR]	29 [25-35]	37 [28-44]	0.174
Number of participants taking part in second interview	3/5	5/10	1.000
Total interviews (n)	8	15	0.144

5.2.2 Programme theory

Figure 4-1 shows the theory of change for the pomVLAD intervention which aimed to define how positive change in primary and secondary outcome measure would be delivered.

Three key short-term outcomes of quality improvement activity, an increase in data completion/quality, and engagement with the dashboard were identified that would lead to increased compliance with the secondary outcome measures and a reduction in risk-adjusted postoperative morbidity.

Initial contact with local sites built interest in the dashboard and the data it displays. Local investigators reported accessing the dashboard around the time of the national PQIP meeting when the project was first introduced.

> *"When it all came after the meeting, I did have a look at it before consenting as well as after the meeting. But all the information is there, but we haven't really gone ahead and looked at it and then seen what can be done to change things which need to be. So I think it will be more of allocating some time." INT 6 (Anaesthetist, Intervention site)*

Site initiation phone calls also helped to build initial engagement with local data.

> *"The most obvious local prompt [to engage] has been the pomVLAD stuff. So that's what the last couple of weeks has made me look at the dashboard a lot more, because I was looking at that dashboard, if you see what I mean…And frankly, you know, with our numbers that we're recruiting, if we had say two to three a week, you know, you need a couple of months to go by before anything really changes anyway." INT 14 (Anaesthetist, Intervention site)*

> *"I think it's pretty good; I had a quick look before this phone call. I think the only thing I can't find in the dashboard is our total numbers, there's lots of percentages but no, so that's a percentage of what." INT 12 (Anaesthetist, Intervention site)*

Initial engagement with the dashboard was not always sustained however and some local investigators reported accessing it infrequently during their normal day to day work. Additional contact with sites when arranging interviews stimulated further engagement.

> *"No, hand on heart, not a lot. I mean I've really only properly looked at [the pomVLAD dashboard] … when I was aware it was available I looked at it out of interest…Then when you requested this phone call I looked at it much more seriously because I thought well I'd better know what I'm talking about slightly so I guess, if I'm honest, no I haven't looked at it on a regular basis." INT 4 (Research nurse, Intervention site)*

Local investigators described limited engagement from senior management which reduced the impact of PQIP and pomVLAD. Participants acknowledged other national studies took time to become established and gain engagement from senior management. Clinicians felt this engagement was required to successfully deliver change projects.

> *"…on other projects and certainly around NELA there is now much better understanding of the importance of the data collection and the use of it. And the … I think it's now part of the everyday language of the hospital… right through to the chief exec. So I think it's that … in some ways I mean PQIP trying to up its own profile with … in the senior executive teams of acute trusts would be … this is all really … it's there, it exists and it's really important data that you can use for quality improvement work in your hospital. But it does require, not just clinician buy in, but it requires general management allocating time to it and using it." INT 3 (Anaesthetist, Control site)*

> *"There is a little bit of futile cycling where we all discuss it, agree what needs to be done but maybe not all the people who need to be in the room are there. And it is actually quite, you almost need the improvement teams and the general managers to come to you to join that final piece together so that actually a QI Project can work." INT 3 (Anaesthetist, Control site)*

Some participants were frustrated the link between data collection for research and subsequent implementation of quality improvement was not understood by senior management, and that their role was not fully appreciated.

> *"I don't know. I mean it would be great if it would be ... I mean what I can see happening is, for example if as a result of PQIP and the data collected and quality improvement initiatives were identified and implemented and it saves the trust some money and they were obviously then aware of that they'd be rubbing their hands with glee and congratulating, quite rightly, the team who had driven that change. But whether they would ever connect that with the fact that this data came to light because of a research study that we were running is very unlikely.*

> *They just don't put the two ends together, they wouldn't match the ... they wouldn't." INT 4 (Research nurse, Intervention site)*

Despite this, clinicians saw the value of a clinical dataset which they felt offers better data quality than administrative datasets, which are frequently used by organisational analysts.

> *"Because they spend huge amounts of time in general management analysing data that's collected from HES and various other sources within the hospital. But some of this clinical data is actually far more robust than the data that they analyse because the data collection is that much more careful and is done often by clinicians or in this case the research nurses. So I think the quality of the data is good." INT 3 (Anaesthetist, Control site)*

Local investigators felt contact between the central project team and senior management highlight the importance of their work.

> *"… it makes me wonder whether we can ask study teams, such as yourself, to inform our, if it's not the chief exec, but some of the higher management levels in the trust about our participation. Because I... whether actually study teams directly approach, emailing the higher management and saying thank you for your trust participation, we are now on this list of blah, blah, blah. At which point they'd probably look at the email and go oh really I didn't know we ... but it's sort of ... I guess it might help." INT 4 (Research nurse, Intervention site)*

As well as contact with senior management, some investigators suggested 'experts' from the central PQIP team attending local site meetings may improve management engagement.

> *"What was helpful from one other study which was prevent I think, one of their research fellows came out and spoke to us at an audit day, just sort of for half for an hour. Because we'd done some local presenting on it and bits and pieces but actually having someone from the centre conducting it coming out seemed to spark a bit more interest. Albeit that was towards the end of their study, so there was ... it was not the ideal time to do it but to be fair to us that was the first time we'd asked them to come out. That seemed to raise a bit of awareness because people understood what was going on.*
>
> *Would that be helpful? Yeah it probably would but it's quite an amount of work for you guys to do, but I think that would give the management an idea of what we're up to and why it's important and the fact it's not some sort of local whim. It's a national study that's potentially ... well*

should guide a big improvement at a local level as well." INT 7 (Anaesthetist, Intervention site)

Sites included in the intervention group of the study generally viewed their participation as a positive opportunity.

"Yeah, no, again that was another … we were pretty chuffed. We were aware we were just one of 10 sites that had the opportunity to be involved in that and therefore didn't squander the opportunity. So, yeah, we've been trying to get our heads around this VLAD stuff and how it … it wasn't too hard but I think it was a good way of presenting the data and I say it's very visually quick to see isn't it whether you're above the expected or … and to make morbidity outcomes and things." INT 4 (Research nurse, Intervention site)

However, this was not universal and some investigators felt engagement with the study was not relevant to their role.

"… until I got your message and linked it up what it was [I wasn't aware of it]… because we don't do research into it." INT 9 (Research nurse, Intervention site)

Clinicians generally considered the process measures recommended within the pomVLAD dashboard were useful and provided QI opportunities for clinical teams.

"I think they're appropriate. I think the nasogastric tubes don't exist, the drains do but I think it would be interesting to look at the drain behaviour of the different surgeons you know it gives you an opportunity to ask questions. So we've got about a twenty percent drain rate roughly and well is that everyone or is there a variation in practise, I think the dreaming is fantastic for the ward, for the ward nurses." INT 12 (Anaesthetist, Intervention site)

"I think those recommendations are noted. If ... they should ... they're probably in place already but it's always a good aide memoire to have it before, you see, because you do lose sight of one or two things don't you? And yes every clinical situation is different. But I think the recommendations do help a lot of surgeons to keep it foremost in their mind because sometimes care varies from one day to the next doesn't it?" INT 1 (Surgeon, Intervention site)

"I want to get the surgeons; I mean in my mind this is just tailormade for surgeons to do QI and I haven't got the time or don't think I should be the one who's doing all the QI side of it I can't possibly take it on so I'm just waiting for the opportunity to get surgeons involved." INT 11 (Anaesthetist, Intervention site)

DrEaMing as a postoperative outcome measure particularly captured the interest of local investigators as a target for quality improvement. The inclusion of DrEaMing as one of the national improvement priorities, identified in the first annual PQIP report increased this enthusiasm.

"We presented the data and we were quite shocked that we aren't achieving the dreams that ... particularly focusing on the dreams because that's the one thing that stood out for us in colorectal surgery. We've always felt that we provide a high quality service, it just [the data] highlights the fact that we couldn't ... all these we need to improve in…

"So the starting point will still be your dashboard… and the dreams I think is quite an important agent to identifying that we have a good approach to delivering dreams but if we haven't why is that? Could we link that back to lack of nursing staff? Could it be linked back to change of staff, [lack of] juniors or registrars doing the ward rounds? There's a lot of things that you can infer from the data and you go back and understand why we've not achieved that and aim to improve it, where things we've identified a shortfall of care we can improve. Does that make sense?" INT 1 (Surgeon, Intervention site)

"I think we are doing very well in four of the parameters that you've mentioned, except for the dreaming. But even the dreaming, I think the early mobilisation is a problem. I think it's, I mean this has been identified and we have looked at it to see what can be done to improve." INT 6 (Anaesthetist, Intervention site)

"I think the dreaming is fantastic for the ward, for the ward nurses." INT 12 (Anaesthetist, Intervention site)

Despite the enthusiasm for DrEaMing as an outcome measure there was some concern about its definition, how the data is captured and whether that might influence apparent performance.

"I think the definition of eating may need looking at because some people don't eat because they don't want to or some people have a mouth full of something and what is, I don't know this is picked up from actually being on the wards when you ask people are you eating and they just go I don't feel hungry. So there's not feeling hungry and not eating versus feeling sick and not eating and I don't think they're the same." INT 12 (Anaesthetist, Intervention site)

"And all areas where we could improve, my only worry is you know things like the whole eating thing on day two, three. And it's a tricky one really, because then you are looking at and thinking, oh gosh is that down to my interpretation of the data, do you know what I mean. That's my … or the questions, because for example you know when you say, have they started eating say diet on day one isn't it, based on post op." INT 5 (Research nurse, Intervention site)

The use of dial displays, with colour coding according to performance, supported identification of areas for improvement.

"So you look at these things and you think oh good, we're in … I was going to say something so yeah, green is good isn't it. Orange is sort of ok-ish and then the red/pink is sort of less than … so you obviously focus on those and you think oh, well why aren't we doing this quite so well and yes, and then you want to try and … yes, I think there's a human nature isn't there to try and then improve or try so that the next time you're in the green." INT 4 (Research nurse, Intervention site)

"I am kind of just obviously looking at the pomVLAD and the enhanced recovering stuff that you can … you know you can see what areas that need improving… But for me personally, you know I can look at it … I do find it quite straightforward and simple to see the areas that we need to as a trust obviously improve it." INT 5 (Research nurse, Intervention site)

However, despite the potential to identify areas for improvement, the impact of the dashboard was reduced by limited quality improvement activity.

"we haven't used it enough yet specifically for QI… hopefully I'll be able to tell you in a couple of months once I can get these surgeons up and at it." INT 11 (Anaesthetist, Intervention site)

"I think the answer is yes it is useful but I mean I think it's going to be, I think the absolute value of it is going to be okay if we have a meeting

next week where we go well our temperature on arrival in recovery is fluctuating between sixty and eighty percent should we try to get that higher and so I don't think we're going to know for another year the value of the dashboard. So, if they result in you doing a change in practise that results in an improvement then the dashboard's valuable, if it doesn't then it's of less value." INT 12 (Anaesthetist, Intervention site)

One potential reason for limited improvement activity was the perceived need for 'bedding in' period prior to clinicians acting on results.

"It's still data collection; we are just beginning talks about where we present the data. Having said that we've done brief presentations you know sort of three to five minutes to make people aware of what's going on but we've been waiting for a year of data to do a bigger presentation and I think that's happening next week at a joint surgical anaesthetic meeting." INT 12 (Anaesthetist, Intervention site)

Other reasons for limited improvement work included the perception that the enhanced recovery recommendations were not be suitable for all surgery types and were simply not deliverable in the local context.

"The gynaecology and the orthopaedic one that's gone into colorectal has different bits attached to it and I'm more of a fan that it's attention to detail rather than the individual aspects. And I know your little five dials going across the top, plenty of ours won't get at least three of those just because of the operation they're having. It's not feasible to get them, it's just not possible." INT 7 (Anaesthetist, Intervention site)

"But it just makes you look at them [enhanced recovery recommendations] with a degree of caution with it because having an awareness when we see a particular low percentage on those dashboard outcomes that that's there for a reason. And we're never going to get particularly high scores on those which just means you pay less attention to it for that speciality. Which I think when you're talking about altering the metrics that are displayed and customising it, if you're able to do that per speciality, per speciality output which is perhaps even more preferable. Because there's three or four there that I would do away with from the word go for a couple of our specialities because they're not ... they're completely unachievable at the moment." INT 7 (Anaesthetist, Intervention site)

Dashboard engagement and its effect on data quality

Poor performance identified through the pomVLAD dashboard stimulated discussion about data quality and led to local investigation of data quality.

> *"So initially we picked on a temperature one a couple of months ago just because from the report we seemed to be doing quite well. And then I think on the pomVLAD dashboard we looked like we'd dropped and we were going to try and find out why we dropped and what had happened…*

> *Our sort of internal record-keeping suggested that there was a slight difference between the two. And it also ... the other thing we were trying to do at the time, and we still are trying to do is to try and convince our theatres to effectively buy more under body warming mattresses. INT 7 (Anaesthetist, Intervention site)*

> *'they presented their Audit and relayed it back to PQIP obviously for the drinks which was great, and is going to look into how things can be changed, whether it's documentation, because I think it's a combination, documentation what they could do to improve that, as then should the drinks be prescribed and just spreading the word really for nurses to document that they have been given, or just, so looking at ways of improving that." INT 5 (Research nurse, Intervention site)*

> *"… my only worry is you know things like the whole eating thing on day two, three. And it's a tricky one really, because then you are looking at and thinking, oh gosh is that down to my interpretation of the data, do you know what I mean. That's my … or the questions, because for example you know when you say, have they started eating say diet on day one isn't it, based on post op." INT 5 (Research nurse, Intervention site)*

> *"We did discuss it last time we had a meeting with the anaesthetists and the research nurses. Some of it I think will be to do with data so I've seen we're getting thirty percent for interoperative warming and that just won't be true unfortunately, is that process or is that the outcome and the temperature and recover?*

> *"What we do in our trust is we warm our fluids in a warming cabinet and then give them like that as a money saving thing and we very rarely use a ranger activity warmer. So I never tick active warming of fluids but they're using warmed fluids at forty degrees but maybe we should so*

everyone gets that and everyone gets a bear hugger." INT 11 (Anaesthetist, Intervention site)

Discussion sometimes identified errors in local documentation that were easily corrected. Occasionally further investigation was needed to fully understand results.

"And then various people before said that everyone's got a temperature about 36.5 and what have you when they get to recovery and then PQIP suggests that quite not so much, in pomVLAD I think suggested it was much lower than we thought, it was dropping. So it was again trying to confirm that which was accurate basically. Is it our local records that are accurate or what our research nurses are putting in which may be more accurate than what we are recording ourselves." INT 7 (Anaesthetist, Intervention site)

"They also looked at DrEaM, how they can change things. And also the preoperative carbohydrate load. I think more with the documentation, it's been agreed in the surgical department. The problem is the documentation not exactly how it is practiced. I think the idea of following exactly what is prescribed might be slightly the issue, but physically it is not documented properly. And when it gets filled in it doesn't get transferred what happened exactly because it's not documented. So [consultant surgeon (surgical lead)] has agreed that documentation is a problem. They are looking at that aspect as a team." INT 6 (Anaesthetist, Intervention site)

"So that makes it easier for us in terms of collecting PQIP data because it's documented a bit more clearly now... No I think, from what I gather from what they've said is that it was happening but it just wasn't being documented." INT 10.2 (Research nurse, Intervention site)

The pomVLAD dashboard simplified the presentation of and led to clearer understanding of complex morbidity data. Participants felt it could feed in to regular morbidity and mortality departmental meetings.

"Where we are in the PQIP cycle, is just behind that, you know. NELA's been running longer, and I think we get there ... and I think the pomVLAD data helps with that, I really do. Because that as a graphic, is far easier to understand that complexities about how the morbidity data is presented. Because it is a bit hard to get your head around, well, you've got you know, morbidity, and then what is severe morbidity? And that

crossover between the two ways of documenting it." INT 14 (Anaesthetist, Intervention site)

"… because obviously the pomVLAD is real time and we can get ... we can track it, from an audit point of view it makes it relatively straightforward to see how you're doing and get almost a carrot to encourage people to carry on." INT 7 (Anaesthetist, Intervention site)

"Well at the moment your dashboard is probably what we're going to be using to measure things as it stands. The data ... what I would like to do is to bring it into a continuous sort of monthly audit that we monitor. But it's feeding back that data and looking at it as a source of looking at quality of ...

We've always felt that we provide a high quality service, it [the data] just highlights the fact that we couldn't ... all these we need to improve in. You have to keep an eye on these things and these things it must be ever present in your mind, in your ... it must be there in front of you to be looked at. And your dashboard is probably going to be a way of doing that. It could be a part of mortality, morbidity discussion we have on a monthly basis…

And identify that fairly early on and then monitor, if we've put in place the right bit of training, right level of staffing levels then has it made a difference or not. It's common-sense things but as a graph of quality of care I think it's a very good, it would be a very good indicator tool as to where we're falling short." INT 1 (Surgeon, Intervention site)

However, the need for accurate outcome data was emphasised by one clinician.

"For us the other very important aspect is the accuracy of our morbidity data, because we, three consultants, myself ... [two consultant colleagues] and myself, lock all the records, so we review the patient notes before every one of our PQIP records is locked." INT 14 (Anaesthetist, Intervention site)

The appropriateness of measuring the POMS at postoperative day 7 was raised by another investigator, with concern that later time points may be more suitable depending on the surgery performed. This led to local discussion about how pre-existing work could be improved.

“Because we hope ... we're in this discussion about morbidity as a useful marker but it depends at what stage of the morbidity you're trying to look at. And for us that becomes speciality and procedure specific based on likely the recovery times such that 30-day morbidities a bit of a poor outcome marker if you're expected to take a year to recover from your surgery. And therefore probably what's more important is morbidity, or mortality and morbidity at a much later date.

And we're ... so we're sort of going right back into our data to figure out which is the critical points from our morbidity, mortality and at what ... not necessarily what processes. But you need to measure the processes for a part of it but what do we think is going to affect those markers and how do we measure them better.” INT 7 (Anaesthetist, Intervention site)

5.2.3 Perceptions of the pomVLAD intervention

Some local investigators felt that the pomVLAD study provided an opportunity to engage surgical colleagues in quality improvement work.

“I mean in my mind this is just tailormade for surgeons to do QI and I haven't got the time or don’t think I should be the one who's doing all the QI side of it I can't possibly take it on so I'm just waiting for the opportunity to get surgeons involved…

I think one of the reasons maybe you know when I send out the report it's all mixed in with other specialties, anaesthetic perspective things. I think the other thing that the pomVLAD is much more to do with the surgical side and I think it will get their attention so I think in theory yes it will be an advantage but in practise I've yet to get my surgeons chomping at the bit for it. INT 11 (Anaesthetist, Intervention site)

“The surgical lead is friendly and receptive to it [PQIP]. But there haven't ... we haven't engaged that much with them, and our coming presentation is an effort to try and improve that, but I think probably the pomVLAD data may well help from that perspective… but I was going to use that as an opportunity to highlight grade three, grade four and grade five morbidity that we've had. Just so we can use that as an opportunity, as a point just to highlight that these things are happening.” INT 14 (Anaesthetist, Intervention site)

Morbidity was generally perceived as a useful outcome measure to report, and likely to be more relevant than mortality in the context of a monitoring tool.

> *"… if you're entering say a hundred or a bit over a hundred cases a year and you've got a two percent mortality rate unless it's risk adjusted morbidity it's not going to be useful for five years or more." INT 12 (Anaesthetist, Intervention site)*

> *"Mortality is done to death really if you ask me. There's so many end points, by and large the major end point of most situations is mortality. I think you're best to focus on morbidity really because there are really predictive models in place for mortality, and that's one on the preoperative situation of the patient… Because if you look at mortality in surgery, even for colorectal it is so extremely small. I think you're going to lose ... if you do have that within your study parameters I'm not sure if you're really going to gain anything out of it really... But I think what you have focused on is worth focusing on, morbidity, absolutely." INT 1 (Surgeon, Intervention site)*

> *"…but it certainly having morbidity, grading it, and identifying morbidity that you care about more, and it's therefore deemed as you know, an indicator of poor practice, potentially, or practice you need to look at. It's certainly something you should do." INT 14 (Anaesthetist, Intervention site)*

The ability to examine organ specific morbidity within the POMS was thought to be particularly useful and provided an opportunity to make targeted improvements.

> *"Yeah, but you think again that that … so the overall outcome morbidity data is useful and important. But actually if you were able to then maybe drill down into the data a little bit more to see where that morbidly is. So is it like you said, is it a problem with wound infections, is it a problem with respiratory infections, that that would be helpful…*
>
> *… then your intervention is much more intelligent then.*
>
> *Absolutely, it would make it [implementing change] far easier. It would be far more meaningful, make it far easier. And actually the response to it, would be quicker." INT 15 (Surgeon, Control site)*

Despite morbidity being considered a useful outcome, the range of morbidity and complication measures reported in the PQIP study confused some local investigators.

> *"So, you know, that's more consistent I guess, with what we do in M&M meetings. Whereas day seven poms is just, it's just identifying some stuff that you could call morbidity. Whereas the other one you get to decide what you call morbidity, based on your clinical experience. So it's ... those two systems aren't the same, they're not even particularly that similar…And so that's where the confusion comes from really…*
>
> *The difficulties, it's getting your head around that POMS major stuff. I'm probably about to say I'm getting it wrong now, but ... so you have your day seven POMS data don't you, and it's one of those that POMS will be defined as being major, depending on how it's cross referenced against the Clavien-Dindo scoring system isn't it? And that is the Wong et al. who publish that...*
>
> *For example, there's no GI POMS that's seen as being POMS major, for example. And it's just getting that ... it's been mapped by somebody, it's just a concept to get your head around, so I think all of it is a little bit confusing." INT 14 (Anaesthetist, Intervention site)*

Participants felt the pomVLAD dashboard would allow timely identification of poor performance and provide a simple way to monitor care delivery.

> *"Well at the moment your dashboard is probably what we're going to be using to measure things as it stands. The data ... what I would like to do is to bring it into a continuous sort of monthly audit that we monitor. But it's feeding back that data and looking at it as a source of looking at quality of ...*
>
> *"Theoretically it could change quite quickly and I'm not sure what action I'd take on a report that I would think is … I mean for example if it hadn't been in the main report and that's all I was looking at, goodness knows what could have happened since May in terms of our practice. So it's almost … it's too historical then. I think it's got to be very current and very dynamic and therefore personally I think I'd take much … give much more credence to looking at it on the dashboard on lunch … presumably … I don't know how often … I mean it's telling me here that this is up to July 2018 so I'm assuming it's within a … is that the end of July? I'm assuming within two weeks of this data being analysed." INT 4 (Research nurse, Intervention site)*

If accessed regularly, the dashboard may change the way local teams interact with their data and reduce their reliance on static reports.

> *"I'm a big fan of real time, or semi real time dashboards. It sounds fairly ... I mean to me it's fairly obvious that having yearly reports are great and they're very useful but you are looking back over the last 12 months and it doesn't really give you the opportunity to intervene on a weekly basis or a monthly basis to change something that's obviously not performing as it should. Whereas the closest you can get them down to real time week by week, month by month I personally think you've got more opportunity to intervene early if you see something that isn't performing as you'd expect, rather than waiting for a three month, six month, yearly report. " INT 7 (Anaesthetist, Intervention site)*

> *"Well that is brilliant. I think better than I thought. But that's the sort of thing I thought. I thought it was live data and that's fantastic. And on that basis to be honest I don't think I'd ever even look at it in a report." INT 4 (Research nurse, Intervention site)*

Although to realise this potential regular engagement was needed and the impact would be greatest when data are acted on and improvement projects implemented.

> *"I think if it used correctly, I think it's going to be useful actually. There's going to be a bit more information and much more earlier. So I think it will be useful actually in fact. But the only thing is you need to look at the information first. That we haven't done so far actually." INT 6 (Anaesthetist, Intervention site)*

> *"Yes, I suppose it has added, I kind of feel that … you know for it to really add to it, there is something it needs to be acted upon, do you know what I mean?*

> *There is no point in collecting the data and you guys doing what you are doing, without you know the surgeons then going back, and feeding back to their colleagues. Who are actually really interested in the findings you know, so yeah I think for that reason it could do, as in you know pomVLAD could add to it definitely." INT 5 (Research nurse, Intervention site)*

One clinician raised concerns over the performance and accuracy of risk-adjustment although other participants did not discuss this.

"Well, I'm always ... for my cases, I'm always questioning risk adjusted models, purely on the fact that do you have enough data points on a grand scale that will give you the power for the details that you want…

What other variables which ... like chemical and haematological markers in itself, how are you going to use that in the model itself? Because there are obviously other parameters apart from the appearance of quality, all the things that you're collecting, there are parameters which are just as good indicators. For example, you look at someone's serum albumin for example, we know the outcome and they're going to be in hospital for two weeks, you see what I mean?

Obviously I'll be very focused on colorectal and I do think in the current climate what's great about the study is that we have moved on five, ten, last five to ten years when it comes to perioperative care. And we don't have a risk adjusted model for the current level of care that we have. It would have been ... previous models would have been based on laparotomy as opposed to laparoscopic resection. All models of care leaving NGT within five days and keeping drains in and things have moved on so much that at this point in time, while for that you're probably going to get much advantage ensuring that the model works, I think it's still worth doing because of where we are in the care, perioperative care provisions." INT 1 (Surgeon, Intervention site)

Despite this concern, risk-adjustment of outcomes was considered important by another participant, particularly when comparing performance between institutions.

"Well I think what you have got … what would be ultimately useful is to understand the case mix that is going on in that hospital. Because our case mix will be hugely different to the next hospital down the road, because we have got loads of centralised, sub-specialised services and do operations that aren't done anywhere else in the region. So comparing that data set in terms of just looking at complications, comparing it to a unit where the upper GI activity for example, mainly gall bladder surgery. Whereas here it's 50 esophagectomies a year, and you can't … it's comparing apples and pears really, it's not really like for like.

… you need to stratify it really don't you, to say that we have got … here is the risk that we have got, so here is the … you know based on the cardio pulmonary exercise tests, anaerobic thresholds or something, something to stratify the risk. Or you know ASA grades to keep it very crude, I don't know it's something that would stratify the risk. And the length … you know the operative time as well, so you can see that these

are long major operations that are going on, compared to … I don't know but you get the idea, you are talking about trying to stratify the risk.

Yeah the outcomes are then stratified as well according to the risk of that operation risk, length of time of surgery and complexity of the surgery. That would then mean that I could compare our data with a similar unit, and then I really do get an average that I can benchmark to." INT 1 (Surgeon, Intervention site)

5.2.4 How do local investigators use the dashboard

Investigators thought the pomVLAD and other PQIP dashboards were useful for obtaining a quick overview of care delivery.

"Yes, yes mainly because it [pomVLAD dashboard] is the best one for, well it's potentially the best one for being, you know using observed versus expected and because it looks fantastic for us." INT 12 (Anaesthetist, Intervention site)

"I think we just flick through them all [the dashboards]… Yes, if you've got the thing open just before you close it you quickly flick through to go how're we doing." INT 12 (Anaesthetist, Intervention site)

"I do more sort of on an ad hoc, frequent basis, look at the dashboards, because that's your sort of immediate quick look. All three methods of looking at data [reports, dashboard and presentations] are sort of useful in their own way, I think" INT 14 (Anaesthetist, Intervention site)

However, access was generally infrequent and often related to other activity on the PQIP website such as inputting patient data.

"A few times [looking at the pomVLAD dashboard] nothing … just out of interest really, just to have a look at it. But not too regularly I would say.

Yes, well again I have looked at them [PQIP dashboards], so I am kind of aware of them. Just to kind of get an idea just of things at [NHS trust], but not too regularly again to be honest with you." INT 5 (Research nurse, Intervention site)

"I've got to be honest to say I look at them [the dashboards] if I'm putting patients in and I've got spare time I have a quick look but I don't look, I don't personally look on a regular basis." INT 12 (Anaesthetist, Intervention site)

"No, hand on heart, not a lot. I mean I've really only properly looked at it … when I was aware it was available I looked at it out of interest." INT 4 (Research nurse, Intervention site)

The availability of quarterly PQIP reports reduced the engagement of some participants with the dashboard who felt it did not add additional information to that contained in quarterly reports.

"I don't even know where to find the dashboard to be fair, because I've looked look at the report, and the report is so well written and formatted and I know where to find it on the site, that's all I've used." INT 13 (Research nurse, Control site)

"I must admit I don't actually. I would like to. The trainee lead who did, they did a bit. I tended to use the reports so my involvement in reviewing the data and analysing it and distributing it tends to be report based." INT 3 (Anaesthetist, Control site)

Despite infrequent access, some participants felt real time dashboards were an exciting concept, although lack of previous experience and familiarity in using them for monitoring care delivery limited their impact.

"No, the dashboards are brilliant, you know this is the first kind of study showing live reporting. As a concept it is brilliant, no doubt about it. So very useful and so it's fantastic really most of the things that come." INT 2 (Anaesthetist, Control site)

"Looking at the dashboards that's what really … I think the quarterly report might be useful, I mean the quarterly report will also say what the dashboard says of course. But you know it depends on the timeline of what we are doing, a few things we want monthly sort of update. Then obviously the dashboard is updated monthly, so we will get a quicker feedback on a monthly basis then. So, I would say we would probably look at both really." INT 2 (Anaesthetist, Control site)

"The dashboards are useful for what's there but the reports tend to give ... they give a bigger picture that we can pick things out of a little bit easier at the moment which might just reflect our not great usage of dashboards in the past." INT 4 (Research nurse, Intervention site)

5.2.5 Strategies to build and maintain engagement

Involvement of local investigators with the data collection process increased engagement with data displays. Participants reported accessing the dashboard around the time of data entry but admitted limited access at other times. Lack of time available during normal day to day work was highlighted as a key factor limiting data review.

"I mean what you tend to do is whenever you've had a session sticking data in you have a quick look at the dashboard just to see how it's going." INT 12 (Anaesthetist, Intervention site)

"Yes in the way we look at it but I think the difficulty is time and you know you look at something when you open it up, so for example if I do some note checking to close down records and I've got five minutes at the end of that I'll have a look at it, if I input data at some point I may have a quick glance at it but I've just got to be honest, because of time constraints I don't look at something regularly that isn't in front of my face." INT 12.2 (Anaesthetist, Intervention site)

"You know I couldn't say hand on heart that I do it on a regular basis. But if I am on the PQIP site because I have been uploading or whatever. Then I might, you know or I've had an email today to say that, the report is out or whatever. So I would definitely have a look at it yeah, I think there is probably a lot more there that I could look at. You know I think there is quite a lot there on the dashboard, you know including research and things like that, that I don't currently use maybe." INT 5 (Research nurse, Intervention site)

The release of quarterly PQIP reports acted as a trigger for participants to engage with their local data but engagement was limited between report cycles.

"… clearly when a report comes out you look at the report." INT 12 (Anaesthetist, Intervention site)

“And it's like I know a report's coming out, I won't particularly look at the data, I'll wait for the report to come out and I'll then use as a prompt to you know, do another presentation to one of our clinical specialists' meetings or send them out around, you know, the generic updates, for my consultant colleagues, for each report that comes out…And I use the reports to trigger me to do that, as a general update.” INT 14 (Anaesthetist, Intervention site)

“I look at it very occasionally actually. Then I do look when the quarterly reports come, or if somebody is asking me something, then I do look at the dashboard. Or if it gets like that from PQIP team, or when the information comes from the PQIP team, then I will cross check, to see what's happening to our place. More than that I don't look into that, every week or once in a fortnight, by looking at that, I think it might be more useful actually.” INT 6 (Anaesthetist, Intervention site)

“I guess when we’re made aware that the actual report is active, I know you can’t do it more frequently but I do look forward to the reports coming out. That gives us an opportunity to sort of … we always say that we don’t see back to the actual clinical stuff…

So I would probably not look at the data in between those three months. But that email was my trigger to think oh right latest reports are up, let's have a look at them.” INT 4 (Research nurse, Intervention site)

“The research nurses obviously flag it up to our anaesthetists and surgeons and we say what do you reckon to the latest report, have the you seen it, are you going to be telling your colleagues, what did they say when you presented it to the department and so we’re interested in what’s happening to the reports and we make sure we are sort of … it’s fed back to us. But the dashboard I’ve been less used to using I guess or just I don’t have the time really to look at it on a regular basis.” INT 4 (Research nurse, Intervention site)

“… just at the time that the reports come out, which is quarterly isn’t it, so yes, maybe at other times, because, you can have, you can fall whatever, five, ten minutes, if you’ve got that email and you’re thinking, right okay, I’ll just have a quick look, you would, I have got that time to do that, but off my own back, it wouldn’t necessarily be something that I’d say, Oh I’ll do that.” INT 5 (Research nurse, Intervention site)

The inclusion of PQIP and pomVLAD presentation at local clinical governance meetings triggered engagement. If meetings fell between the PQIP quarterly report cycle some clinicians reported increased utilisation of dashboard data to ensure they were presenting up to date results.

> *"And I use the reports to trigger me to do that, as a general update. But then, between those times, you know, if we can have a PQIP meeting, let's say, or have a meeting with my surgeons, to talk about stuff, or ... for whatever reason, if there's another smaller prompt, I'll look at the dashboard data instead." INT 14 (Anaesthetist, Intervention site)*

Implementation of local quality improvement projects related to the PQIP dashboards stimulated engagement. Participants felt dashboards were particularly useful for monitoring the impact of improvement activity.

> *"So I think that was highlighted in one of the reports and then [trainee] who was the trainee with me, they went through the data much more meticulously and break it all down and looked at it. Then we presented that and then took it from there..*
>
> *So when we looked at the data and tried to understand it, we also looked at, or we surveyed the general or colorectal anaesthetists about the preferred analgesic technique and found actually that, particularly for laparoscopic colorectal work, and found that there was quite wide variability. And one of my colleagues had done a separate audit actually, not using PQIP data and had done a literature review and felt that we should be increasing our spinal opioids use and dose. So that's what we agreed amongst the five of us that we would do. And so we've got a more uniform analgesic technique of spinal diamorphine at higher dose for laparoscopic resection and that that seems to have made a difference. But it all stemmed from the data coming out of PQIP." INT 3 (Anaesthetist, Control site)*
>
> *"… we've had four medical students join me for their Research Speciality Unit, and I'm obviously left to supervise them, so I looked at PQIP and thought oh we can do something here, around pre op, perioperative anaemia because of a lot of the international initiatives that have come out from PQIP, looking at enhanced recovery, and around obviously pre-operative risk. So I've looked at the dashboard and actually the medical students that I'm working with over next year, I was looking at the small little projects around those three areas, so I've looked at it, but it's not part, I feel it's not part of my remit. I know, it's easy to say, it's nothing to*

do with me and healthcare is everything to do with everybody but if I spent time doing that side of things, I wouldn't get my day job done." INT 13 (Research nurse, Control site)

The release of the PQIP annual report in April 2018 and two collaborative events held to coincide with it stimulated engagement with local data.

"I think, we have the PQIP annual day in London so I attended that. So, when the annual report was published… yeah so that identified a few things individual for the trust because that's what the report said. And a few of that were national, sort of common to all trusts like anaemia, HbA1c where pretty much everyone was poor." INT 2 (Anaesthetist, Control site)

"A bit variable. At the minute I spend a bit longer, but there are sometimes I hardly based on a few emails, either coming from the hospital or coming from PQIP team… Like when I got the annual report for about a month or so, around that time, been very busy actually. A good few hours a week I would say. It does take time for preparing and presentation, that kind of stuff." INT 6 (Anaesthetist, Intervention site)

"And then the other thing was we're outlying with regards to patients on opioids for longer in other patients. So we need to look at that as to why it's the case… That was in the annual report." INT 8 (Anaesthetist, Control site)

"I looked at the Annual report and then I looked at our most recent individual reports, to see whether there were any trends or anything that we needed to look at, and around the perioperative iron, sorry, anaemia, we've got a Blood Conservation team here at the Trust, a Blood Management team, so in collaboration with [named clinician], who manages that team, they're obviously around perioperative anaemia management… because of optimising peoples haemoglobin through some of the PQIP data, and National data." INT 13 (Research nurse, Control site)

Establishing a team to share the workload of data collection and QI activity helped reduce the burden on individuals and supported ongoing engagement.

"What we've got is three core consultants who're spending time on PQIP, then I would say another five or six who're helping out and then we've got between three and five trainees who are contributing to PQIP and so we've effectively got a team…So the consultants primarily do the checking and final inputting of data and checking of data and the trainees do the day to day management, on a Friday a trainee looks up next week's PQIP patients and sends round an email highlighting what people need to do on what day." INT 12 (Anaesthetist, Intervention site)

"So that would be myself, there's probably about four research nurses or practitioners working on the trial. And then there's probably about four to five anaesthetists that work on the trial.

Yeah, we try to discuss things that we think we may be doing subjectively at our meetings so that everybody is on the same page. INT 10 (Research nurse, Intervention site)

Regular, monthly or bi-monthly meetings of these teams were implemented in three interview sites. Participants felt these meetings promoted engagement, improved data quality, and generated enthusiasm for improvement work.

"Well I think locally what we've decided is we're going to have a two-monthly meeting for two hours when we're going to sit down and look at the data and we found out that every time we get together and meet we get enthusiastic and we talk about what QI we can [do] and it happens whereas if we don't meet it doesn't happen." INT 11 (Anaesthetist, Intervention site)

"So, we try and do a monthly or two monthly PQIP meeting with everybody involved. So, before that meeting I would tend to look at the data and ... because usually we just go like the central PQIP projects that could be done there. Or any issues that we're having with recruiting or follow ups that kind of thing. So that would prompt me to look at the data." INT 10 (Research nurse, Intervention site)

"Probably before, so we have a monthly PQIP meeting that's primarily the PQIP team so I'd probably look at it once a month before that meeting." INT 12 (Anaesthetist, Intervention site)

"Once in two weeks I do open, you know open up the site to see the dashboard, not the dashboard, the data entry to make sure the nurses have sort of completed things. But I tend to meet the nurses once a month, the research nurses once a month, just to troubleshoot, if they

have got queries with data collection and things like, you know what does this mean, and what data should we input. And queries like that, so I mean if I am collecting … if I am doing quality improvement in that for example, we did do a quality improvement in that, with regards to documentation on objective risk assessment obviously." INT 2 (Anaesthetist, Control site)

Participants circulated PQIP quarterly reports by email to try and maintain colleague engagement. Some participants also included summaries or key points related to the report.

"Primarily just emailing round reports to people and it's trying to get, the main bit, the people we've been communicating with are the ward nurses, the senior nurse in surgery and the colorectal surgeons." INT 12 (Anaesthetist, Intervention site)

"Yeah, so I do a quick summary, five key points, email for each report to my department, I print out the report, and circulate it. And I send a summary to our surgical lead. I send updates and summaries to the research nurses as well, just so they can know how we're doing." INT 14 (Anaesthetist, Intervention site)

"The quarterly reports, yeah absolutely, yeah absolutely. And then I tend to circulate it around the colorectal surgeons and anaesthetists who are involved]. So I try to circulate them around the department, sort of keep them updated as to what is happening, yeah absolutely." INT 2 (Anaesthetist, Control site)

"Yeah, so when the report came out I emailed, put an email round, gave the sort of salient points in the email and then ... with a copy of the report. And then also a lot of the information from the report went up on our PQIP notice board." INT 8 (Anaesthetist, Control site)

"But other information, quarterly reports and annual report as well we have shared with the management, both on the research side, as well as the anaesthetic managers, surgical managers, medical directors. All of them are in the loop, and the associate medical director who covers our division as well in the group." INT 6 (Anaesthetist, Intervention site)

However, some participants reflected on the limited feedback this contact received from colleagues.

"So, our PI has sent them out via email to some of the Matrons at the beginning, but, I don't think there was a massive response." INT 10 (Research nurse, Intervention site)

"So, we've used it like that and we've presented and then when the reports come out your, the ones direct from PQIP, they get emailed very broadly and I did last time write an A4 summary of it as an email to send with it, as well as attaching the report. And that gets mixed ... well it doesn't get any negative response but the vast majority is no response and a few people are interested and want a bit more information." INT 3 (Anaesthetist, Control site)

Local data were also presented at clinical governance meetings with the aim of maintaining engagement. All sites reported presentation of local results at one or more meetings to a range of audiences, including anaesthetists, surgeons, research nurses and management colleagues.

"[Presented] to the anaesthetists I have but I haven't taken it any further than that at the moment… I'd probably do it just after the reports; I won't present everything I'll just try and choose a couple of things that they need to focus on." INT 11 (Anaesthetist, Intervention site)

"Correct in terms of using the data, well anybody who's got access to it obviously but the main people who're looking at it are three of the consultants and it's regularly presented at a clinical effectiveness meeting…

There are sessions where the general surgeons are there, if we have a general surgeon and anaesthetic/intensivist we usually put in our NELA and PQIP stuff there and then the report's circulated by email as it comes out.'' INT 12 (Anaesthetist, Intervention site)

"Saying that, we have PQIP and perioperative medicine meetings where the whole team gets together, maybe twice a year. So, it's formal meetings outside of information communication, emails etc." INT 14 (Anaesthetist, Intervention site)

"Yeah, in the month of May there was a combined QI meeting between the surgeons and the anaesthetists, so we presented the whole data, you know detailed discussion on those things. And, I guess the plan is, you know the six months or quarterly, you know within the anaesthetic department, we would be presenting it. I did make an initial presentation last year when PQIP was started just to raise awareness." INT 2 (Anaesthetist, Control site)

"Yeah, so both myself and we had a, for two years we had a trainee lead but they've just moved on in August. But they and myself had been at least twice to the surgical audit meeting and three times, I think, to the anaesthetic audit meeting and presented updates. And we did have an issue, for example that was shown through the PQIP data or high pain scores in recovery." INT 3 (Anaesthetist, Control site)

Despite these presentations, cross-specialty engagement was often limited. This was particularly true of individuals outside local PQIP teams who rarely, if at all, accessed and used data for improvement work.

"Yes, I've got the anaesthetic department updated and I'm waiting for a lot with the surgeons to go and describe this to them, the trouble is know full well only two or three of them will be there and it's virtually impossible and emailing them doesn't work, so I probably need to find some time just to go round and actually show them on my laptop." INT 11 (Anaesthetist, Intervention site)

"I tend to do it as a presentation at audit is the main way but I've distributed it and then if other people ask for the data, then send it to them. And encourage people to have a long answer, they can look at it themselves but I don't honestly think that any of my surgical colleagues do that. It just tends to be the trainees helping the project with me." INT 3 (Anaesthetist, Control site)

Participants also used less formal communication methods to engage colleagues, including informal conversations and using notice boards to disseminate results.

"…the people we've been communicating with are the ward nurses, the senior nurse in surgery and the colorectal surgeons." INT 12 (Anaesthetist, Intervention site)

"I thought I'll present it to individual surgeons at various points while trying to sort of encouraging them to get on board." INT 8 (Anaesthetist, Control site)

"…what I have been doing is I've been printing off the information and putting it on all the notice boards…It's in the department, I also put a smaller notice board up in the ... like a room where the surgeons write their reports in theatre. I put it in there as well". INT 8 (Anaesthetist, Control site)

"We do sort of meet once every couple of weeks informally and kept it that way, we've settled. For example, we operate together, we will talk about what's going on with this and so we sit down with coffee and he lets me know what's going on." INT 9 (Research nurse, Intervention site)

One site shared anonymised clinician outcomes to stimulate behaviour change and maintain engagement with local data.

" I think our approach to QI are fairly basic so I think they'll be a meeting where we go okay these are your temperatures and then you go shall we share individual patients temperatures and that's our usual approach. So, you anonymously show what each individual's outcomes are and then you let the individuals know that the group doesn't know who's who and it's a really simple way of people changing behaviour if you know somebody's watching you, you change." INT 12 (Anaesthetist, Intervention site)

As well as developing engagement within hospitals, participants discussed whether local area networks, where good practice and successful improvement work could be shared might be beneficial.

"We have an informal network. So, we have a perioperative medicine meeting in a pub once or twice a year and the people who attend that are all involved in PQIP. Then I think with the GIRFT stuff we're starting our visits in six weeks' time… do you do PQIP you know and what's your PQIP data like." INT 12 (Anaesthetist, Intervention site)

"The other thing about PQIP is that we are maybe using PQIP as a basis for a regional QI network. So, we are starting to get together with at least one other Trust who is interested and will hopefully suck in another two trusts which are in our region, to start looking at well how are you doing with these and how is everything going in terms of your PQIP outcomes.

Just as a starting point. And say well what have you done about yours and could we learn from that etc etc and share some of the learning and maybe try and standardise some of the practice across the region if we can crack it, that would be amazing. A big goal to set yourself. But that is what we are trying to do." INT 15 (Surgeon, Control site)

The integration of data collection into a pilot of perioperative medicine ward rounds at one site increased interest in care delivery and patient outcomes. Data collected through PQIP and pomVLAD were also used to evaluate the impact of the service.

"… the pomVLAD dashboard definitely has led us to it but actually another development has also led us to it that we started doing the perioperative medicine round on the colorectal ward. So definitely the PQIP data has made us aware of it but I think you're made even more aware of things when you actually see the patients." INT 12 (Anaesthetist, Intervention site)

"I know our Medical Director is aware of it, partly because it ties into our new perioperative medicine ward rounds, that we're piloting at the moment. And very much, you know, the PQIP data, and collecting the PQIP data is going to be partly one of the roles of the perioperative medicine ward rounds really." INT 14 (Anaesthetist, Intervention site)

"So, part of that is they're collecting data about what their impact is. And we will hope with that team, we'll see better dreaming rates, and I know that's what their focus is… I have to see all PQIP patients as part of the perioperative medicine ward round, you hope by looking at these things, you'd improve these things at the same time." INT 14 (Anaesthetist, Intervention site)

Sites who were able to engage medical students and trainees in local PQIP data collection and improvement projects reported a positive impact.

"Yeah, so both myself and we had a, for two years we had a trainee lead but they've just moved on in August. But they and myself had been at least twice to the surgical audit meeting and three times, I think, to the anaesthetic audit meeting and presented updates. And we did have an issue, for example that was shown through the PQIP data or high pain scores in recovery." INT 3 (Anaesthetist, Control site)

"I've got an F1, not an F1 a CT1 just started now, they've done an actual... they've done a year of [quality improvement training]. So, they're here now to help me with PQIP. So, I'm hoping that will be a good stimulus to get things ... to get looking at some more information and get some... QI projects out of it but I've been trying to sort of get them ... I've been trying to get them to look at." INT 8 (Anaesthetist, Control site)

"Well actually very interestingly, there was, there's a Registrar here, who became involved in PQIP a little bit at the beginning of the year, and then they used the information or looked at the information from PQIP and did their own Audit on Pre-Op Carbohydrate Loading, because I think it was showing that we weren't particularly good at that. So, they looked into that and investigated, and did present at the Anaesthetic meeting and fed that back, and they also, there was a big meeting on Friday, which it was for the Colorectal Surgeons and the Anaesthetists, some ward staff were there, I'm trying to think what that was called, it was the GIRFT [Getting it Right First Time], so they went through that." INT 5 (Research nurse, Intervention site)

Some participants who had not managed to engage trainees acknowledged their involvement would support data collection and QI activity.

"What we'd like to do is get trainees involved with it as well and get them tooling around on their own doing, not quite post-op ward rounds but not far off post-op ward rounds, specifically around this data collection and data entry. And take a load off our research nurses and if we can do that." INT 7 (Anaesthetist, Intervention site)

One participant tried to encourage colleagues to engage with local data through its use for annual appraisal, although this strategy failed.

"Because I have to produce quite a lot of information for my appraisal from it which was quite useful. And I've said, if you've got the name of your patients, I can get them their Bauer satisfaction score from it. But I've had no ... no one's asked for it.

I recognise that's really difficult now because it's dependent on the surgeon and everything else that goes on. But this is something where you can actually get proper feedback on what they thought of ... were they sick, were they in pain, what did they think of the service providers? But that ... I've tried to push that and I've shown them what I've produced but no one's asked me. I think they're speaking next week, I'll get a bit more interest from them." INT 8 (Anaesthetist, Control site)

The Bauer Patient Satisfaction Questionnaire is a tool used to assess patient comfort and satisfaction specifically related to anaesthesia care. It includes questions on anaesthesia-related discomfort (e.g., drowsiness, pain, thirst) and overall satisfaction with various aspects of anaesthesia services, with responses using Likert scales.[224]

5.2.6 Barriers to quality improvement activity

Baseline period of data collection

Participants felt a baseline period of data collection was needed to enable them to understand local care delivery prior to implementing change.

> *"Yeah, well we've had a discussion about that, but I guess to just put it into ... describe it now, you know, partly pomVLAD came out and it was ... you know, we saw it as a good thing, but a bit of an unknown. And then you sort of getting like for PQIP projects full stop, mentally I've decided to wait for a period of time before I pay them much interest, just to let it bed in." INT 14 (Anaesthetist, Intervention site)*

> *"So, for the first year we were getting baseline data on board and we're only a couple of months past that now and what I'm trying to do now is get some enthusiasm from the surgical teams and the anaesthetic teams to do some QI. So I don't really have much to tell management about that except to show what our baseline data is and I've done my best to just do that as well as I can." INT 11 (Anaesthetist, Intervention site)*

The small number of patients being recruited at some sites also contributed to the perceived need for this baseline period. Participants described large apparent swings in performance due to limited patient recruitment and therefore a period of baseline data collection would help them better understand true performance.

> *"it's part of the feeling about local data, so the fact if we're only having three or four patients through, you know, something happens to one of them, that's 25% change.*

> *So we're always going to have very large swings in you know, what our rates of whatever are. Just because one patient, declaring one severity of symptom, is going to be maybe a quarter of the patients, well as low*

as a quarter of the patients, maybe an eighth to a tenth of the patients. So you know, they're always going to have a significant impact." INT 14 (Anaesthetist, Intervention site)

"Yes, if, the first thing that a bunch of consultants will do is say well hold on, where's the scientific backgrounds for this, so you know how many was the denominator, what's the numerator, what is the scope of the data that's being collected i.e. can I actually draw any conclusions from it? Statistically. Because setting off on a tangent with wrong data would be almost as bad as not taking any action at all really I guess. It just wouldn't be right." INT 15 (Surgeon, Control site)

As local patient cohorts grew clinicians felt data became more relevant although some participants identified ongoing issues with data collection affecting the use of data to drive improvement.

"it's because the data's sort of becoming more mature and therefore more relevant. I mean, you sit down and they say "oh yeah, we've had like 50 patients and only 20 of them done a Bauer questionnaire and people go "okay, whatever, I'm not listening to that".

But you know, when you say you've done 100 and whatever patients, and 70% have done the Bauer questionnaire, because I think that's roughly what our rates are, then people will be ... you'll start to listen to it more. And I think that's even before we get into that tipping point now. INT 14 (Anaesthetist, Intervention site)

"The second thing I'd say is that data collection and data input, I think continued to be a challenge. I was speaking to [anaesthetic lead/PI] the other day about PQIP and data collection and data input, and to do it consistently and produce really good data is always a challenge whatever sphere you are in. It requires a lot of discipline, dedication and people around it to make it part of their job. I am not sure that has happened yet with PQIP so I think you know the data entry I think has fallen off a little bit and unless you are producing decent amounts of data it doesn't have any meaning." INT 15 (Surgeon, Control site)

Lack of dedicated time to deliver quality improvement

A major barrier to QI activity was lack of time available during day-to-day work to deliver quality improvement activity. Local investigators felt improvement activity was not valued by their trust and senior management and QI activity often had to be delivered before obtaining designated time in job plans.

> *"Yes in the way we look at it but I think the difficulty is time and you know you look at something when you open it up, so for example if I do some note checking to close down records and I've got five minutes at the end of that I'll have a look at it, if I input data at some point I may have a quick glance at it but I've just got to be honest, because of time constraints I don't look at something regularly that isn't in front of my face." INT 12 (Anaesthetist, Intervention site)*

> *"I can make sure you recruit etc., and all these sorts of problems but it's then having the time then to try and push the agenda of PQIP to try and get people to start doing QI." INT 8 (Anaesthetist, Control site)*

> *"If there was more time available I'd like to spend more time looking at the data itself. But it is very helpful to have the report and to have the pre-prepared presentations because they, that can be quite a burden otherwise" INT 3 (Anaesthetist, Control site)*

> *"Yeah, I mean, if you, if you go and talk to the, you know audit department lead, they will agree and concur and you know agree with the value of the data and all that. But I think when it comes to reality, it's ... I think it's a question of resources and manpower for them really… they were overwhelmed with their routine firefighting really." INT 2 (Anaesthetist, Control site)*

> *"I think you've got to demonstrate it before people will give you money so you have to do the work within your existing job plan and then I think over time will demonstrate will actually people need PAs for this and I expect that will then be forthcoming." INT 12 (Anaesthetist, Intervention site)*

> *"No, it has not been allocated. We have been talking about this. Obviously, the money is so tight within the trust, so people are trying to get more out of whatever is available. So, there is no allocation so far. I have raised it formally at least a couple of times so far, but it has not been allocated. I'm doing it out of my own interest basically actually…*

potentially the dashboard is more interactive and ... it can be looked more often, but I don't know how much time we have available to act on it. I'm not saying that it's not useful, it's certainly useful as I mentioned, but putting that into practice involves a lot more engagement, between the PQIP team, the local PQIP team, and the clinicians and managers. The current working of the department doesn't allow that time commitment from anybody I think." INT 6 (Anaesthetist, Intervention site)

"I think in principle they [surgical colleagues] do agree and they all want to but again it's just you know, time and resources for them as well and because they have an extra pressure to perform and make sure they are within the targets of achieving their activity..." INT 2 (Anaesthetist, Control site)

"I suppose it totally but in a busy work schedule, remembering even every 3 months to go and have a look at the dashboard and see what it is showing etc without a prompt or something landing on ... I respond to things that land on my desk normally and my desk is, I am looking at it now, there's around about 50 sets of case notes to be reviewed, there's 100 emails arrive each day in my inbox, so do you see what I mean, that actually making it part of your business to go and look at this stuff needs some kind of prompt or a kind of did you know, have you looked at your PQIP this month or something like that." INT 15 (Surgeon, Control site)

"So our PI is trying to use the data to do some Quality Improvement at the moment but I think with everything, I think they've got a lot of other stuff on at the moment, so I think it's been quite difficult in terms of getting people engaged in Quality Improvement, because obviously it's difficult to get everybody engaged when you don't really have time to push it yourself." INT 10 (Research nurse, Intervention site)

Where clinicians did manage to obtain time in their job plan allocated to PQIP and quality improvement work, this did not reflect the true time commitment required and clinicians were forced to go above and beyond to deliver the work.

"I'm going to plan for 0.25 for it, that's a couple of ... that's two hours a week, that's what I'm down for.

… so our Trust basically states that if you're a core investigator or project principle investigator, to an externally funded research project, and each project comes with its ... an equivalent to .25, you can use in your job plan.

And so you can use those to make up that one additional, but all my research stuff is essentially above and beyond, because I've got other roles, to fill up that one. I'm not actually paid or funded for it… what would have to happen is that my CD and the management have to agree that they see value in me continuing to do PQIP above and beyond, and therefore losing clinical sessions to it.

And that is, just isn't something that I've negotiated with them at the moment. And I guess that's an ongoing debate around the usefulness to local Trusts, or these type of dashboard projects, be it NELA or be it PQIP." INT 14 (Anaesthetist, Intervention site)

Participants felt allocated time was needed to support quality improvement activity, although this did not necessarily need to be clinician time. Involvement of administrators or junior managers was suggested as an option to improve the way data are used to improve care.

"I don't think we've got the capacity to do that, because Quality Improvement isn't research per se, I don't, I know it's all part of the clinical effectiveness cycle, but if I took time out to actually implement the Quality Improvement side of things, I wouldn't get the research done, so I think there needs to be some ringfenced time, somewhere along the line, for the Anaesthetist and their trainees, to take on the QI side of things." INT 13 (Research nurse, Control site)

"I guess as clinicians we tend to focus on clinicians. And actually… it's the management and senior management that has the ability to allow time…

Yes well they have the ability to ... because they've got a big general management team at all sorts of levels, and it doesn't take ... all you need is a named person for which PQIP is a dedicated part of their working week. So they don't need to employ another person, they just need to find a junior general manager in the surgical division who they ... who has got time in their job to take on PQIP as well as all the other things they do. Because it's not full time, if they spent two to ... two hours a week, four hours a week maybe within their job checking, analysing, producing data. Talking to the clinical leads, the reports ... and distributing reports, that would make a massive difference I think." INT 3 (Anaesthetist, Control site)

Difficulty engaging colleagues

Difficulty engaging colleagues was another major barrier to implementing QI. This included participants struggling to engage colleagues within their own department and across specialties.

> *"I mean I guess there are very high performing trusts where everyone's up and can do about it and there are other trusts probably like mine where everyone's just got their noses to the grindstone a little bit and it makes me feel bad that I'm a trust where I can't engage the surgeons that easily but I ran another for four years and came up against the same thing." INT 11 (Anaesthetist, Intervention site)*

> *"We have got surgeons, there is a rep, and so the surgical representative the involvement is getting better I would say. And hardly any involvement in the first year, but there is a bit more buy in and involvement this year, but there is still a long way to go from the surgical side." INT 2 (Anaesthetist, Control site)*

> *"I mean when I first proposed that we take part in PQIP I went to see our colorectal surgical lead who was really, really negative and their first comments were ... because I asked whether they would be the surgical lead. And they said I wish you lot would just leave us alone to get on with our jobs. So I knew there was going to be a bit of a struggle there." INT 3 (Anaesthetist, Control site)*

> *"… also the responses coming when you engage with people, you know, how things can be done differently. Like for example our department, at the last meeting I pointed out that preop risk assessment needs to be done differently, and these are the ways to do it, we try to agree how to do it, but there is not much response from my department. So those things do make it difficult, can happen that when I'm trying to do things on my own time, to make a difference. Whereas if the engagement is not coming, it's not just a question of time alone, also how well people are engaging, not just to the clinical people but also management." INT 6 (Anaesthetist, Intervention site)*

> *"I don't think there was massive ... I think from surgeon's point of view they weren't too fussed about it. I think there was a few people interested in, anaesthetists interested but there wasn't very many surgeons interested in using the data." INT 10 (Research nurse, Intervention site)*

"So what I'm trying to say is to spend time exclusively for this, to extract the information and then spread out to the department it takes a lot of effort. And then the second bit is how much the clinicians on the floor, my department isn't very much engaged with it, so I wasn't able to convince so far, about preoperative assessment. Whereas we're able to do a few things better, differently, as compared to what we have done a few years ago…

More than that, I'm unable to engage people, unable to change people so far." INT 6 (Anaesthetist, Intervention site)

"So our PI is trying to use the data to do some Quality Improvement at the moment but I think with everything, I think they've got a lot of other stuff on at the moment, so I think it's been quite difficult in terms of getting people engaged in Quality Improvement, because obviously it's difficult to get everybody engaged when you don't really have time to push it yourself." INT 10 (Research nurse, Intervention site)

Difficulty accessing and analysing data sources outside the main PQIP programme

Participants stated that difficulty accessing and analysing data outside those provided by the PQIP team presented a barrier to improvement work. This included both data within the PQIP programme and linking other data sources routinely collected within the NHS such as administrative data.

"… I've probably got too much on my plate at the moment…So when I have got time I would love to go on and do data queries and show the surgeons how to do it but I'm finding it too time consuming and so it's just been I've pushed it to the side for now…

So what I want to do for the surgeons is stand up and demonstrate how to do it but I've got no confidence that I'll be able to do that at the moment so it's going to be a hard sell but I will try and get that sorted out.

I feel frustrated with myself because I'm sure it's doable but in the time slot I've had I haven't managed to make it work yet." INT 11 (Anaesthetist, Intervention site)

"Because there's about four different databases for each individual speciality as far as I can see, and they don't talk to each other and there's

no way to communicate between them or extract the data that you want from them. So, if you go for something like the database that we've got for cancers. It's obviously possible to get data out of and then link it with our systems, so you end up collecting it in three different databases and repositories which just is a horrible duplication and triplication of work." INT 7 (Anaesthetist, Intervention site)

Failure to collect data in an accessible format that could then be used easily for QI acted as an additional barrier.

"… so those two systems work together to give us the data that we need. It's not easy to interrogate however, apart from ... it's just like you flipping your notes, it's just like almost like a PDF really if you ask me where things are scanned into it. So it's a little bit difficult to interrogate but if further data is required then it means obviously the old-fashioned way of going through the notes and going through the electronic notes and extracting the data. That's certainly feasible. I'm looking at how we structure our data to collect data going forward and our perioperative quality. Obviously, that's a big, big project and it's not, you appreciate it's not a priority for the trust at the moment. They're just concerned about hitting that target of all data, of the people existence, how we get there, how the data is recorded obviously is not their worry." INT 1 (Surgeon, Intervention site)

Limited senior management support for quality improvement

Limited support from management teams acted as a barrier to QI. Participants reflected on the need for this support and engagement to deliver positive change.

"I think we're just left alone to get on with it rather than we're supported." INT 12 (Anaesthetist, Intervention site)

"From a management perspective, I don't see much enthusiasm to it. As far as I can tell really, it's about clinicians trying to use that methodology to try and improve things". INT 14 (Anaesthetist, Intervention site)

"I mean so with all of those we are looking then ... so we've got the outline now of a pathway, which this meeting on Monday is partly to get the buy in of the surgical, clinical director and of the management team to say we, as clinicians can't make this just happen out of the blue. It needs to be a directorate of action and we can then monitor our progress with it." INT 3 (Anaesthetist, Control site)

"There is a little bit of futile cycling where we all discuss it, agree what needs to be done but maybe not all the people who need to be in the room are there. And it is actually quite, you almost need the improvement teams and the general managers to come to you to join that final piece together so that actually a QI Project can work. Even though you've got the people who know what to do in there, actually making sure the process happens is a whole different ball game. INT 3 (Anaesthetist, Control site)

"It's very hands off [management support/engagement]. We have support ... we have research nurses that we wouldn't be able to do it really without them. So, it helps produce the data but when it comes to actual QI and actually using it there's not a lot of pressure or sort of support that helps us keep generating QI from it. No, I wouldn't say that's ... it's quite hands offish, sort of left to get on with it on our own really." INT 8 (Anaesthetist, Control site)

"Well, I think they're a bit disappointed because they wanted more management to be there, so they could hear what needed to be done and just ways of moving forward, I think it was a shame, I think they were a bit disappointed. There was somebody there, who wasn't a senior Manager but was from management, who was going, his aim was obviously to pass it all back to management, but yes, there was, they were quite limited really." INT 5 (Research nurse, Intervention site)

"The potential is not realised actually, particularly, I mean, more and more people are aware, so now I don't even speak to people, trying to convince them, so they do know. But as for the management side, it does not come across yet, I feel that's my impression in my set up actually." INT 6 (Anaesthetist, Intervention site)

"From a management perspective, I don't see much enthusiasm to it. As far as I can tell really, it's about clinicians trying to use that methodology to try and improve things, but that isn't part of the QI projects that I've personally done and local projects that I've done." INT 14 (Anaesthetist, Intervention site)

This lack of support was sometimes perceived to be due to conflict between clinical and managerial priorities for improvement.

"I mean it's down to the initiative of the individuals. Organisationally is there a push to improve quality improvement, yes in general. But it's more generic on the lines of the trust agenda, with nothing specific to PQIP really. But if you go with any ideas of quality improvement

especially if it has got, you know financial savings attached to it, and certainly yeah I have never come across them being obstructive as such, but the interest does mainly kind of depend on, you know the agenda you take forward." INT 2 (Anaesthetist, Control site)

"[Our trust is] very much a trust that is focused on quality and that extends all across the board from safe care to high quality surgery, to getting it right first time. But also on data capture. By enlarge data capture is focused on trust performance figures, things like safety thermometer or falls and bundles for care for sepsis or pneumonia. All those various things. But when it comes to other quality measures, which is what I was alluding to earlier that PQIP has more built into it about dreams for example, are they getting up and out of bed and eating and so forth early. All that data I'm afraid isn't captured at another level as it were of quality." INT 1 (Surgeon, Intervention site)

Existing data dashboards available in institutions are rarely clinician facing tools resulting in a lack of experience with using them limiting their impact.

"We do [have some other non-PQIP dashboards available to monitor hospital performance] but I'm not familiar with them because they're used primarily as a management tool not a clinical tool." INT 12 (Anaesthetist, Intervention site)

"there is a dashboard that the trust board looks at so it's length of stay, readmissions, deaths and a few other things but it's not really used by the clinicians." INT 12 (Anaesthetist, Intervention site)

"I mean, I'm sure there's loads of managerial dashboards, that I have got access to. But from what's really keenly publicised to us, from anaesthesia perspective really, it's never ... there is in house dashboards that are created by the clinical governance lead, which is relatively ad hoc, depending on what they're wanting to look at. INT 14 (Anaesthetist, Intervention site)

"sometimes the parameters that the management team want to monitor are different to what the clinicians want to monitor. The management team have the resources and the personnel to keep these dashboards alive and running and maybe the clinicians don't have access to that in an equal way." INT 3 (Anaesthetist, Control site)

"so we've been trying to do something along the dashboard front for a while that would be in, not public access, but available to all staff within

trust. And try and make dashboards that will be relevant to their area and we've been trying to do that for a couple of years and have basically hit various technological, IT, management stonewalls. So there's no sort of interactive dashboard that's available. I believe at management level they have the availability of some dashboards but I wouldn't like to say what they are, nor how accurate they are, nor how often they're used. So I think at clinician level this is probably the first that we've had with any sort of clinical relevance to them." INT 7 (Anaesthetist, Intervention site)

Lack of formal quality improvement structures

A lack of formal quality improvement structure within sites acted as a barrier to improvement activity. Participants felt improvement work was dependant on individual clinicians to deliver and that support to develop suitable data collection tools and select robust outcome measures was needed.

"It's exceptionally dependent upon you and other clinicians to do everything for it. So there doesn't seem to a great way of building a more sort of rigorous structure for picking all your metrics and you know, your process measures and outcome measures etc. Which is one of the sort of attractiveness of PQIP projects, because a lot of that side of stuff is done for us really, isn't it?" INT 14 (Anaesthetist, Intervention site)

"No, not by a long way. I mean I think ... well I think in order to build it in and sustain it is required a dedicated, doesn't have to be a highly banded person, but a dedicated person that physically sits within the department but, maybe, part of the management team who can do the data, not even so much collection because that can be quite difficult. It can be a clinical thing but can ... when it is captured electronically can access it and turn it into reports and work directly with the clinicians." INT 3 (Anaesthetist, Control site)

Some participants felt using clinician time to deliver QI activity rather than them being involved in data collection or analyses was a more appropriate use of resources.

"So I think clinical time ... I think our department is fair, I mean you'd always want more time for that but it seems a waste to me if you're asking senior clinicians, giving them enough time to actually sit down, look at spreadsheets and produce reports. Whereas really they should be interpreting them and driving initiatives and change and projects and

there should be ... it doesn't require that level of hourly pay to produce the reports. So I think it's more what we probably lack is the administrative support to produce them." INT 3 (Anaesthetist, Control site)

"I think from a data point of view we've been very good at collecting data but I think it has been suggested by yourselves, in terms of what we do with the data that requires a much wider discussion with the colorectal team and the anaesthetic department. Because there's still some way to go in getting the culture change towards a better team." INT 1 (Surgeon, Intervention site)

And without a structure providing time, financial and methodological support in place sustained improvement was not deliverable.

"So I think that's probably where a lot of these things get held up and people will try, well what they will do is predictable. They will come up with an action plan saying here's a few things that we need to do and then unless you've got highly motivated people with a lot of time on their hands to get meetings together to do this, to do that, and with an actual methodology that means that you can achieve that change. It won't change, or it'll change very little or it'll change for a little while and then it will get worse and it'll go back to where it was again. So I am speaking as someone who has been in a very senior leadership role for quite a few years and has made quite a lot of high level changes and then have struggled with other changes and I guess I've sometimes not understood why something has worked well and why something is so incredibly difficult to change." INT 15 (Surgeon, Control site)

Local improvement priorities not aligning with the national agenda

Another perceived barrier to QI was when national improvement agenda did not fit with local priorities, leading to study outputs being considered irrelevant to the local context.

"Whilst I remember, there's the bit I don't like, is you often produce posters as well, and little summaries. And quite often, those posters and summaries are what PQIP's agenda is, not what our agenda is.

And that's fine, but we find it useless because they're not our agenda. So, at the moment, we've got ... you know, there's things they're trying to push obviously, the HbA1cs, also, you know, completion of enhanced

recovery I think, trying to push you know, the anaemia bit. And I guess these are all tied into national things that I'd want to be improved to be honest, but I can't exactly remember where they've all come from." INT 14 (Anaesthetist, Intervention site)

One participant also suggested this conflict between national and local priorities was relevant to the pomVLAD dashboard.

"And I know your little five dials going across the top, plenty of ours won't get at least three of those just because of the operation they're having. It's not feasible to get them, it's just not possible." INT 7 (Anaesthetist, Intervention site)

And as a result the relevance of the dashboard was reduced, with the local investigator not pursuing improvement.

"But it just makes you look at them with a degree of caution with it because having an awareness when we see a particular low percentage on those dashboard outcomes that that's there for a reason. And we're never going to get particularly high scores on those which just means you pay less attention to it for that speciality…

they're not overly interested on what's on pomVLAD because the bits we're looking at are ... the metrics aren't ideal for us. So it's sort of given us a push to what we need to do and what we might do with our dashboards." INT 7 (Anaesthetist, Intervention site)

The competition for engagement between different national projects such as PQIP and NELA was a potential barrier to QI, with more established national audit projects perhaps reduing engagement with PQIP and the pomVLAD intervention.

"I think for the first time, we're now getting a large enough dataset, where it feels relevant. And that maybe some ... the other aspect of it is the best form that we present in, is our clinical effectiveness meetings, and like it or lump it, NELA takes over slightly. Because until now, surgeons are more interested in it, and it's been going longer." INT 14 (Anaesthetist, Intervention site)

Staff turnover and poor continuity in clinical teams delivering patient care

Staff turnover acted as an obstacle to delivery of QI in some institutions, resulting in problems with the day-to-day running of the PQIP study.

> *"I think, because there's been a few changes in staffing, and going forward we're going to have another few chances in staffing because there's a lot of admin to ... associated with PQIP as well. So it probably is going to get a bit more difficult actually in the next few months and particularly over holiday periods when we don't have as many ... like Christmas and Easter that kind of thing when there's not as many people on the ground to help with the follow-ups and also just recruitment in general. And it can be quite difficult to keep up with the amount that we're supposed to be recruiting." INT 10 (Research nurse, Intervention site)*

> *"I think as time has gone on and we've changed Principle Investigator a couple of times, that the support and the importance of it, has become less of a priority, and I feel that my team, I'm managing my team to obviously collect the information, the patients are kind enough to take part, we've got a duty of care to do something with the result, and I feel that we're not doing that." INT 10 (Research nurse, Intervention site)*

Changes in nursing staff and junior doctors caring for patients also affected participants ability to implement QI.

> *"So even from a general surgical point of view really but leaving that to a side we are not reaching our target when it comes to dreams… That's due to absences and so forth, that's due to people who are supposed to be dealing with it, dealing with enhanced recovery pathways etc. They've just not been around.*

> *This is becoming ... understanding what the enhanced recovery is because we have a different nurse every other day. Understanding ... juniors understanding what's happening. Juniors often changing, we don't necessarily have the same registrar on the wards on a daily basis. So I'm looking at overall processes because at the end, it's all really ... it's to get them up and about quickly to reduce SSIs [surgical site infections] amongst other things." INT 1 (Surgeon, Intervention site)*

> *"And then your trainees of course they ... I mean some of our trainees stay for up to two years but that's unusual. The majority are six months*

to a year. So they may get one small project done in that time but it's difficult to keep that going over without somebody whose job it is to be monitoring it all the time." INT 3 (Anaesthetist, Control site)

Good or satisfactory performance reduces the need to improve

Where performance was considered acceptable (either based on local views or in comparison to national average) the drive to improve was reduced.

"At the moment, there hasn't been any major triggers for us to go and look at the data I don't think there's been alarming triggers if you like." INT 2 (Anaesthetist, Control site)

"Part of our issue is we seem to do quite well so everyone seems quite happy, rather than if we said oh we're rubbish at this. It's how the NHS used to work, isn't it? You get told when you're bad at something and you have to improve, that seems to be how QI works in the NHS. But because our numbers isn't ... in PQIP, the results are actually quite reasonable, it's quite difficult to ... so it's a bit like less of a stick try and get people to sort of try and work on things and improve." INT 8 (Anaesthetist, Control site)

"[the recent report] was quite positive, there wasn't any particular, you know set of actions that needed to come out of it, if you see what I mean. It seemed to be reasonably quite a … there was no concerns raised from it. So the question is then, what if everything is going okay, almost all things are going okay, then do you need any specific actions related to that, or any specific QI work streams… Because if the only benchmark is the national average, so if you are above the national average then no one sits there feeling a need to get involved in a QI project." INT 15 (Surgeon, Control site)

5.2.7 Facilitators to quality improvement activity

Regular meetings to review data entry and performance

Regular meetings to review data entry and performance generated enthusiasm amongst local investigators to undertake QI work.

"Well, I think locally what we've decided is we're going to have a two-monthly meeting for two hours when we're going to sit down and look at the data and we found out that every time we get together and meet we get enthusiastic and we talk about what QI we can and it happens whereas if we don't meet it doesn't happen. So, prompts would be good I don't think people will get fatigued by them I think that'd be a good idea." INT 11 (Anaesthetist, Intervention site)

These meetings also gave participants the opportunity to discuss interpretation of variables and identify issues that may affect data quality.

"I think that question is coming up at most of our meetings. I guess if something ... so we're prompted by the data that we get to ask the question and then the first question we ask from that is how are we documenting it? Or how are we documenting it on the CRF and are we looking in all the right places to document that?

… But it's usually stuff that's a bit more like the clavien-dindo sometimes, like between number one, number two can be a bit difficult to decide which patient it is. But we've come up with a strategy for figuring that out, but that was an example of something that we had been documenting the way we thought, but then the anaesthetists said oh maybe document it ... it should be documented this way." INT 10 (Research nurse, Intervention site)

Endorsement and promotion of the project by national bodies

The perceived importance given to studies by Royal Colleges such as the Royal College of Anaesthetists and the Royal College of Surgeons was suggested as a facilitator of QI. Participants felt that a high-level push from these bodies increases local engagement and buy in from clinical teams.

"… something trickled down from up high from the college that would be a massive advantage even to the trainees because the trainees can drive this, it doesn't need the consultants and in many ways it's hard to change the consultants and they're just up to their necks in it quite honestly. The trainees and the fellows I thought would be biting my arm off to get involved in this but it might be that it's because it's coming from next door from a different department rather than coming down from up high from the college, that would be a massive advantage if you could do that because that'll be all the incentive they need." INT 11 (Anaesthetist, Intervention site)

The Royal College of Anaesthetists taking a leading role in evolving the field of Perioperative Medicine was seen as positive and had led to a change in clinician mindset about how patients should be assessed and cared for.

> *"Yeah, yeah absolutely I mean it, yeah I mean it's, I mean it's partly PQIP and part, partly of course, the overall ... overall Peri-Operative medicine push from the Royal College, perhaps that has also made a, playing a bit of a role but there's certainly a shift in mindset with regards to pre-assessment or assessing a patient or the way we manage patients I think, there's more emphasis on the patient experience side of things and the obligations definitely yes." INT 2 (Anaesthetist, Control site)*

Greater involvement and promotion of PQIP by the Royal College of Surgeons may encourage engagement amongst surgical colleagues resulting in more multidisciplinary QI.

> *"I mean I think a common problem for a lot of these projects is the ... is how the surgeons here are informed about it. They're quite a traditional bunch like that. So, if the Royal College of Surgeons ... I think for emergency laparotomy the Royal College of Surgeons very nearly embrace that as their major project. But I think they then went with gallbladders or something in the end. So ... but at the same time they are now, I think quite well engaged with emergency laparotomy and therefore that comes out of the Royal College of Surgeons.*
>
> *And I think the same for PQIP, if the Royal College of Surgeons can push it a bit more and recommend, particularly to its trainees and obviously to the consultants that they embrace and use this data more, and are they fully aware of it? Because I think to be ... even the surgeon who was so negative at the beginning is now actually ... and didn't want any of his patients included. So, for the first six to nine months or even maybe the first year none of his patients are in there but when he started seeing the data he now wants his patients in there." INT 3 (Anaesthetist, Control site)*
>
> *"I don't know, I don't know how much support you've had a level from the College of Surgeons, whether there's any information goes into surgical journals or surgical newsletters that sort of thing, may become ... because I think that would help if it came through their own channels rather than just through saying I need to keep steering and telling them about various projects." INT 8 (Anaesthetist, Control site)*

Other national projects, perceived to be led by anaesthetists, took time to bed in and receive the full attention of surgical colleagues. It was considered this process was likely to occur with PQIP and that given time it would gain wider acceptance.

> *"NELA crept in and no one's said anything against it by and large. It was done by the anaesthetists but the surgeons looked at it and said alright, hang on a minute, the data is there, we need to improve. I suspect PQIP would take on that evolution at some point as well. So, I suspect on a sort of national importance you guys should be working on that specifically. Of course, it all depends on what they generate on a national collective level…*
>
> *I think whether surgeons accept where it comes from or not is probably irrelevant. I think it's probably going to take on a very national importance really." INT 1 (Surgeon, Intervention site)*

The visit of an external organisation (Getting it Right First Time) to one trust acted as a stimulus for investigators to engage with their data and to explore QI opportunities.

> *"We had the Getting it Right First Time (GIRFT), the information from PQIP was presented. So that's ... I'm hoping that's going to be, yes, there were a lot of things said in that meeting, apart from PQIP information. But there are a few things that they want to act upon, like the preop carbohydrate loading, for all patients, and then getting how the information is actually gathered to reflect clinical practice. So those things hopefully, I'm hoping that it will change." INT 6 (Anaesthetist, Intervention site)*
>
> *"… because of the whole idea of the GIRFT was obviously to improve outcomes and care and things, so it was all really relevant. So that was really good, because that was purely PQIP stuff… it was up to date, relevant, for everybody in the room and yes, so that was quite a positive thing really…*
>
> *I think for those who were in the room and I think they could see definitely, see the benefit of it [PQIP] yes, and I think because they were all there for a common aim, and PQIP was very relevant to that because it gives them the information they need." INT 5 (Research nurse, Intervention site)*

The potential for data collected in studies such as PQIP to feed into evidence presented at external visits was also discussed.

> *"I guess so what matrix [of complications/morbidity] do people want to know, what's useful for the trust. What is useful for CQC kind of reports and things like that, because you want to be producing something that you can show as a trust to any regulator, show that you are doing okay, whether it's the Royal College, the CQC or whoever, the NHS whoever it is. So that would be my view, and yeah so I think the matrix are … a lot of them are very useful, but there is one or two I think that … and maybe they could expand the scope of the matrix as well." INT 15 (Surgeon, Control site)*

Shared learning of successful improvement projects

The sharing of successful improvement work from sites recruiting to the PQIP study was suggested as a facilitator of quality improvement. There was an acknowledgment of siloed working within the NHS resulting in different institutions going through similar improvement cycles. Facilitating shared learning between sites about how and why improvement project did or did not improve patient care was considered potentially beneficial.

> *"I think it might be worth trying to find some good news stories of places who've done something or done something with their data and made an improvement or places who've managed to get funding and then also had a success in improving their outcomes." INT 12 (Anaesthetist, Intervention site)*

> *"Things like knowing ... it's like if ... positive deviance stuff is really good but it almost needs a bit more detail to work out why someone is ... it would be nice to know why someone is a positive deviant. So, to get maybe something I would quite like to see would be a small little report to say from ... so maybe the PQIP person in a hospital to say why they think going on a PQIP positive deviant in that area. So it gives people ideas of maybe things that can be done to improve. Because a lot of these things are quite simple sometimes. The changes that individual undertake that somehow seem to make quite a big difference." INT 8 (Anaesthetist, Control site)*

"Perhaps if other hospitals are doing QI projects and they're going quite well, maybe including in the email when you've done the quarterly reports, examples of what the hospitals are doing, because maybe give an example of what other hospitals are doing, might inspire a little bit and give people ideas." INT 10 (Research nurse, Intervention site)

Local networks were suggested as a solution to address the issue of siloed working. I Participants felt these networks could facilitate regional QI activity and increase the success of implementation.

"… I think it's just, it'd be good to see, to share case studies. I saw on the website, was it Salford or one of the centres, shared a case study, as to what improvements, I think that it would be interesting to obviously it's a National study, is, it's that networking side of things, it's learning from other people are doing, and whether that would be of use, because rather than think up our own changes that we need to make. If there was some changes that people have done, up country maybe, that have made a difference to share that, rather than reinventing the wheel every time." INT 13 (Research nurse, Control site)

"I think it will probably just happen [develop some local networks] but we're quite competitive. I understand PQIP don't go for the negative, it's go for the positive as a national principle which I completely agree. I think locally and because we're all good I think the competitive you know we are linked stays better than yours is a driver in particular for the colorectal where you go it's the same procedures. So, for thoracic not everybody does thoracic, hepatobiliary not everybody does hepatobiliary but everybody does colorectal." INT 12 (Anaesthetist, Intervention site)

The impact local networks could have was not limited to just clinical applications, but may also support clinicians to obtain parity with local colleagues in relation to time allocated to undertake this important work.

"So, for example I know [other NHS organisation] there are specific PAs for the PI and PQIP so I think that gives you, when you have your job plan review that gives you an opportunity to say look this is a really important thing, this is how down the road somebody gets a PA for that." INT 12 (Anaesthetist, Intervention site)

Some areas had already started to develop links between local hospitals.

"The other thing about PQIP is that we are maybe using PQIP as a basis for a regional QI network. So, we are starting to get together with at least one other Trust who is interested and will hopefully suck in another two trusts which are in our region, to start looking at well how are you doing with these and how is everything going in terms of your PQIP outcomes. Just as a starting point. And say well what have you done about yours and could we learn from that etc etc and share some of the learning and maybe try and standardise some of the practice across the region if we can crack it, that would be amazing. A big goal to set yourself. But that is what we are trying to do." INT 15 (Surgeon, Control site)

Communication from national project team to senior trust management

Direct contact between the central PQIP team and local senior management was generally viewed as positive and helped to raise the profile of local work. Participants felt this external communication had a greater impact than local investigators themselves trying to engage colleagues.

"I think one that did wake things up was the sending of the annual report to the chief executives and the medical directors because that suddenly got the medical director wanting to know what's going on, if I send something to my medical director it goes in his bin pretty quickly." INT 12 (Anaesthetist, Intervention site)

"What was helpful from one other study which was PREVENT I think, one of their research fellows came out and spoke to us at an audit day, just sort of for half for an hour. Because we'd done some local presenting on it and bits and pieces but actually having someone from the centre conducting it coming out seemed to spark a bit more interest… That seemed to raise a bit of awareness because people understood what was going on.

Would that be helpful? Yeah, it probably would but it's quite an amount of work for you guys to do, but I think that would give the management an idea of what we're up to and why it's important and the fact it's not some sort of local whim. It's a national study that's potentially ... well should guide a big improvement at a local level as well." INT 6 (Anaesthetist, Intervention site)

This direct contact could raise the profile of the work local investigators were doing and potentially support their requests for additional funding and support.

"I mean our research team here, the research nurses work phenomenally hard on all of this without any ... we've basically got one and a half nurses involved in this and they've got other projects themselves as well. And actually trying to get additional funding for them to get the appropriate time they need, and they're doing it well as they are but they are rushed off their feet. And having someone from the centre themselves coming out and showing what we do to the management might just give them a little shove to try and free up a little bit more time for them." INT 6 (Anaesthetist, Intervention site)

"… it makes me wonder whether we can ask study teams, such as yourself, to inform our, if it's not the chief exec, but some of the higher management levels in the trust about our participation. Because I ... whether actually study teams directly approach, emailing the higher management and saying thank you for your trust participation, we are now on this list of blah, blah, blah. At which point they'd probably look at the email and go oh really I didn't know we ... but it's sort of ... I guess it might help." INT 4 (Research nurse, Intervention site)

Time allocated in job plans to undertake QI work

Participants frequently discussed the pressure they felt to deliver clinical work and few had time specifically allocated to improvement implementation. Many felt that allocated time was key to support them in delivering high quality, sustained improvement in patient care. Without such time, many felt they moved through improvement cycles with little success. Financial incentives were seen as one way to influence trust priorities.

"Yeah, I think yeah absolutely. Time is, time yeah, time involvement will definitely you know, improve things and if only you know all this became you some kind of a best practice standard, that would straightaway focus everyone and ... because then I think accountability comes for everyone, clinicians, managers and everyone but yeah, it will help, it will help… I think ringfenced time will help" INT 2 (Anaesthetist, Control site)

Participants were interested in improvement work and felt that if people were afforded time in their day-to-day work to deliver work, it would be result in positive change.

"I think the people that gain that time back would want to do something with the data, whether they'd be allowed to, as in would they then get peeled off, want to do something more clinical that would be the battle I think. But I think most of the people that are interested would want to stay and do something with it. It's whether they'd be ... you know what it's like you have half an hour back because you've done something quicker, you have to do something else. Rather than being given that time to do something that you're actually interested in. INT 6 (Anaesthetist, Intervention site)

The need for protected time did not just related to senior clinicians but also non-clinical staff and trainees.

"They were yeah, they were really proactive, and I think they possibly had the time to be able to do it, because they were a trainee, I think it needs a lot of time dedicated to it, to make, you've only got to make small changes but I think you need to be quite a high profile in the Trust, and liked, and just needs somebody with a high motivation and a high time ability to drive it forward, to make those small changes." INT 13 (Research nurse, Control site)

"And what I'm trying to see is whether we can actually get a bit of dedicated time out of named members of staff in the general management team to help actually drive the changes that come out of the collecting and the showing of the data. So that's my plan for Monday morning…" INT 3 (Anaesthetist, Control site)

Trainee engagement in data collection and QI implementation

Surgical and anaesthetic trainees involvement in PQIP reduced the data collection burden on research nurses and senior clinicians and led to successful QI implementation in some sites with trainees leading specific projects.

"Yes, it's been me with a team of research managers who've been absolutely fantastic and they're the only reasons it's ever got up and running and kept running and the trainees which have helped with weekend data collection and that sort of thing." INT 11 (Anaesthetist, Intervention site)

“So we've got three, four trainees, who are part of the PQIP team, and one of them, very early on, we sort of thought we should be trying to improve our pain scores, because I think we got 21% severe pain, in recovery, although still very high satisfaction rates.

So, we thought that's potentially one of the things we can do. As part of that we're standardising, not standardising, we're producing general SOPs or guidelines for anaesthesia type, and they want to look at the pain scores more accurately really, than what the ... you can really get from PQIP study in terms of linking it to the types of anaesthesia. Getting in mind all that data themselves, from the database, and the downloadable components of it.” INT 14 (Anaesthetist, Intervention site)

“So, I think that was highlighted in one of the reports and then the trainee with me, they went through the data much more meticulously and break it all down and looked at it. Then we presented that and then took it from there…” INT 3 (Anaesthetist, Control site)

“I've got an F1, not an F1 a CT1 just started now, [they’ve] done an actual ... [they’ve] done a year of [quality improvement training]…

So [they’re] here now to help me with PQIP. So I'm hoping that will be a good stimulus to get things ... to get looking at some more information and get some... QI projects out of it but I've been trying to sort of get them ... I've been trying to get them to look at.” INT 8 (Anaesthetist, Control site)

“Well actually very interestingly, there was, there’s a Registrar here, who became involved in PQIP a little bit at the beginning of the year, and then [they] used the information or looked at the information from PQIP and did [their] own Audit on Pre Op Carbohydrate Loading, because I think it was showing that we weren’t particularly good at that. So [they] looked into that and investigated, and [they] did present at the Anaesthetic meeting and fed that back, and [they] also, there was a big meeting on Friday, which it was for the Colorectal Surgeons and the Anaesthetists, some ward staff were there, I’m trying to think what that was called, it was the GIRFT [Getting it Right First Time], so they went through that. So anyway.. so basically [they] presented her Audit and relayed it back to PQIP obviously for the drinks which was great, and is going to look into how things can be changed…” INT 5 (Research nurse, Intervention site)

A culture of quality improvement within an institution

A culture of improvement and institutional drive to enhance patient care was considered a facilitator or QI activity. One participant reflected on the presence of 'QI days' within their trust, where traditional audit or clinical governance meetings had been renamed. They felt this had resulted in a shift of focus away from audit towards quality improvement.

> *"Well, I think we have got a dedicated QI days, QI days you know the clinical governance or quality improvement half days we call it every month, half day as it is in many trusts... So, I think from the research side we are getting adequate support, and QI it's mainly down to the initiative of consultants and colleagues really. So, whenever there is a QI we are supported generally and encouraged." INT 2 (Anaesthetist, Control site)*

> *"No, it's always, I mean last, I don't know, six, seven years, I think the Trust changed the name of Clinical Governance meetings to Quality Improvement half days, that's all…*
>
> *I think the terminology change did help because lots of people were doing stuff which were not really quality improvement, so there's a bit of a, sort of rigorous reporting but because the quality Improvement agenda, the half-days and action plans have to be reported to the Trust which goes to the E-Governance and Trust Department and so there's a bit of a Governance that are passed to ... Governance and Audit as to what activity is done." INT 2 (Anaesthetist, Control site)*

The same trust had also established an annual quality improvement competition, where projects were presented and prizes awarded. Recognition of local investigators effort at the local level had a positive impact and acted as an incentive for trainee engagement.

> *"Our Trust does have a yearly sort of Quality Improvement poster project and a competition to ... wide open to the whole of the Trust and they give some prizes so with a bit of financial rewards, small rewards for the Department… That, that's always been there and yeah so we, we ... last, last year we got the second prize in the Trust...*

Yeah, yeah that's a hospital-wide presentation day, where the Chief Executive comes and sort of deliver, delivers the prize, if you like. So, there was a keen trainee involvement because that's obviously a Trust-wide recognition." INT 2 (Anaesthetist, Control site)

The identification of quality improvement leads within the department led to feeling of trust support for QI work.

"Well, we have a ... I think the regional QI lead is one of my colleagues. So, QI is heavily backed I believe, by the Trust. It's certainly seen that methodology to improve patient outcomes, patient welfare, etc., as something that it wants to invest in." INT 14 (Anaesthetist, Intervention site)

And there was an acknowledgment from participants of the need for dedicated staff and funding to support QI activity.

"… well it would all have to be to do with funding really, because, or like extra staff to help with Quality Improvement, I can't see things being done without extra staff being around just to assist with the Quality Improvement and implement it." INT 10 (Research nurse, Intervention site)

Support from management and divisional leads for quality improvement

Engagement and support from divisional leads and non-clinical management staff within surgical and anaesthetic directorates was viewed as a facilitator of improvement work. Where principle investigators felt supported and had identified interested colleagues improvement activity was generally more successful.

"I mean I would be leading it mainly because I have got the interest and the stakes there. But the clinical director is fully supportive of this, and so what we have done is sort of we had the trainee induction last week, so we have included that in the trainee sort of induction pack involvement in that…

But overall you know we have identified a few colleagues who are keen at getting involved. Some of the things we … especially the pre operative risk assessment we were able to straightaway agree on a few things and come up with a strategy, and that's sort of … we are implementing that

straightaway. Other things like HbA1c is pretty straightforward and anaemia pathway and prehabilitation, that's going to take months. That's slowly progressing." INT 2 (Anaesthetist, Control site)

"Then I approached the colorectal surgeon… and he was happy to be the lead and he, to be brutally honest, he is very helpful at times. When something ... like come Monday I've sort of pre-warned him that one of the things that I'm going to be pushing for is the management team to, and the surgeons, to take on the anaemia issue that PQIP has highlighted. And I think he will give significant support at that meeting because he's the CD for surgery for that." INT 3 (Anaesthetist, Control site)

"By and large from a national viewpoint we're one of the best run trusts and the quality speaks for itself, we have our challenges. But when it comes to taking quality and focusing on the data capture, if we wanted to introduce something else, for example the trust are very, very supportive. We have a very good relationship with the managers and in terms of quality they are very, very focused on." INT 1 (Surgeon, Intervention site)

Education in improvement science and experience of delivering successful QI

Participants who had undergone QI training or had experience of delivering successful improvement work found it easier to engage colleagues and lead QI activity.

"I have no official status in terms of title but I'm continuously doing forms of quality improvement and I'm sort of an unofficial, I don't have a title but I'm the person people go to if they're doing any perioperative medicine or surgical pathway improvement." INT 12 (Anaesthetist, Intervention site)

"We had a number of Trust fellows who were QI fellows, who are involved in Trust wide projects, and certainly I don't know the structure behind that, I just know it's around, because a colleague of mine, so I think [they're] the regional lead, but [they're] certainly the Trust lead for it.

So [they're] very much involved in that side, educating some of those fellows, and they actually go through proper courses etc., for that educational sort of screens, so they can do the QI work properly." INT 14 (Anaesthetist, Intervention site)

The need to build capacity to deliver QI within organisations was identified as important.

"Actually, what the capability to actually improve, is not there I think in most organisations, most Trusts, most departments. So, I think what needs to be bolted onto a lot of these massive laudable efforts that are going on to produce data for people and show you what's going on and show you where your problems are is to increase the capability to change. Which sometimes might need more resource and sometimes it won't do, it just means you've got to, you know you've got to introduce that knowledge and capability into that organisation or into that department…

So that's the first thing I would say is that where's the capability to back up the areas that you've highlighted to say okay we will now be able to support you to make some changes because this is how it was done in Trust A and B and look, there's a methodology." INT 15 (Surgeon, Control site)

Development of QI teams

The establishment of local teams, either to work on specific projects, or to ensure the day to day running of the PQIP study reduced the burden on individuals and resulted in positive change.

"So when we looked at the data and tried to understand it, we also looked at, or we surveyed the general or colorectal anaesthetists about the preferred analgesic technique and found actually that, particularly for laparoscopic colorectal work, and found that there was quite wide variability. And one of my colleagues had done a separate audit actually, not using PQIP data and had done a literature review and felt that we should be increasing our spinal opioids use and dose. So that's what we agreed amongst the five of us that we would do. And so we've got a more uniform analgesic technique of spinal diamorphine at higher dose for laparoscopic resection and that that seems to have made a difference. But it all stemmed from the data coming out of PQIP." INT 3 (Anaesthetist, Control site)

"Yes, so a couple of colleagues of mine, I wasn't involved in the initial anaemia pathway, but they have been setting that up over a number of years, probably before and then alongside PQIP before it became a sort of national priority within PQIP. But it was struggling to get any patients down the pathway and certainly PQIP has helped them. So I get regular requests from them to share the PQIP data with them to see how their pathway is going. So again that's increased the awareness of PQIP and the information in it is trusted because it's on the whole collected so carefully, and it's a clinical dataset. So yes, that is, that is working quite well." INT 3 (Anaesthetist, Control site)

Comparison with external sites

Providing a comparison of performance against external sites was considered a potential drive to improve. Some participants felt that being able to compare their own results with similar institutions would be helpful, either locally or more widely. It was acknowledged however, that if a hospital was delivering processes at or above the national average this may limit improvement activity.

"I mean the comparison, the national comparison is very helpful it tells us where we are, you know compared to national levels and we can sort of try and work out you know where there is scope for improvement. And where we consider … can look at ourselves and say, look okay we are doing a reasonable job because we are at or above national standards. I mean that's a useful thing, the national standards and the comparative graph with the national standards." INT 2 (Anaesthetist, Control site)

"From the graph [VLAD] perspective, you know, I think although it's very simple, I'd say that's one of its strengths. It would be useful to know where we are in comparison to other places, and I guess you could talk a little bit about maybe moving in that direction, you know, the final plot type idea. That's always quite useful, because there's always confidence around these things." INT 14 (Anaesthetist, Intervention site)

"So either national or if you want to sort of band Trusts into sort of DGH's and major teaching hospitals or something. So you are comparing like with like it would be quite interesting." INT 4 (Research nurse, Intervention site)

"I think what's transpired from it is that we have been quite reassured that we are not an outlier in terms of the outcomes that you are measuring on PQIP. I think it's been quite reassuring to know where we stand basically as a Trust and where we stand compared to our peers in other Trusts etc. So that's been one very useful thing because you need to be confronted with that data. But then it has highlighted a couple of areas where we do need to do some work. What you then need is something around that to say well yes he identified there were two areas that you need to improve on. The whole point is improvement but then it needs people with improvement methodology who can then undertake whatever it is that you need to do to sort it out." INT 15 (Surgeon, Control site)

Identification of poor performance was considered a drive to improve, either through comparison with national averages or identification of 'outlier' trusts using statistical methods.

"In other words are you saying if we got wind that … if the reports were showing that there were areas that needed improvement would we act on that? Absolutely." INT 4 (Research nurse, Intervention site)

"I think personally I'm not uncomfortable with that [outlier identification] because if we're an outlier I would want to know and then the next is actually we've probably, our bit that we've not used this for yet is you know we're not the best now at length of stay for say our elective colorectal. So you know at some point we need to use this to go to our surgeons and go so what is it we're doing you know, what bits of this and I think it's if you have a confidence interval that was benchmarking, I know we're not allowed to use the word benchmarking in PQIP, I think it would be of value." INT 12 (Anaesthetist, Intervention site)

Collection of novel data

The availability of data that was perceived to be interesting and novel, such as postoperative thirst, acted as a facilitator for QI.

"I think from the anaesthetists the things that people have found interesting is the amount of thirst, temperature in recovery so I think those are the sort of simple things that we weren't looking at beforehand." INT 12 (Anaesthetist, Intervention site)

"I think it's locally it has just been reporting them getting used to 6-monthly reports and audits of the PQIP data, and then people thinking actually you know are any of my patients in there. Because the performance is, for several of the measures, some of them aren't immediate things a real problem for us but for the DrEaMing has been quite good so they like to see that obviously because that on the whole looks quite good. They like to see their patients in there and included in it. But it has, the PQIP data has definitely been positive in driving our initiative to improve anaemia and it highlights it regularly that it is a real problem. So we have a, we do now have a referral pathway for anaemia and IV Iron is established. Though our problem still remains that patients are identified way too late for it to impact on their preoperative haemoglobin. So the number of patients who are having IV Irons is slowly increasing." INT 3 (Anaesthetist, Control site)

Availability of real time data was also considered to better support QI.

"I have said data that only appears once every blue moon will always have a limited impact. If you produce data for people that's real time, there's up to the minute and they are faced with it every month then that is something that you probably will find that needs to change a lot faster than reports that only come out every so often. But that's life, do you know what I mean, we've got national bowel cancer and all sorts of big reports that come out every year. They do only come by the nature of it, they do only come out once a year and you have to accept that but it is not the optimum way of doing it." INT 15 (Surgeon, Control site)

Involvement of research nurses in data collection process

The role of research nurses in patient recruitment and data collection was viewed as positive and vital in successfully delivering the PQIP study.

"I mean, the benefit of ... for a local Trust, benefit of PQIP is the fact that you're giving us an easy way, really, of collecting all this data that hasn't been collected before. And yeah, being sold as research, so we get this CRN funding, so we have research analysis, doing the questionnaires, and some of the data collection." INT 14 (Anaesthetist, Intervention site)

"I think the Research, I think the help with the Research Nurses is really, really useful, as they're the ones, they're the backbone really, they're the ones doing most of the work." INT 2 (Anaesthetist, Control site)

"I mean our research team here, the research nurses work phenomenally hard on all of this without any ... we've basically got one and a half nurses involved in this and they've got other projects themselves as well. And actually trying to get additional funding for them to get the appropriate time they need, and they're doing it well as they are but they are rushed off their feet." INT 6 (Anaesthetist, Intervention site)

Despite their involvement being valued by clinicians, research nurses did not always feel their role was valued within their organisation.

"… the Trust is not hugely supportive of research, we're underfunded and understaffed and a bit stretched…

It's utterly frustrating from my end but our funding has been reduced and our workload has been increased. And one of the what, looking back on it was probably a luxury, is being able to keep an eye and monitor our PQIP performance and everything has had to go by the wayside" INT 4 (Research nurse, Intervention site)

Financial incentives to direct improvement priorities

Financial incentives have the potential to facilitate improvement activity through determining institution priorities. However, incentives provided in the care of some patients, through the use of best practice tariffs for example, may exacerbate poor care and limit improvement activity to other groups.

"…they've approved a new Data Collection Nurse because we were asking for a Data Collection Nurse for nearly five, six years without much success really and the NELA BPT straightaway sorted it out, because they were able to fund it because ... so that's what happens and as always they tend to fill the template, you know, the Fracture Neck of Femur Data Collection Nurse and that happens unhindered because we've got £1,300 extra for each patient which is the BPT (best practice tariff)…

Yeah, I think yeah absolutely. Time is, time yeah, time involvement will definitely you know improve things and if only you know all this became you some kind of a best practice standard, that would straightaway focus everyone and ... because then I think accountability comes for everyone, clinicians, managers and everyone but yeah, it will help, it will help." INT 2 (Anaesthetist, Control site)

5.2.8 Participants perceptions of their role in quality improvement

Delivery of quality improvement was often considered to the responsibility of clinicians, with the onus often falling on anaesthetists in the setting of PQIP and pomVLAD. Participants expressed frustration and sometimes guilt at being unable to deliver successful QI.

"The QI we have started it's all anaesthesia related." INT 11 (Anaesthetist, Intervention site)

"I mean I would be leading it mainly because I have got the interest and the stakes there." INT 2 (Anaesthetist, Control site)

"I think part of it's my ... I think I will be partly to blame in not going after them more and saying look at this data. Because that's the thing I'm trying to get across is I'm PI for it but I'm not ... I don't have time with everything else I do to actually then start creating a few [QI projects] that I need other people to collect the data on" INT 8 (Anaesthetist, Control site)

Due to time constraints clinicians often felt unable to deliver QI work themselves and instead saw their role as one to engage and stimulate colleagues to implement QI.

"I want to get the surgeons, I mean in my mind this is just tailormade for surgeons to do QI and I haven't got the time or don't think I should be the one who's doing all the QI side of it I can't possibly take it on so I'm just waiting for the opportunity to get surgeons involved." INT 11 (Anaesthetist, Intervention site)

"But I think just looking at the data doesn't achieve anything I think it's going to be using it as I say for demonstrating to the ward nurses so showing it to the ward nurses or showing it to the surgeons that are meeting to then reciprocate an okay we need to look at this." INT 12 (Anaesthetist, Intervention site)

QI activity was not considered to be part of research nurses' role. This led to a disconnect where those most invested in data collection and data quality were not involved in using data for improvement.

"My main focus is the data collection, because that's how I see my role, I'm obviously not an Anaesthetist, I'm not a Pain Specialist Nurse, I feel that we need to give them the information, and let them make the improvements. So I'm obviously, we're the conduit collecting the information that we need to do something with and it's a bit frustrating because I'd like to see things change, not that it's my field of expertise, but if you're doing a study, it's going to improve things, you want to see improvements, so it's a little bit frustrating." INT 13 (Research nurse, Control site)

"I mean I think, to be frank, I mean obviously as a research nurse you could make the argument that did I ... is it even necessary that I was ever aware of it, to be honest. Because I can't implement it and I can't ... as I said to you earlier, it was interesting from my selfish perspective to just be appraised of what our data is showing. But actually if you were being really callous about it, did I ever really need to know? And I'm still ... the little bits I do glean about the data is that the clinical team are looking at it. I mean I think as a trust, and as ... the clinical team who are in a position to implement improvements are absolutely looking at the data." INT 4 (Research nurse, Intervention site)

"I don't think we've got the capacity to do that, because Quality Improvement isn't research per se, I don't, I know it's all part of the clinical effectiveness cycle, but if I took time out to actually implement the Quality Improvement side of things, I wouldn't get the research done, so I think there needs to be some ringfenced time, somewhere along the line, for the Anaesthetist and their trainees, to take on the QI side of things." INT 13 (Research nurse, Control site)

"yeah, when they were attached to the emails I used to look, go through them. But I know they put highlights on the emails that tell you the main points that have come up for your site. I do read those but I don't, if I'm honest, I don't have time to go on and check the reports. If I was involved in a QI project I probably would but because I'm mainly responsible for running the trials and not really doing QI from them, I don't really feel the need to." INT 10 (Research nurse, Control site)

"…all the quality improvement is what the doctors work on because they're trying to improve their practice with PQIP whereas we're just gathering information. And we aren't changes practice in that sense. I know we're collecting the data and helping to, but it is the doctors that are running with change in practice. So I don't think it would benefit us." INT 9 (Research nurse, Intervention site)

Despite quality improvement work not being directly relevant to their role, some research nurses expressed interest in implementing QI if their responsibilities allowed time for this work.

"Yeah definitely, I think if there was time to do it [get involved in quality improvement], we probably would be very interested in doing it." INT 10 (Research nurse, Intervenion site)

5.2.9 Recommendations for improving the pomVLAD dashboard

Easier data export to enable own analyses

Participants discussed their difficulties obtaining raw data to analyse for local QI projects. Improving the ease of data access would particularly support QI for data not currently reported within pomVLAD or elsewhere in the PQIP study.

"Well going on the reports we produce a couple of times I've allocated an hour to sit with your users guides about how to do my own data query and I haven't quite got there yet. I get to a certain point but I've found it quite difficult and it's on my to-do list is to phone up and get some kind of action on-line advice because I've not managed to make it, I don't find it very user friendly at all because there's specific data queries so I find it really tricky…

it's not intuitive at all. So what I want to do for the surgeons is stand up and demonstrate how to do it but I've got no confidence that I'll be able to do that at the moment so it's going to be a hard sell but I will try and get that sorted out." INT 11 (Anaesthetist, Intervention site)

Including the number of patients as well as percentage

Due to the small number of patients recruited each week in some sites, participants felt that making the denominator clear for percentage calculations would help understanding of large changes in apparent performance. One potential solution suggested to these large swings was the use of moving average charts.

"I think the only thing I can't find in the dashboard is our total numbers, there's lots of percentages but no, so that's a percentage of what.

I wonder if that could go in your basic demographic. So you've got acquisition cases per week, I could add them up but I'd have to go out of the dashboard to find the total number and then also I'd have to look at what are complete. That's the only thing when I looked at it I thought I would've liked to have seen." INT 12 (Anaesthetist, Intervention site)

"Because if it's small numbers, we're doing just colorectal, things can bounce around hugely month by month or certainly week by week. So that sort of moving average is a really important component of it I think." INT 3 (Anaesthetist, Control site)

Comparison with local sites

Participants reported that visual comparison with the performance of local sites would potentially improve dashboards and stimulate engagement. This was considered to be particularly relevant in the context of establishing local networks.

"I think these are things that you need to get people to want locally so I think you should probably be open to something like if say a group of five hospitals go can you give us a dashboard that compares us all I think that's something to think about but that's got to be driven by people wanting it locally." INT 12 (Anaesthetist, Intervention site)

In the setting of pomVLAD one investigator felt that providing a comparison may support interpretation.

> *"Because the one thing that is good about reports is the comparison for the national average as well. I mean this dashboard for the pomVLAD is really good but then I had to go elsewhere and look at other documents to find out sort of how that compares with the national rate but that's what's missing from this dashboard at the moment. If we can get context…*
>
> *Well in that case, if it's an option, if it's not set in stone in its current format it would be fantastic. Even if it's one further click, see your own data and then there's some way of looking at that comparison with national or like … similar Trust sort of data would be really helpful." INT 4 (Research nurse, Intervention site)*

Ability to customise dashboard for local priorities

The ability of local investigators to customise dashboard displays to display local improvement priorities was suggested as a way to improve dashboard utilisation.

> *"What is a lot more useful for us is knowing that every anaemic patient is at least being given one IV infusion or being considered for it. But that type of data we can't get, you know, a separate local project. So that, just as an example, okay, a lot of their stuff is what's coming through on the posters and the top of the reports. But it's not that helpful for us.*
>
> *So ... but I guess every Trust, you're going to have, we appreciate are going to have different agendas and you can't really clearly produce reports that are specific to the agenda of that Trust. That's just a bit of generic feedback, is what they say ... aren't that useful…*
>
> *But yeah, if you have that customisation ability, to enable us to look at a few add-ons, which are, as I said, you know, bit more specific to our local agenda. And also plough almost you know, PQIP pretty reports, that aren't more specific to what I want to be presenting to my local department, you know, that's going to provide a lot more usability, and make it far more relevant." INT 14 (Anaesthetist, Intervention site)*
>
> *"I have ... yeah I absolutely ... I'm sure that it [customizability] would. Again, mainly because you're then selecting the markers or the metrics that you feel you can, or that you need to improve upon. And it then may*

also allow you to change them over time such that you can focus your efforts for a given period on one thing that you decided that we need to change. And once you feel you've got a good handle on that, you can then move on to the next thing as it were and tweak your dashboard to allow you to reflect that. And then see how you manage that with time, rather than potentially staying with the same one the whole time." INT 7 (Anaesthetist, Intervention site)

Provide ability to identify individual cases within dashboard

Another improvement suggested by participants was to improve the way individual cases are identified within the pomVLAD dashboard facilitating easier local investigation of morbidity and mortality after surgery. One investigator also felt that identification of high-risk patients as data are entered may improve care.

"Yeah, so if in the pre op if we input data and there is a flag coming up, say for example of … flagged up this patient, is high risk and is likely to do such and such a thing, I mean the research nurses are there for data collection, then they are not at the clinical input. But if there is a way to feed that back to the clinical team on a live basis, that would be fantastic. But it's probably asking for too much, but that would be great…

Yeah, the, the, the only sort of logistical tricky, difficulty we have is because I think when we have to identify patients, we have to go by the name really because you know the ID, the ID tracking number in your PQIP is different to our hospital ID." INT 2 (Anaesthetist, Control site)

Display constituent parts of DrEaMing outcome

Reporting performance of individual components that make up the DrEaMing outcome (drinking, eating and mobilising) was suggested as a way to improve understanding of local data. Some participants reported that performance in delivery of DrEaMing was often heavily influenced by a single component, and without displaying all three it was not clear which this was.

"And I think it was a bit confusing because I can't remember. I can't actually remember the ... something like, something was placed over something else, and it would be easier if it was extricated and you could see the individual component of the thing.

Yeah I think that was the dreaming part wasn't it? I think, yeah" INT 4 (Research nurse, Intervention site)

Improve information about risk-adjustment process and VLAD interpretation

Despite many investigators feeling the VLAD display was intuitive, some participants wanted additional information about the risk-adjustment process and interpretation of the VLAD. Improving the written resources available was suggested to increase local understanding of what trends in the VLAD may mean.

"I sort of get the feeling it's in development. I mean, after my conversation with you, I was a lot happier. I mean, I guess you have only got 10 pilot sites, what you could do is actually get up a pilot site specific explanation of it.

So you could critique the graph. I mean, from an educational perspective, there's probably a little bit more you could do around your blog website, in describing how it's calculated, where the original dataset and risk prediction came from. How you interpret the graph generally. I mean, I think you've done a good start, but I was still left with questions that you answered quite easily and very thoroughly, so clearly the knowledge is there." INT 14 (Anaesthetist, Intervention site)

"And then it comes down to specific graph is saying well, I see our trend is going in that direction, you know, what does that mean, how should I interpret that, that would be the next steps? And you've only got 10 pilot sites at the moment, I'm just giving your work a test, but I don't see it outside the realms of possibility that you should get slightly more site specific feedback to each site, as part of the project, so we can interpret it better." INT 14 (Anaesthetist, Intervention site)

Another possible solution to aid interpretation was the addition of control limits or confidence intervals.

“From the graph perspective, you know, I think although it's very simple, I'd say that's one of its strengths. It would be useful to know where we are in comparison to other places, and I guess you could talk a little bit about maybe moving in that direction, you know, the final plot type idea. That's always quite useful, because there's always confidence around these things. So I now look at a chart and go "oh, we'd got worse, and now we're getting better and better and better and better". And that could be because the risk prediction is slightly wrong. It could be because we are getting a bit better, but actually, you know, if you put that in the context of confidence intervals, you know, we could actually be ... that trend, it could be no different, it could be within the realms of possibility or variation in competence, etc., that it's an area we should actually look at, and should still consider going along the zero axis if you see what I mean…

You know, is it a clear difference? So we have to care about it now. Or is it a trend, and that's a term I hate, to significance? But then when you get a number of dots, and there's a persistence of time running, you know, above where you'd ideally want to be, and you've got to start looking at it that little bit more.” INT 14 (Anaesthetist, Intervention site)

Chapter 6 Discussion and conclusions

6.1 Summary of findings

Despite quantitative analysis not demonstrating a significant change in the primary outcome, risk-adjusted Postoperative Morbidity Ratio (PMR), the qualitative analysis revealed that local investigators, which included anaesthetists, surgeons, and research nurses, generally perceived pomVLAD positively and valued the real-time feedback of patient outcomes. Participants identified several implementation challenges that might reduce the effectiveness of the monitoring tool, including difficulties in accessing and interpreting data, concerns about the accuracy and completeness of data, and a perceived burden associated with using the monitoring tool. Local investigators also expressed difficulty in engaging colleagues with their QI endeavours as well as finding the time to deliver QI activity amongst busy clinical workloads. In order to successfully implement a large-scale monitoring programme many of these barriers must be considered at the design and implementation phase.

I will now discuss the findings in the context of the research questions set out in Chapter 4.

6.2 Discussion of findings in context of research questions

6.2.1 Research question 1 and 2

Does the implementation of an online dashboard reporting risk-adjusted postoperative morbidity outcomes and perioperative care recommendations in near real time result in a reduction in the incidence of postoperative morbidity in NHS hospitals recruiting to the PQIP study?

Does reporting of risk-adjusted morbidity outcomes and performance in delivery of a range of perioperative care processes result in greater compliance with those care processes?

The implementation of the pomVLAD dashboard did not reduce the incidence of postoperative morbidity or improve adherence to recommended perioperative care processes during the 12-month follow up period. Potential reasons for this may have included the relatively short implementation period and the capacity of local teams to respond to rapid feedback. Within the NHS there can be significant inertia to change, with the time taken for clinical policies to be approved or for procurement services to obtain and commission new equipment extending to a period of months or even years. An example of the potential impact of these issues was demonstrated when local investigators in one site discussed their efforts to improve intraoperative warming of patients. They discussed the challenge of moving away from one type of warming device to another that they thought would be more effective. Despite this drive to improve, there was difficulty in gaining approvals and funding for the new equipment resulting in no real change during the implementation period.

There was evidence that reporting of performance led to local discussion about potential reasons for poor adherence with the suggested process measures. Consideration was given to both data quality concerns and to true poor adherence. Positive performance also acted as a facilitator for discussion of results within sites.

Another factor that may have reduced the impact of the dashboard was the capability of clinical teams to respond to real-time feedback. Whilst local investigators generally perceived the rapid availability of outcome data as positive, their capacity to act on this data was limited. Clinicians frequently cited a lack of designated time in their day-to-day work to deliver meaningful change. Instead, the pressures of direct clinical care took priority. They also expressed a feeling of lack of support from management, especially when clinicians' priorities did not align with those of senior management or wider NHS priorities.

The DrEaMing process measure was one that particularly captured the interest of local investigators. Many were or planned to undertake QI related to it, reflecting success of its promotion as a national improvement priority by the PQIP study and the publication of final incentives for its delivery in 2023.[52] This accomplishment was in part related to a perception that DrEaMing was something local investigators were able to directly impact, and that it was within their control to improve with little external support. Implementing improvement did not require significant investment or wholesale change of care pathways which contributed to enthusiasm around the measure and its improvement.

6.2.2 Research question 3

Does the availability of near real time risk-adjusted morbidity data increase engagement of clinicians with their local data?

It was not possible to directly quantify user access to the pomVLAD dashboard using login data, so engagement could only be assessed subjectively through qualitative data obtained from the interviews.

The availability of near real-time risk-adjusted morbidity and associated process of care data prompted some discussion among healthcare professionals about performance metrics, data quality, and patient outcomes. In certain cases, clinicians examined discrepancies between dashboard data and their perceived

performance or internal records, which contributed to improved understanding of local data but did not consistently lead to wider engagement or action.

The dashboard's ability to simplify complex morbidity data into an understandable format supported clinician engagement to some extent, making it easier to interpret the data. However, translation of this insight into tangible changes in practice was limited. Some initiatives were reported, including proposals for medical equipment, minor adjustments to documentation, and small changes to perioperative protocols, but these were generally isolated and not sustained.

Despite limited improvement activity directly aimed at reducing postoperative morbidity, the dashboard was viewed by some as a useful tool for monitoring and audit, rather than a catalyst for broader quality improvement efforts.

6.2.3 Research question 4

What is the programme theory supporting the pomVLAD intervention?

The programme theory is grounded in the belief that providing clinicians and healthcare teams with access to near real-time, risk-adjusted morbidity data can enhance engagement with local data, stimulate QI initiatives, and thereby improve patient outcomes. This theory is supported by several key components and assumptions:

Initial engagement and awareness

The introduction of the pomVLAD dashboard at national meetings and through site initiation phone calls builds initial interest and awareness among clinicians. This initial engagement is crucial for sparking curiosity and encouraging the exploration of local data.

Data accessibility and presentation

The dashboard provides a user-friendly interface for clinicians to access complex morbidity data. Features such as dial displays with color coding facilitate easy identification of areas needing improvement, making the data more actionable for clinicians.

Stimulating quality improvement initiatives

By highlighting discrepancies and areas of underperformance, the pomVLAD dashboard encourages clinicians to initiate QI projects. These projects are aimed at addressing specific issues identified through the dashboard, such as postoperative outcomes like DrEaMing and other process measures that offer opportunities for clinical improvement.

Engagement from senior management

Successful delivery of change projects and sustained improvement efforts requires engagement and support from senior management. The programme theory acknowledges the need for this engagement to allocate time and resources effectively for QI initiatives.

Continuous feedback loop

The dashboard's near real-time data serves as a continuous feedback mechanism for clinicians and healthcare teams. This allows for ongoing monitoring of performance improvements and the effectiveness of implemented QI projects.

Educational and supportive role of the central team

The involvement of 'experts' from the central PQIP team in local site meetings and the direct communication with senior management highlight the importance of the study and its findings. This support helps to raise awareness and potentially increase engagement from both clinical and managerial staff.

Better data quality over administrative datasets

Clinicians recognise the clinical dataset provided by the pomVLAD dashboard as superior in quality to administrative datasets typically used by organisational analysts. This high-quality data is seen as more reliable for informing clinical practice and QI efforts.

Customisation and relevance to clinical practice

The programme theory suggests that the dashboard should present data that is relevant and customisable to different specialties and clinical practices. This ensures that the metrics and recommendations are applicable and achievable, thereby fostering more targeted and effective QI initiatives.

6.2.4 Research question 5

What are staff members' perceptions of the pomVLAD intervention?

Staff members' perceptions were broadly positive, highlighting its potential to engage healthcare professionals, particularly surgeons, in QI efforts. The intervention was viewed as particularly well-suited for surgical engagement, given its focus on surgical outcomes and the opportunity it presents for targeted improvements. Morbidity, rather than mortality, was seen as a more relevant and actionable outcome measure, offering a meaningful basis for monitoring and improving patient care.

The dashboard's ability to provide timely feedback was highly valued, as it allowed for the immediate identification of areas needing attention, contrasting with the limitations of static, historical reports. This real-time aspect was believed to hold the potential to transform how teams interact with and respond to their performance data. However, realising this potential hinged on regular engagement with the dashboard and subsequent action. Staff emphasised that the impact of the dashboard would be maximised when data are not only accessed but also acted upon through the implementation of improvement projects.

Despite the enthusiasm, some staff members expressed confusion over the range of morbidity measures reported across PQIP as a whole and the concept of risk adjustment, suggesting a need for clearer communication and education on these topics. Concerns were also raised about the performance and accuracy of risk-adjusted models, especially their ability to reflect the case mix of different hospitals when making comparisons.

6.2.5 Research question 6

Do local investigators use the pomVLAD dashboard and PQIP quarterly reports? If so, how?

Local investigators used the pomVLAD dashboard and PQIP quarterly reports, but their engagement levels and the manner in which they utilised these resources varied. Investigators found the pomVLAD and other PQIP dashboards useful for obtaining a quick overview of care delivery. The dashboard was particularly valued for its potential to highlight the performance of care delivery in a visually appealing and straightforward manner. However, access to the dashboard tended to be infrequent and often occured incidentally, such as when investigators were already on the PQIP website for other reasons, such as inputting patient data. Many acknowledged only looking at the dashboards out of interest or when they had spare time, rather than as part of a regular review process.

The production and release of quarterly PQIP reports seemed to reduce the engagement of some participants with the dashboard. A few investigators expressed a preference for the reports over the dashboard, citing the reports' comprehensive analysis and ease of access as reasons for their preference. They felt that the reports offered a well-structured and familiar format for reviewing data, which they relied on for disseminating findings within their teams. This preference suggests that while the concept of real-time dashboards is exciting and recognised as innovative, the actual impact and usage of them are limited by a lack of familiarity and established habits.

Despite this, there is a recognition of the value of real-time data provided by the dashboards, with some investigators acknowledging their potential for offering quicker feedback compared to the more comprehensive but less frequent quarterly reports. This suggests a growing interest in leveraging real-time data for more timely monitoring and improvement of care delivery, although the full potential of this approach has yet to be realised by many investigators.

6.2.6 Research question 7

How do staff members describe their experiences of using the pomVLAD dashboard?

Staff members described their experiences of using the pomVLAD dashboard as generally positive, valuing it for the quick and visual overview it provides of care delivery. They appreciated the dashboard's potential, particularly for its ability to highlight areas of care that are performing well or need improvement. This visual and immediate representation was cited as a significant advantage, making the dashboard potentially the best tool for a quick assessment of performance.

However, the experiences shared by staff also highlighted a pattern of infrequent access to the pomVLAD dashboard. Many staff members indicated that their engagement with the dashboard occured sporadically, often tied to other activities on the PQIP website, such as inputting patient data. This incidental use suggests that while the dashboard was recognised for its utility, it did not become a regular part of most staff members' routines. Some even expressed a lack of familiarity with accessing the dashboard, relying instead on the more traditional format of PQIP quarterly reports for in-depth analysis and review.

Despite this infrequent access, the concept of real-time dashboards was met with enthusiasm. Staff members acknowledged the innovative nature of live reporting and the unique benefits it offers for monitoring care delivery. However, they also noted that their engagement with and impact of the dashboard was limited by a lack

of previous experience and established habits in using such tools for continuous care monitoring.

6.2.7 Research question 8

What factors act as barriers to site collaborators engaging with their local data and pomVLAD dashboard and using them in quality improvement initiatives?

Site collaborators face several barriers to engaging with their local data and the pomVLAD dashboard, impeding their ability to utilise these tools. One primary barrier is the perceived need for a baseline period of data collection to understand local care delivery before implementing changes. This perception, coupled with the small number of patients being recruited at some sites, leads to large apparent swings in performance, making it challenging to interpret the data reliably. Consequently, some clinicians waited for data to become more mature and relevant before taking action. Ongoing issues with data collection, particularly related to turn over or loss of staff supporting data collection reduced the use of data to drive improvement.

A significant barrier to QI activity repeatedly discussed was a lack of dedicated time during day-to-day work. Local investigators felt improvement activity was undervalued by their trust and senior management, with QI activity typically delivered outside designated job plans. This lack of allocated time for QI, combined with the challenge of engaging colleagues across departments and specialties, severely restricted the capacity of sites to undertake meaningful improvement work. Additionally, difficulties in accessing and analysing data sources outside the main PQIP programme, including administrative data routinely collected within the NHS, present further obstacles.

Limited senior management support for QI emerged as another major barrier, with a perceived disconnect between clinical and managerial priorities for improvement. The absence of formal QI structures within sites exacerbated this issue, making

improvement work heavily dependent on individual clinicians. This lack of support and structure meant that even when clinicians managed to secure job planned time for PQIP and QI work, it often did not reflect the true time commitment required, forcing them to go above and beyond their job plans to deliver the work.

Where national improvement agendas did not align with local priorities the perceived value of the pomVLAD dashboard was diminished. There was also a perception that some national audit projects are effectively in competition for engagement, potentially reducing interest in PQIP and the pomVLAD dashboard. Staff turnover and poor continuity in clinical teams further hinder the day-to-day running of the PQIP study and the implementation of QI initiatives.

Lastly, a perceived good or satisfactory performance, either based on local views or in comparison to the national average, reduced the drive to improve. If performance was deemed acceptable, there was less incentive for site collaborators to engage deeply with local data and the pomVLAD dashboard for QI efforts. These barriers collectively contributed to a challenging environment for leveraging local data and the pomVLAD dashboard in quality improvement initiatives.

6.2.8 Research question 9

Are there factors that act as facilitators to increase engagement and use of the dashboard/reports for quality improvement?

Several facilitators significantly enhanced engagement and utilisation of the dashboard and reports for QI within healthcare settings. Regularly scheduled meetings dedicated to reviewing data entry and performance have proved to be a crucial motivator for local investigators, sparking enthusiasm for QI work. These sessions not only facilitated discussions on data interpretation and identification of data quality issues but also encouraged collaborative brainstorming on potential QI projects.

Support and endorsement from national bodies, such as the Royal College of Anaesthetists and the Royal College of Surgeons, also emerged as significant facilitators, providing a top-down push that boosted local engagement and multidisciplinary involvement. This external validation played an important role, especially in integrating QI activities into routine practices across different teams.

Shared learning of successful QI projects between sites was suggested as a potential way to address the issue of isolated working within the NHS, and promote a culture of continuous improvement. Investigators also felt the development of local networks and direct communication from the central project team to senior trust management helped to raise the profile of local QI efforts and secure additional support and resources.

Allocating specific time in job plans for QI activities enabled clinicians to engage meaningfully with the data and develop improvement initiatives. The involvement of trainees in data collection and QI implementation reduced the burden on senior clinicians and research nurses whilst also fostering a proactive approach to QI projects.

An institutional culture that prioritised quality improvement, characterised by dedicated QI days, competitions, and identified QI leads, helped to cultivate an environment where improvement initiatives were valued and pursued. The backing of management and divisional leads was equally crucial, providing the necessary support to ensure the success of QI activities.

Education in improvement science and previous successes in QI projects empowered clinicians to effectively engage their colleagues in QI efforts. Establishing QI teams for specific projects or for the day-to-day management of studies like PQIP helped to reduce the burden on individuals, facilitating better engagement and delivery of QI activities.

6.2.9 Research questions 10 and 11

What are PQIP collaborators' recommendations for improving the dashboard? What are PQIP collaborators' recommendations to increase the use of the dashboard to drive quality improvement?

PQIP collaborators provided several insightful recommendations aimed at enhancing the functionality of the pomVLAD dashboard and fostering its use in driving quality improvement initiatives. One key suggestion was to simplify the process of exporting data, enabling users to conduct their own analyses, especially for data not currently reported within pomVLAD or the broader PQIP study. This improvement would support local QI projects by providing easier access to raw data for bespoke analysis. Additionally, incorporating the total number of patients alongside percentages in the dashboard was recommended to add context to large changes in performance metrics, especially in settings where patient recruitment numbers are small. A suggestion to overcome this issue was to implement moving average charts, which would mitigate the impact of limited patient numbers that can lead to large swings in apparent performance when reported as monthly percentages.

The ability to visually compare performance with surrounding hospitals was also suggested as a potential improvement. Such comparisons, particularly relevant in the context of establishing local networks, could stimulate engagement and collaborative improvement efforts. In addition to this, enabling local investigators to customise the dashboard displays to reflect local improvement priorities may further enhance the dashboard's utility, making it more relevant to specific local agendas.

Improvements to facilitate the identification of individual cases within the dashboard was suggested as a way to support easier local investigation of morbidity and mortality post-surgery. This change was implemented during the pilot phase of the dashboard. Where a patient died during their hospital stay a red triangle was placed on the VLAD chart, and hovering over that provided the PQIP CaseId which a local investigator could then use to find the patients details. Reporting constituent parts

of the DrEaMing metric was also suggested as a way to improve the dashboard. Investigators suggested this would provide clearer insights into which areas require focus for improvement, as often this was only one part of the drinking, eating and mobilising bundle of care, rather than all three elements.

To aid the interpretation of the VLAD and understanding of the risk-adjustment process, collaborators recommended improving the written resources available. This included the use of detailed explanations of the risk-adjustment methodology, the origins of the dataset, and guidance on interpreting trends in the VLAD. Another area suggested was adding control limits or confidence intervals to the VLAD to help investigators determine the significance of observed trends.

6.3 Focus and questions for future work

Reflecting on some of the key findings from this work there are multiple areas and questions for future focus. Firstly, the integration and utilisation of the dashboard in QI efforts highlights a critical need for enhancing user experience, particularly in data accessibility and dashboard customisation. The ability to easily export data for local analyses and to tailor the dashboard to reflect local priorities suggests a potential avenue for improving engagement with the dashboard and local data.

The role of regular meetings in generating enthusiasm for QI work underscores the importance of collaborative engagement and continuous dialogue among local collaborators. This raises the question of how virtual or regional networks could be established or strengthened to facilitate shared learning and collaboration beyond the confines of individual hospitals.

Endorsement and active promotion by national bodies, such as the Royal Colleges, was identified as powerful facilitators of engagement. This presents an opportunity to explore more formal partnerships or endorsements that could elevate the profile and perceived importance of QI projects within the surgical and anaesthetic communities. What strategies can be employed to secure such endorsements, and how can they be leveraged to maximise impact?

The challenge of increasing engagement with the dashboard and reports amidst the pressures of clinical duties points to a broader issue of how best to support clinicians to deliver QI, including whether dedicated time for QI activities within job plans may offer benefit. Investigating models for integrating QI work into clinical roles without overburdening staff is essential. Could there be innovative approaches to job planning or financial incentives that better support QI activities? Can the burden of data collection be removed so that clinicians can focus their time on using data for improvement rather than collecting it themselves?

With the roll out of electronic health records across the NHS there is an opportunity to capture large amounts of data during routine patient care. This data capture needs to be in a structured, standardised format however to allow automated analyses and display. If developed, such a system would allow reporting of a large range of healthcare process and outcome metrics to become instantaneous without the need for any documentation beyond that required for standard patient care. This model would allow hospital sites to capture a wider range of process and outcome measures for a greater number of patients, thus reducing concerns about small patient numbers and the representativeness of data.

However, even if this is realised, there remains a significant translation step from data capture, analysis, and display to its use by clinicians and hospital teams for driving improvement in patient care. This step is complex and likely to be highly context specific. What works in one hospital department, may not work in another, even within the same institution, let alone a different hospital. Understanding how local context affects the impact of national QI initiatives, and how this context can be changed to maximise the effect of national programmes on patient outcomes will help improve the impact of such programmes. A ‘one size fits all’ approach at a national level may not provide the scope for local clinicians and leaders to address local priorities. Instead, facilitating and empowering clinicians to understand their local performance and drive their own improvement efforts may yield more success. Delivering such a ‘local feel’ on a national scale may be difficult to do and in isolation without addressing other barriers to quality improvement activity is unlikely to lead to sustainable change.

6.4 Reflexivity

Reflecting on my involvement in this study, I realise that my roles as a clinician, researcher, and a member of the PQIP central team, coupled with my belief in the power of data to drive quality improvement (QI), may have shaped the research and its findings.

As a clinician, I brought an understanding of the complexities and challenges of hospital work, which influenced both the design and the analysis of the study. This insider perspective was invaluable for connecting with participants and interpreting their experiences. However, it also meant that I was predisposed to see the value in multidisciplinary approaches and perioperative coordination, which I assumed would benefit both patients and practitioners. My own personal biases may have influenced the discussions I had around difficulties implementing change in a complex system, as I have at times found it hard to instigate improvement projects as a clinician for varying reasons. I also recognise that my strong belief in data-driven quality improvement may have impacted which aspects of the study I prioritised and how I interpreted the data.

This work was my first experience of conducting interviews and formal qualitative research. I did at times catch myself 'leading' the discussion and as I gained experience of interviewing participants I found this easier to manage and avoid. The process of interviewing participants has improved me as a researcher. I now have a much stronger appreciation of the value of understanding my biases and acknowledging their impact on my work. There is undoubtedly great skill in gathering information about individuals' thoughts and experiences without projecting one's own views. This is something that I became increasingly conscious of during the work, but I acknowledge I still have much to learn.

My role as part of the PQIP central team added another layer of complexity to my work. Striving to maintain a stance as an independent researcher while possessing insider knowledge occasionally led to moments where I found myself advocating for PQIP, perhaps influencing how participants responded to my questions.

Participants often regarded me as a source of advice for improving PQIP implementation, drawing on my insights to enhance their own practices. While this was helpful for building relationships and fostering a collaborative atmosphere, my advocacy for PQIP and the field of perioperative medicine could have subtly nudged the study's findings in certain directions. I continually attempted to check these impulses by engaging with other research team members in critical reflection and analysis. Participants were also aware that the pomVLAD dashboard was ‘my project’ and their responses about its utility and potential may therefore have been skewed. They were, however, honest about their interactions and the potential barriers to using the dashboard, and so I am confident the qualitative analysis provides an honest and balanced evaluation.

6.5 Limitations

This thesis, while providing valuable insights into the implementation of a monitoring tool into clinical practice, has some limitations that should be considered when interpreting its results. One key limitation is the relatively short implementation period, which may have hindered the full realisation of its potential impact. Longer-term studies are necessary to assess whether improvement in outcomes can be observed over a longer period and in other fields of medicine.

Additionally, the study's small sample size, with the intervention limited to ten hospitals, may have constrained the statistical power of the findings. This limitation makes it challenging to detect small but meaningful effects and to generalise the results across a broader range of surgical settings. Consequently, larger-scale studies are important to confirm the intervention's effectiveness and ensure that the findings are applicable to diverse clinical environments. Any studies assessing the implementation of monitoring tools should also include a qualitative evaluation to improve understanding of how clinicians use them and support the translation of data capture and feedback to quality improvement.

The study's focus on a broad spectrum of surgical procedures also poses a limitation, as it restricts the ability to assess the intervention's efficacy within specific surgical specialties or procedures. Future research should aim to investigate the intervention's impact in more targeted settings, providing a clearer understanding of its effectiveness across various surgical domains.

While the pomVLAD study employed rigorous methodological approaches, potential biases may still influence the interpretation of its results. Selection bias is one possible type, as the 20 hospitals (intervention and control sites), were selected for participation based on their study recruitment. The hospitals could therefore be viewed as ‘early adopters’ who were perhaps in a better position to implement the intervention compared to other sites recruiting lower numbers or not recruiting patients at all.

Measurement error is another potential issue, given the reliance on postoperative morbidity data recorded at postoperative day 7. Although the outcome measure used has been validated and demonstrated to have good inter-observer reliability.[65,66]

Lastly, the study may be subject to performance bias, as surgical teams aware of being part of the intervention group might have been more motivated to enhance their performance. This awareness could have led to a Hawthorne effect, where the observed changes in process and outcome may not be sustained over time once the initial enthusiasm and scrutiny diminish. Addressing these biases and limitations in future studies will be crucial to providing a more comprehensive evaluation of the pomVLAD intervention and its long-term effectiveness in improving surgical outcomes.

6.5.1 History bias

A possible source of bias that deserves special consideration is history bias. History biases are “events unrelated to the policy under study that occur before or during the implementation of that policy and that may have a greater effect on the policy’s hoped-for outcome than the policy itself”.[225]

Intervention and control sites that participated in the pomVLAD study were selected from the NHS hospital trusts actively recruiting patients to the wider PQIP study. Two specific aims of this larger study, which the pomVLAD intervention was nested within, were to measure and improve outcome from major surgery in the UK NHS and to implement and evaluate a complex intervention aiming to enhance the use of data for improvement by clinical teams. Whilst the pomVLAD pilot study is an important part of delivering these aims, other forms of data feedback were used throughout the PQIP study.

These alternative methods of data feedback may have reduced the impact of the pomVLAD intervention on both process measure compliance and risk-adjusted postoperative morbidity outcome. Two key aspects are important to highlight:

1. PQIP study quarterly reporting structure and existing data dashboards
2. Annual report, national improvement priorities and collaborative events

Within the PQIP study data feedback to local site investigators was provided on a three-month basis. Reports were in pdf and Microsoft Word® format, available through the PQIP webtool, with sites notified by email when the latest release is available. Report development was led by JB with support and input from members of the PQIP delivery team and preceded the pomVLAD study. Local performance against a range of perioperative process measures aggregated to month were shown, together with a national performance comparison. Specialty specific reports, such as colorectal or hepatobiliary, were provided as well as hospital level data. Appendix E-1 shows an example the PQIP quarterly report provided to sites in July 2018. The content and layout of reports evolved during the study period.

The delivery of carbohydrate loading preoperatively, DrEaMing (drinking, eating, and mobilising) outcome, and unadjusted postoperative morbidity outcome were all reported in both the quarterly reports and existing PQIP dashboard displays.

In addition to the regular reporting structure annual reports showing national trends in process measure and outcome were also produced. The first annual PQIP report, published in April 2018, highlighted five national improvement opportunities.[47] These were:

1. Identification and treatment of preoperative anaemia
2. Measurement of HbA1c and optimisation of blood glucose control in patients diagnosed with diabetes prior to surgery
3. Use of enhanced recovery pathways
4. Individualised postoperative pain management
5. Delivery of DrEaMing within 24 hours of surgery

Two collaborative events, where local site investigators attended lectures and workshops, were held in April 2018 and June 2018. These events showcased the annual report and highlighted the national improvement opportunities.

Identification of the use of enhanced recovery pathways and DrEaMing as national priorities may influence the results of the pomVLAD intervention. Although enhanced recovery pathways vary between hospitals, they are likely to encompass many, if not all, the process measures recommended in the pomVLAD intervention.

6.6 Concluding remarks

This thesis has shown that whilst there is enthusiasm for novel monitoring tools that improve understanding of care outcomes, there remains a significant translation step between data capture and their use to improve the quality-of-care delivery. The capacity of frontline NHS staff to respond to contemporaneous data and deliver quality improvement activity whilst navigating significant time and resource constraints is limited. Novel approaches to reduce the burden of data collection and methods to create a collaborative culture of QI across institutions offer the potential

to increase capacity for QI. Improving education for clinical staff in change management rather than a focus on basic QI methodology may also result in more impactful QI activity.

References

1. Institute of Medicine, Committee on Quality of Healthcare in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. The National Academies Press; 2001. doi:10.17226/10027

2. World Health Organization, Organisation for Economic Co-operation and Development, The World Bank. *Delivering Quality Health Services: A Global Imperative for Universal Health Coverage*.; 2018.

3. Busse R, Klazinga N, Panteli D, Quentin W. *Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies*. Vol 53.; 2019. doi:10.1055/s-0031-1286086

4. Key questions and quality statements - Care Quality Commission. Accessed March 21, 2023. https://www.cqc.org.uk/about-us/how-we-will-regulate/five-key-questions-and-quality-statements

5. The National Quality Board, NHS England. *Shared Commitment to Quality from the National Quality Board - Five Year Forward View*.; 2017.

6. Donabedian A. The quality of medical care. *Science (1979)*. 1978;200(May):387-390.

7. Donabedian A. The Quality of Care: How Can It Be Assessed? *JAMA*. 1988;260(12):1743-1748. doi:10.1001/jama.1988.03410120089033

8. National Institute of Health and Care Excellence. *NICE Strategy 2021 to 2026*.; 2021.

9. Allwood D. *Engaging Clinicians in Quality Improvement through National Clinical Audit*.; 2014.

10. Healthcare Quality Improvement Partnership. A-Z of National Clinical Audits – HQIP. Accessed March 21, 2023. https://www.hqip.org.uk/a-z-of-nca/#.ZBnb8y-l0lI

11. Hospital Episode Statistics (HES) - NHS Digital. Accessed March 21, 2023. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

12. Eugene N, Oliver CM, Bassett MG, et al. Development and internal validation of a novel risk adjustment model for adult patients undergoing emergency laparotomy surgery: the National Emergency Laparotomy Audit risk model. *Br J Anaesth*. 2018;121(4):739-748. doi:10.1016/j.bja.2018.06.026

13. Healthcare Quality Improvement Partnership. *National Bowel Cancer Audit Annual Report 2019 v2.0.*; 2020.

14. Spencer SA. *Hospital Episode Statistics (HES): Improving the Quality and Value of Hospital Data.*; 2011.

15. Thompson MR, Tekkis PP, Stamatakis J, et al. The National Bowel Cancer Audit: The risks and benefits of moving to open reporting of clinical outcomes. *Colorectal Disease*. 2010;12(8):783-791. doi:10.1111/j.1463-1318.2009.02175.x

16. Batalden PB, Davidoff F. What is “quality improvement” and how can it transform healthcare? *Qual Saf Health Care*. 2007;16(1):2-3. doi:10.1136/qshc.2006.022046

17. Reed JE, Card AJ. The problem with plan-do-study-act cycles. *BMJ Qual Saf*. 2016;25(3):147-152. doi:10.1136/bmjqs-2015-005076

18. Wagstaff DT, Bedford J, Moonesinghe SR. Improvement Science in Anaesthesia. *Curr Anesthesiol Rep*. 2017;7(4):432-439. doi:10.1007/s40140-017-0234-5

19. Berwick DM. Continuous Improvement as an Ideal in Health Care. *New England Journal of Medicine*. 1989;320(1):53-56. doi:10.1056/NEJM198901053200110

20. Kohn LT, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health System*.

21. Perla RJ, Provost LP, Parry GJ. Seven Propositions of the Science of Improvement: Exploring Foundations. *Quality Management in Healthcare*. 2013;22(3).

22. Langley GJ, Moen RD, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide : A Practical Approach to Enhancing Organizational Performance*.

23. Ljungqvist O, Scott M, Fearon KC. Enhanced recovery after surgery a review. *JAMA Surg*. 2017;152(3):292-298. doi:10.1001/jamasurg.2016.4952

24. Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*. 2009;360(5):491-499. doi:10.1056/nejmsa0810119

25. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg*. 2009;250(3):363-376. doi:10.1097/SLA.0b013e3181b4148f

26. Moonesinghe SR, McGuckin D, Martin P, Bedford J. The Perioperative Quality Improvement Programme (PQIP patient study): protocol for a UK multicentre, prospective cohort study to measure, report and improve quality and outcomes of major surgery. *Perioperative Medicine*. Published online 2021:under review.

27. Smith PC, Mossialos E, Papanicolas I, Leatherman S. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*. Cambridge University Press; 2010.

28. Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. *Health Equity in England: The Marmot Review 10 Years On - The Health Foundation*.; 2020.

29. Poulton TE, Moonesinghe R, Raine R, Martin P. Socioeconomic deprivation and mortality after emergency laparotomy: an observational epidemiological study. *Br J Anaesth*. 2020;124(1):73-83. doi:10.1016/j.bja.2019.08.022

30. Wan YI, McGuckin D, Fowler AJ, Prowle JR, Pearse RM, Moonesinghe SR. Socioeconomic deprivation and long-term outcomes after elective surgery: analysis of prospective data from two observational studies. *Br J Anaesth*. 2021;126(3):642-651. doi:10.1016/j.bja.2020.10.019

31. Braithwaite J, Wears RL, Hollnagel E. *Resilient Health Care, Volume 3: Reconciling Work-as-Imagined and Work-as-Done*. CRC Press; 2016.

32. Gustafsson UO, Scott MJ, Schwenk W, et al. Guidelines for perioperative care in elective colonic surgery: Enhanced Recovery After Surgery (ERAS) Society recommendations. *Clin Nutr*. 2012;31(6):783-800. doi:10.1016/j.clnu.2012.08.013

33. Ramsay G, Haynes AB, Lipsitz SR, et al. Reducing surgical mortality in Scotland by use of the WHO Surgical Safety Checklist. *British Journal of Surgery*. 2019;106(8):1005-1011. doi:10.1002/bjs.11151

34. Thomas HS, Weiser TG, Drake TM, et al. Pooled analysis of who surgical safety checklist use and mortality after emergency laparotomy. *British Journal of Surgery*. 2019;106(2):e103-e112. doi:10.1002/bjs.11051

35. Weick KE, Sutcliffe KM. *Managing the Unexpected: Sustained Performance in a Complex World*. Wiley; 2015.

36. Aveling EL, McCulloch P, Dixon-Woods M. A qualitative study comparing experiences of the surgical safety checklist in hospitals in high-income and low-income countries. *BMJ Open*. 2013;3(8):e003039. doi:10.1136/bmjopen-2013-003039

37. Beach MC, Inui T, Frankel R, et al. Relationship-centered care: A constructive reframing. *J Gen Intern Med*. 2006;21(SUPPL. 1). doi:10.1111/j.1525-1497.2006.00302.x

38. Steele Gray C, Wodchis WP, Upshur R, et al. Supporting Goal-Oriented Primary Health Care for Seniors with Complex Care Needs Using Mobile Technology: Evaluation and Implementation of the Health System Performance Research Network, Bridgepoint Electronic Patient Reported Outcome Tool. *JMIR Res Protoc*. 2016;5(2):e126. doi:10.2196/resprot.5756

39. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Educ Couns*. 2009;74(3):295-301. doi:10.1016/j.pec.2008.11.015

40. Carmona C, Crutwell J, Burnham M, Polak L. Shared decision-making: Summary of NICE guidance. *The BMJ*. 2021;373:1-3. doi:10.1136/bmj.n1430

41. Epstein RM, Street RL. The values and value of patient-centered care. *Ann Fam Med*. 2011;9(2):100-103. doi:10.1370/afm.1239

42. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. *Br Med J*. 2007;335(7609):24-27. doi:doi: 10.1136/bmj.39246.581169.80

43. Khuri S, Daley J, Henderson W. The Department of Veterans Affairs' NSQIP Peer-Controlled Program for the Measurement and Enhancement of the Quality of Surgical Care. *Ann Surg*. 1998;228(4):491-504.

44. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a New American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg*. 2008;207(5):777-782. doi:10.1016/j.jamcollsurg.2008.07.018

45. Shiloach M, Frencher SKJ, Steeger JE, et al. Toward Robust Information: Data Quality and Inter-Rater Reliability in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg*. 2010;210(1).

46. Khuri SF, Henderson WG, DePalma RG, Mosca C, Healey NA, Kumbhani DJ. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg*. 2005;242(3):326-343. doi:10.1097/01.sla.0000179621.33268.83

47. PQIP project team. *Perioperative Quality Improvement Programme: Annual Report 2017-18*.; 2018. doi:10.1039/C1DT90165F

48. PQIP project team. *Perioperative Quality Improvement Programme: Annual Report 2018-19*.; 2019.

49. PQIP project team. *Periopative Quality Improvement Program: Report 3 - August 2019 to July 2021*.; 2021.

50. PQIP project team. *Perioperative Quality Improvement Programme: National Report 4 - July 2021 to March 2023*.; 2023.

51. Oliver CM, Warnakulasuriya S, McGuckin D, et al. Delivery of drinking, eating and mobilising (DrEaMing) and its association with length of hospital stay after major noncardiac surgery: observational cohort study. *Br J Anaesth*. 2022;129(1):114-126. doi:10.1016/j.bja.2022.03.021

52. NHS England. Commissioning for Quality and Innovation (CQUIN) scheme for 2022/23 Annex: Indicator specifications. 2023;(January).

53. PQIP project team. *Perioperative Quality Improvement Programme - Report 5 March 2023 to March 2024*.; 2024. https://pqip.org.uk/FilesUploaded/PQIP-Report2023-2024.pdf

54. NHS England. Commissioning for Quality and Innovation (CQUIN) scheme for 2022/23 Annex: Indicator specifications. 2023;(January). https://www.england.nhs.uk/wp-content/uploads/2022/12/CQUIN-scheme-for-2023-24-indicator-specifications-version-1.1.pdf

55. Eugene N, Kuryba A, Walker K. *The Second Patient Report of the National Emergency Laparotomy Audit (NELA): Development of the Risk Adjustment Model*.; 2016.

56. NELA Project Team. *Eighth Patient Report of the National Emergency Laparotomy Audit*.; 2023.

57. Bedford JR, Martin P, Spurling LJ, Moonesinghe SR. Identifying and investigating apparent variation in care based on continuous outcome monitoring. 2020. https://data.nela.org.uk/getattachment/Support/NELA-Risk-Adjustment-Model-(1)/NELA-(2020)-Investigating-unusual-results-in-continuous-outcome-monitoring.pdf.aspx?lang=en-GB

58. Oliver CM, Bassett MG, Poulton TE, et al. Organisational factors and mortality after an emergency laparotomy: multilevel analysis of 39 903 National Emergency Laparotomy Audit patients. *Br J Anaesth*. 2018;121(6):1346-1356. doi:10.1016/j.bja.2018.07.040

59. NELA Project Team. *Sixth Patient Report of the National Emergency Laparotomy Audit: December 2018 to November 2019*.; 2020.

60. Pearse RM, Clavien PA, Demartines N, et al. Global patient outcomes after elective surgery: Prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth*. 2016;117(5):601-609. doi:10.1093/bja/aew316

61. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: A 7 day cohort study. *The Lancet*. 2012;380(9847):1059-1065. doi:10.1016/S0140-6736(12)61148-9

62. Moonesinghe SR, Harris S, Mythen MG, et al. Survival after postoperative morbidity: a longitudinal observational cohort study. *Br J Anaesth*. 2014;113(6):977-984. doi:10.1093/bja/aeu224

63. Manku K, Bacchetti P, Leung JM. Prognostic significance of postoperative in-hospital complications in elderly patients. I. Long-term survival. *Anesth Analg*. 2003;96(2):583-589.

64. Stephens TJ, Peden CJ, Pearse RM, et al. Improving care at scale: Process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). *Implementation Science*. 2018;13(1):1-16. doi:10.1186/s13012-018-0840-8

65. Grocott MPW, Browne JP, Van der Meulen J, et al. The Postoperative Morbidity Survey was validated and used to describe morbidity after major surgery. *J Clin Epidemiol*. 2007;60(9):919-928. doi:10.1016/j.jclinepi.2006.12.003

66. Davies SJ, Francis J, Dilley J, Wilson RJT, Howell SJ, Allgar V. Measuring outcomes after major abdominal surgery during hospitalization : reliability and validity of the Postoperative Morbidity Survey. *Perioperative Medicine*. 2013;2(1):1-9.

67. Wong DJN, Oliver CM, Moonesinghe SR. Predicting postoperative morbidity in adult elective surgical patients using the Surgical Outcome Risk Tool (SORT). *Br J Anaesth*. 2017;119(1):95-105. doi:https://dx.doi.org/10.1093/bja/aex117

68. Ahmad T, Bouwman RA, Grigoras I, et al. Use of failure-to-rescue to identify international variation in postoperative care in low-, middle- and high-income

countries: A 7-day cohort study of elective surgery. *Br J Anaesth*. 2017;119(2):258-266. doi:10.1093/bja/aex185

69. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. *New England Journal of Medicine*. 2009;361(14):1368-1375. doi:10.1056/NEJMsa0903048

70. Dimick JB, Pronovost PJ, Cowan JA, Lipsett PA, Stanley JC, Upchurch GR. Variation in postoperative complication rates after high-risk surgery in the United States. *Surgery*. 2003;134(4):534-540; discussion 540-1. doi:10.1016/S0039

71. Montroy J, Breau RH, Cnossen S, et al. Change in adverse events after enrollment in the national surgical quality improvement program: A systematic review and meta-analysis. Shamji M, ed. *PLoS One*. 2016;11(1):1-14. doi:10.1371/journal.pone.0146254

72. Osborne NH, Nicholas LH, Ryan AM, Thumma JR, Dimick JB. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for medicare beneficiaries. *JAMA - Journal of the American Medical Association*. 2015;313(5):496-504. doi:10.1001/jama.2015.25

73. Peden CJ, Stephens T, Martin G, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *The Lancet*. 2019;393(10187):2213-2221. doi:10.1016/S0140-6736(18)32521-2

74. Huddart S, Peden CJ, Swart M, McCormick B, Dickinson M, Mohammed M. Use of a pathway quality improvement care bundle to reduce mortality after emergency laparotomy. *Br J Surg*. 2015;102(1):57-66. doi:http://dx.doi.org/10.1002/bjs.9658

75. Woodcock T, Liberati EG, Dixon-Woods M. A mixed-methods study of challenges experienced by clinical teams in measuring improvement. *BMJ Qual Saf*. 2021;30(2):106-115. doi:10.1136/bmjqs-2018-009048

76. Bennett-Guerrero E, Welsby I, Dunn TJ, et al. The use of a postoperative morbidity survey to evaluate patients with prolonged hospitalization after routine, moderate-risk, elective surgery. *Anesth Analg*. 1999;89(2):514-519. doi:10.1213/00000539-199908000-00050

77. Dindo D, Demartines N, Clavien PA. Classification of Surgical Complications. *Ann Surg*. 2004;240(2):205-213. doi:10.1097/01.sla.0000133083.54934.ae

78. Howes TE, Cook TM, Corrigan LJ, Dalton SJ, Richards SK, Peden CJ. Postoperative morbidity survey, mortality and length of stay following emergency laparotomy. *Anaesthesia*. 2015;70(9):1020-1027. doi:10.1111/anae.12991

79. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217(5):833-842.e3. doi:10.1016/j.jamcollsurg.2013.07.385

80. Stubbs DJ, Davies BM, Bashford T, et al. Identification of factors associated with morbidity and postoperative length of stay in surgically managed chronic subdural haematoma using electronic health records: A retrospective cohort study. *BMJ Open*. 2020;10(6):1-11. doi:10.1136/bmjopen-2020-037385

81. Stubbs DJ, Davies BM, Burnstein R, Joannides AJ, Ercole A. Exploratory Analysis of Preoperative and Postoperative Risk Stratification Tools to Identify Acute Kidney and Myocardial Injury in Patients Undergoing Surgery for Chronic Subdural Haematoma. *J Neurosurg Anesthesiol*. 2022;34(4):439-440. doi:10.1097/ANA.0000000000000796

82. Stubbs DJ, Bowen JL, Furness RC, et al. Development and Validation of an Electronic Postoperative Morbidity Score. *Anesth Analg*. 2019;129(4):935-942. doi:10.1213/ANE.0000000000003953

83. Balvardi S, St-Louis E, Yousef Y, et al. Systematic review of grading systems for adverse surgical outcomes. *Canadian Journal of Surgery*. 2021;64(2):E196-E204. doi:10.1503/cjs.016919

84. Mitropoulos D, Artibani W, Biyani CS, Bjerggaard Jensen J, Rouprêt M, Truss M. Validation of the Clavien–Dindo Grading System in Urology by the European Association of Urology Guidelines Ad Hoc Panel. *Eur Urol Focus*. 2018;4(4):608-613. doi:10.1016/j.euf.2017.02.014

85. Abbott TEF, Patel A, Ahmed T, et al. Prospective observational cohort study on grading the severity of postoperative complications in global surgery research. *British Journal of Surgery*. 2019;106(2):e73-e80. doi:10.1002/bjs.11025

86. Mentula PJ, Leppäniemi A. Applicability of the Clavien-Dindo classification to emergency surgical procedures: a retrospective cohort study on 444 consecutive patients. *Patient Saf Surg*. 2014;8(31):1-7. doi:10.1177/1457496913511019

87. Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes*. 4th ed. Health Administration Press; 2012.

88. Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. *Br Med J*. 1999;318(7197):1515-1520. doi:10.1136/bmj.318.7197.1515

89. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005;24(8):1185-1202. doi:10.1002/sim.1970

90. Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet*. 2001;357:463-467. doi:10.1016/S0140-6736(00)04019-8

91. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet*. 1997;350:1128-1130. doi:http://dx.doi.org/10.1016/S0140-6736%2897%2906507-0

92. Cook DA, Coory M, Webster RA. Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators. *BMJ Qual Saf*. 2011;20(6):469-474. doi:10.1136/bmjqs.2008.031831

93. Steiner SH, Mackay RJ. Monitoring risk-adjusted medical outcomes allowing for changes over time. *Biostatistics*. 2014;15(4):665-676. doi:10.1093/biostatistics/kxt057

94. Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *Br Med J*. 1998;316:1697-1700. doi:10.1136/bmj.316.7146.1697

95. Pagel C, Utley M, Crowe S, et al. Real time monitoring of risk-adjusted paediatric cardiac surgery outcomes using variable life-adjusted display: implementation in three UK centres. *Heart*. 2013;99:1445-1450. doi:https://dx.doi.org/10.1136/heartjnl-2013-303671

96. Stark J, Gallivan S, Lovegrove J, et al. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet*. 2000;355(9208):1004-1007. doi:10.1016/S0140-6736(00)90001-1

97. Sherlaw-Johnson C, Lovegrove J, Treasure T, Gallivan S. Likely variations in perioperative mortality associated with cardiac surgery: when does high

mortality reflect bad practice? *Heart*. 2000;84(1):79-82. doi:10.1136/heart.84.1.79

98. Duckett SJ, Coory M, Sketcher-Baker K. Identifying variations in quality of care in Queensland hospitals. *Medical Journal of Australia*. 2007;187(10):571-575.

99. Steiner SH, Jones M. Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart. *Stat Med*. 2010;29(4):444-454. doi:10.1002/sim.3788

100. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med*. 2007;35(4):1091-1098. doi:10.1097/01.CCM.0000259468.24532.44

101. Martin P, Bedford JR, Spurling LJ, Moonesinghe SR. What is a EWMA chart and how do I read it? 2020. https://data.nela.org.uk/getattachment/Support/NELA-Risk-Adjustment-Model-(1)/NELA-(Sep-2020)-What-is-a-EWMA-chart-(1).pdf.aspx?lang=en-GB

102. Steiner SH, Woodall WH. Debate: what is the best method to monitor surgical performance? *BMC Surg*. 2016;16(1):15. doi:10.1186/s12893-016-0131-8

103. Montgomery DC. *Introduction To Statistical Quality Control.* Vol 10. 8th ed. Wiley; 2019.

104. Mohammed MA, Rathbone A, Myers P, Patel D, Onions H, Stevens A. An investigation into general practitioners associated with high patient mortality flagged up through the Shipman inquiry: retrospective analysis of routine data. *Br Med J*. 2004;328:1474-1477. doi:10.1136/bmj.328.7454.1474

105. Wagstaff D, Moonesinghe SR, Fulop NJ, Vindrola-Padros C. Qualitative process evaluation of the Perioperative Quality Improvement Programme (PQIP): Study protocol. *BMJ Open*. 2019;9(7). doi:10.1136/bmjopen-2019-030214

106. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medial contexts. *Stat Methods Med Res*. 2003;12(2):147-170. doi:10.1177/096228020301200205

107. Spiegelhalter D, Sherlaw-Johnson C, Bardsley M, Blunt I, Wood C, Grigg O. Statistical methods for healthcare regulation: rating, screening and surveillance. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY*. 2012;175(1):1-25. doi:10.1111/j.1467-985X.2011.01010.x

108. Covidence. Accessed September 21, 2018. https://www.covidence.org/home

109. Google Translate. Accessed February 21, 2019. https://translate.google.com/

110. QSR International. NVivo 12 Qualitative data analysis software. Published online 2018.

111. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*. 2013;13(1):1-8. doi:10.1186/1471-2288-13-117

112. Long AF, Godfrey M, Randall T, Brettle A, Grant M. *HCPRDU Evaluation Tool for Mixed Methods Studies*.; 2002.

113. Pluye P, Nha Hong Q. Combining the Power of Stories and the Power of Numbers: Mixed Methods Research and Mixed Studies Reviews. *Annual*

Reviews of Public Health. 2014;35:29-45. doi:10.1146/annurev-publhealth-032013-182440

114. Long A, Godfrey M, Randall T, Brettle A, Grant M. *Developing Evidence Based Social Care Policy and Practice. Part 3: Feasibility of Undertaking Systematic Reviews in Social Care.*; 2002. doi:10.5999/aps.2016.43.1.128

115. Pace R, Pluye P, Bartlett G, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud*. 2012;49(1):47-53. doi:10.1016/j.ijnurstu.2011.07.002

116. Souto RQ, Khanassov V, Hong QN, Bush PL, Vedel I, Pluye P. Systematic mixed studies reviews: Updating results on the reliability and efficiency of the mixed methods appraisal tool. *Int J Nurs Stud*. 2015;52(1):500-501. doi:10.1016/j.ijnurstu.2014.08.010

117. Pluye P, Gagnon MP, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. *Int J Nurs Stud*. 2009;46(4):529-546. doi:10.1016/j.ijnurstu.2009.01.009

118. Mixed methods appraisal tool. Accessed October 17, 2019. http://mixedmethodsappraisaltoolpublic.pbworks.com

119. Albert AA, Walter JA, Arnrich B, et al. On-line variable live-adjusted displays with internal and external risk-adjusted mortalities. A valuable method for benchmarking and early detection of unfavourable trends in cardiac surgery. *European Journal of Cardio-thoracic Surgery*. 2004;25(3):312-319. doi:10.1016/j.ejcts.2003.12.009

120. Albert AA, Walter JA, Arnrich B, Rosendahl UP, Ennker J. Handling of information and quality assessment in cardiac surgery - Evaluation of clinic

specific techniques. *Z Herz Thorax Gefasschir*. 2004;18(1):7-12. doi:http://dx.doi.org/10.1007/s00398-004-0429-1

121. Arrowsmith JE, Powell SJ, Nashef SAM. Local clinical quality monitoring for detection of excess operative deaths. *Anaesthesia*. 2006;61(5):423-426. doi:10.1111/j.1365-2044.2006.04587.x

122. Belliveau D, Burton HJ, O'Blenes SB, Warren AE, Hancock Friesen CL. Real-time complication monitoring in pediatric cardiac surgery. *Annals of Thoracic Surgery*. 2012;94(5):1596-1602. doi:10.1016/j.athoracsur.2012.05.103

123. Borracci RA, Rubio M, Guerrero RAA, Barrero C, Mauro V, Fairman E. Continuous monitoring of Cardiac Surgery Quality Indicators. *Arch Cardiol Mex*. 2007;77(4):275-283.

124. Brunelli A, Refai M, Salati M, Pompili C, Sabbatini A. Standardized combined outcome index as an instrument for monitoring performance after pulmonary resection. *Ann Thorac Surg*. 2011;92:272-277. doi:https://dx.doi.org/10.1016/j.athoracsur.2011.03.038

125. Clarke ALL, Shearer W, McMillan AJ, Ireland PD. Investigating apparent variation in quality of care: the critical role of clinician engagement. *Medical Journal of Australia*. 2010;193(8):S111-S113.

126. Collett D, Sibanda N, Pioli S, Bradley JA, Rudge C. The UK scheme for mandatory continuous monitoring of early transplant outcome in all kidney transplant centers. *Transplantation*. 2009;88(8):970-975. doi:10.1097/TP.0b013e3181b997de

127. Driessen SRC, Van Zwet EW, Haazebroek P, et al. A dynamic quality assessment tool for laparoscopic hysterectomy to measure surgical outcomes. *Am J Obstet Gynecol*. 2016;215(6):754.e1-754.e8. doi:10.1016/j.ajog.2016.07.004

128. Fusco D, Barone AP, Sorge C, et al. P.Re.Val.E.: outcome research program for the evaluation of health care quality in Lazio, Italy. *BMC Health Serv Res*. 2012;12:25. doi:http://dx.doi.org/10.1186/1472-6963-12-25

129. Kuhan G, McCollum DP, Renwick PM, Chetter IC, McCollum PT. Variable life adjusted display methodology for continuous performance monitoring of carotid endarterectomy. *Ann R Coll Surg Engl*. 2018;100(1):63-66. doi:http://dx.doi.org/10.1308/rcsann.2017.0170

130. Lovegrove J, Sherlaw-Johnson C, Valencia O, Treasure T, Gallivan S. Monitoring the performance of cardiac surgeons. *Journal of the Operational Research Society*. 1999;50(7):684-689.

131. Morton AP, Clements ACA, Doidge SR, Stackelroth J, Curtis M, Whitby M. Surveillance of healthcare-acquired infections in Queensland, Australia: Data and lessons from the first 5 years. *Infect Control Hosp Epidemiol*. 2008;29(8):695-701. doi:10.1086/589904

132. Patella M, Sandri A, Pompili C, et al. Real-time monitoring of a video-assisted thoracoscopic surgery lobectomy programme using a specific cardiopulmonary complications risk-adjusted control chart. *European Journal of Cardio-thoracic Surgery*. 2016;49(4):1070-1074. doi:10.1093/ejcts/ezv294

133. Roberts G, Thorburn G, Collins D, Smailes S, Dziewulski P. Development and implementation of prospective outcome monitoring in a regional burns service using cumulative sum (CUSUM) techniques. *Burns*. 2013;39(4):571-576. doi:http://dx.doi.org/10.1016/j.burns.2012.10.018

134. Sketcher-Baker KM, Kamp MC, Connors JA, Martin DJ, Collins JE. Using the quality improvement cycle on clinical indicators - improve or remove? *Medical Journal of Australia*. 2010;193(8 Suppl):S104-106.

135. Snyder JJ, Salkowski N, Zaun D, et al. New quality monitoring tools provided by the scientific registry of transplant recipients: CUSUM. *American Journal of Transplantation*. 2014;14(3):515-523. doi:10.1111/ajt.12628

136. Vasilakis C, Wilson APR, Haddad FS. Automating the monitoring of surgical site infections using variable life-adjusted display charts. *Journal of Hospital Infection*. 2011;79(2):119-124. doi:http://dx.doi.org/10.1016/j.jhin.2011.05.015

137. Williams H, Gwyn R, Smith A, Dramis A, Lewis J. Variable life-adjusted display (VLAD) for hip fracture patients: a prospective trial. *European Journal of Orthopaedic Surgery & Traumatology*. 2015;25(6):1047-1050. doi:https://dx.doi.org/10.1007/s00590-015-1620-7

138. Brunelli A, Rocco G, Van Raemdonck D, Varela G, Dahan M. Lessons learned from the European thoracic surgery database: The composite performance score. *European Journal of Surgical Oncology*. 2010;36(SUPPL. 1):S93-S99. doi:10.1016/j.ejso.2010.06.023

139. Brunelli A, Berrisford RG, Rocco G, Varela G. The European Thoracic Database project: composite performance score to measure quality of care after major lung resection. *European Journal of Cardio-thoracic Surgery*. 2009;35(5):769-774. doi:10.1016/j.ejcts.2009.01.037

140. O'Brien SM, Shahian DM, DeLong ER, et al. Quality Measurement in Adult Cardiac Surgery: Part 2-Statistical Considerations in Composite Measure Scoring and Provider Rating. *Annals of Thoracic Surgery*. 2007;83(4 SUPPL.). doi:10.1016/j.athoracsur.2007.01.055

141. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55-63. doi:10.7326/M14-0697

142. Pagel C, Brown KL, Crowe S, Utley M, Cunningham D, Tsang VT. A mortality risk model to adjust for case mix in UK paediatric cardiac surgery. *Health Services and Delivery Research*. 2013;1(1):1-133. doi:10.3310/hsdr01010

143. Sergeant P, de Worm E, Meyns B. Single centre, single domain validation of the EuroSCORE on a consecutive sample of primary and repeat CABG. *European Journal of Cardio-thoracic Surgery*. 2001;20(6):1176-1182. doi:10.1016/S1010-7940(01)01013-2

144. Davies G. Queensland Public Hospitals Commission of Inquiry: Report Bundaberg Base Hospital. Published online 2005:1-537.

145. Van Der Weyden MB. The Bundaberg hospial scandal: The need for reform in Queensland and beyond. *Medical Journal of Australia*. 2006;184(6):284-285. doi:10.5694/j.1326-5377.2006.tb00196.x

146. Daly MP, Cleary MI, McCormack LJ. The Bundaberg emergency response team. *Australian Health Review*. 2012;36(4):384. doi:10.1071/ah11061

147. Roberts G, Thorburn G, Smailes S, Dziewulski P. Open access software tool for CUSUM analysis in burns – a freely available download for prospective outcome monitoring. *Scars Burn Heal*. 2016;2(1). doi:10.1177/2059513116642396

148. Shah A. Using data for improvement. *Br Med J*. 2019;364:l189. doi:10.1136/bmj.l189

149. Groven S, Eken T, Skaga NO, Roise O, Naess PA, Gaarder C. Long-lasting performance improvement after formalization of a dedicated trauma service. *Journal of Trauma - Injury, Infection and Critical Care*. 2011;70(3):569-574. doi:http://dx.doi.org/10.1097/TA.0b013e31820d1a9b

150. Daly MP, Cleary MI, McCormack LJ. The Bundaberg emergency response team. *Australian Health Review*. 2012;36(4):384. doi:10.1071/ah11061

151. Health Service Directive: Queensland Government. *Guideline for Variable Life Adjusted Display and Other National Patient Safety Indicators*.; 2017.

152. Centre for Healthcare Improvement. *Report on the VLAD Laparoscopic Cholecystectomy Indicator Review: Summary of Activities*.; 2012.

153. Centre for Healthcare Improvement. *Report on the VLAD Heart Failure Indicator Review Final Summary of Activities*.; 2012.

154. Centre for Healthcare Improvement. *Paediatric Tonsillectomy and Adenoidectomy*.; 2011.

155. Patient Safety Unit: Queensland Health. *Report on the Stroke VLAD Indicator Review*.; 2012.

156. Patient Safety Unit: Queensland Health. *Report on the Orthopaedic VLAD Indicator Review: Summary of Activity*.; 2012.

157. Patient Safety Unit: Queensland Health. *Report on the Obstetric VLAD Indicator Review: Summary of Activity*.; 2014.

158. Patient Safety Unit: Queensland Health. *Report on the Acute Myocardial Infarction VLAD Indicator Review: Summary of Activity*.; 2012.

159. Renzi C, Asta F, Fusco D, Agabiti N, Davoli M, Perucci CA. Does public reporting improve the quality of hospital care for acute myocardial infarction? Results from a regional outcome evaluation program in Italy. *International Journal for Quality in Health Care*. 2014;26(3):223-230. doi:10.1093/intqhc/mzu041

160. Renzi C, Sorge C, Fusco D, Agabiti N, Davoli M, Perucci CA. Reporting of quality indicators and improvement in hospital performance: The P.Re.Val.E. regional outcome evaluation program. *Health Serv Res*. 2012;47(5):1880-1901. doi:10.1111/j.1475-6773.2012.01401.x

161. Berta P, Seghieri C, Vittadini G. Comparing health outcomes among hospitals: The experience of the Lombardy Region. *Health Care Manag Sci*. 2013;16:245-257. doi:10.1007/s10729-013-9227-1

162. Ventura M, Fusco D, Bontempi K, Colais P, Davoli M. Regional Outcome Evaluation Program (P.Re. Val.E.): Reduction of inequality in access to effective health care in the Lazio region of Italy (2012-2015). *PLoS One*. 2018;13(3):1-14. doi:10.1371/journal.pone.0194972

163. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*.; 2001. doi:10.1007/978-1-4757-3462-1

164. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997;16(9):965-980. doi:10.1002/(SICI)1097-0258(19970515)16:9<965::AID-SIM509>3.0.CO;2-O

165. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78(1):1-3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

166. Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. *British Journal of Surgery*. 1991;78(3):355-360. doi:10.1002/bjs.1800780327

167. Gilhooly D, Moonesinghe SR. The Perioperative Quality Improvement Programme: improving outcomes. *Br J Hosp Med*. 2018;79(2):117. doi:10.1093/bja/aex137

168. Sahni A, Moonesinghe R. The Perioperative Quality Improvement Programme. *The Bulletin of the Royal College of Surgeons of England*. 2018;100(8):335-337. doi:10.1308/rcsbull.2018.335

169. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. *British Journal of Surgery*. 1998;85(9):1217-1220. doi:10.1046/j.1365-2168.1998.00840.x

170. Midwinter MJ, Tytherleigh M, Ashley S. Estimation of mortality and morbidity risk in vascular surgery using POSSUM and the Portsmouth predictor equation. *British Journal of Surgery*. 1999;86(4):471-474. doi:10.1046/j.1365-2168.1999.01112.x

171. Tekkis PP, Prytherch DR, Kocher HM, et al. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). *BRITISH JOURNAL OF SURGERY*. 2004;91(9):1174-1182. doi:10.1002/bjs.4430

172. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MP. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. *Anesthesiology*. 2013;119(4):959-981. doi:10.1097/ALN.0b013e3182a4e94d

173. Bagnall NM, Pring ET, Malietzis G, et al. Perioperative risk prediction in the era of enhanced recovery: a comparison of POSSUM, ACPGBI, and E-PASS scoring systems in major surgical procedures of the colorectal surgeon. *Int J Colorectal Dis*. 2018;33(11):1627-1634. doi:10.1007/s00384-018-3141-4

174. Protopapa KL, Simpson JC, Smith NCE, Moonesinghe SR. Development and validation of the Surgical Outcome Risk Tool (SORT). *British Journal of Surgery*. 2014;101(13):1774-1783. doi:10.1002/bjs.9638

175. Tu J V., Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. *J Am Coll Cardiol*. 2001;37(4):992-997. doi:10.1016/S0735-1097(01)01109-3

176. Walker K, Finan PJ, Van Der Meulen JH. Model for risk adjustment of postoperative mortality in patients with colorectal cancer. *British Journal of Surgery*. 2015;102(3):269-280. doi:10.1002/bjs.9696

177. Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *British Journal of Surgery*. 2002;89(6):763-768. doi:10.1046/j.1365-2168.2002.02080.x

178. Canet J, Gallart L, Gomar C, et al. Prediction of Postoperative Pulmonary Complications in a Population-based Surgical Cohort. *Anesthesiology*. 2010;113(6):1338-1350. doi:10.1097/ACO.0b013e32835e80d8

179. Sauerbrei W. The Use of Resampling Methods to Simplify Regression Models in Medical Statistics. *Journal of the Royal Statistical Society*. 1999;48(3):313-329. doi:10.1111/1467-9876.00155

180. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*. 1986;1(1):54-75.

181. Austin PC, Tu J V. Bootstrap Methods for Developing Predictive Models. *Am Stat*. 2004;58(2):131-137. doi:10.1198/0003130043277

182. Hurvich CM, Tsai C ling. Regression and Time Series Model Selection in Small Samples. *Biometrika*. 1989;76(2):297-307.

183. van Houwelingen JC. Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy. *Stat Neerl*. 2001;55(1):17-34. doi:10.1111/1467-9574.00154

184. Moons KGM, Donders ART, Steyerberg EW, Harrell FE. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: A clinical example. *J Clin Epidemiol*. 2004;57(12):1262-1270. doi:10.1016/j.jclinepi.2004.01.020

185. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: A simulation study based on clinical data. *Stat Med*. 2002;21(24):3803-3822. doi:10.1002/sim.1422

186. Harrell FE, Margolis P a, Gove S, et al. Development of a clinical prediction model for an ordinal outcome: the World Health Organization Multicentre Study of Clinical Signs and Etiological agents of Pneumonia, Sepsis and Meningitis in Young Infants. WHO/ARI Young Infant Multicentre Study Group. *Stat Med*. 1998;17(8):909-944. doi:10.1002/(SICI)1097-0258(19980430)17:8<909::AID-SIM753>3.0.CO;2-O

187. Steyerberg EWÃ, Eijkemans MJC, Habbema JDF. Application of shrinkage techniques in logistic regression analysis : a case study. 2001;55(1):76-88.

188. Hurvich CM, Tsai C -L. A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection. *J Time Ser Anal*. 1993;14(3):271-279. doi:10.1111/j.1467-9892.1993.tb00144.x

189. Jones DR, Copeland GP, De Cossart L. Comparison of POSSUM with APACHE II for prediction of outcome from a surgical high-dependency unit. *British Journal of Surgery*. 1992;79(12):1293-1296. doi:10.1002/bjs.1800791216

190. Brooks MJ, Sutton R, Sarin S. Comparison of Surgical Risk Score, POSSUM and p-POSSUM in higher-risk surgical patients. *British Journal of Surgery*. 2005;92(10):1288-1292. doi:10.1002/bjs.5058

191. González-Martínez S, Martín-Baranera M, Martí-Saurí I, Borrell-Grau N, Pueyo-Zurdo JM. Comparison of the risk prediction systems POSSUM and P-POSSUM with the Surgical Risk Scale: A prospective cohort study of 721 prospective patients. *International Journal of Surgery*. 2016;29:19-24. doi:10.1016/j.ijsu.2016.03.005

192. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: Patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg*. 2013;217(2):336-346. doi:10.1016/j.jamcollsurg.2013.02.027

193. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*. 1996;58(1):267-288.

194. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x

195. Pavlou M, Ambler G, Seaman S, De iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016;35(7):1159-1177. doi:10.1002/sim.6782

196. Langley A, Denis JL. Beyond evidence: The micropolitics of improvement. *BMJ Qual Saf*. 2011;20(SUPPL. 1):43-46. doi:10.1136/bmjqs.2010.046482

197. Peden CJ, Stephens T, Martin G, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *The Lancet*. 2019;393(10187):2213-2221. doi:10.1016/S0140-6736(18)32521-2

198. Stephens TJ, Peden CJ, Pearse RM, et al. Improving care at scale: Process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). *Implementation Science*. 2018;13(1):1-16. doi:10.1186/s13012-018-0840-8

199. Wholey J. *Evaulation and Effective Public Management*. Little, Brown; 1983.

200. McNall MA, Welch VE, Ruh KL, Mildner CA, Soto T. The use of rapid-feedback evaluation methods to improve the retention rates of an HIV/AIDS

healthcare intervention. *Eval Program Plann*. 2004;27(3):287-294. doi:10.1016/j.evalprogplan.2004.04.003

201. Sonnichsen RC. High Impact Internal Evaluation: A Practitioner's Guide to Evaluating and Consultant Inside Organizations. In: ; 2000. doi:10.4135/9781483328485

202. Soumerai SB, Starr D, Majumdar SR. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. *Prev Chronic Dis*. 2015;12(6):1-17. doi:10.5888/pcd12.150187

203. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-355. doi:10.1093/ije/dyw098

204. Bernal JL, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *Int J Epidemiol*. 2018;47(6):2082-2093. doi:10.1093/ije/dyy135

205. Soumerai SB, Ceccarelli R, Koppel R. False Dichotomies and Health Policy Research Designs: Randomized Trials Are Not Always the Answer. *J Gen Intern Med*. 2017;32(2):204-209. doi:10.1007/s11606-016-3841-9

206. Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: A methodological systematic review. *BMJ Glob Health*. 2020;5(10). doi:10.1136/bmjgh-2020-003567

207. McNall M, Foster-Fishman PG. Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation*. 2007;28(2):151-168. doi:10.1177/1098214007300895

208. R Core Team (2021). R: A language and environment for statistical computing.

209. Bedford J. pomVLAD: Near real-time reporting of risk-adjusted outcomes. 2018. https://pomvlad.blog

210. Cripps P, Cripps M. netsolving. Accessed May 24, 2022. https://www.netsolving.com

211. Lassen K. Consensus Review of Optimal Perioperative Care in Colorectal Surgery. *Archives of Surgery*. 2009;144(10):961. doi:10.1001/archsurg.2009.170

212. Ljungqvist O, Scott M, Fearon KC. Enhanced Recovery After Surgery. *JAMA Surg*. 2017;152(3):292. doi:10.1001/jamasurg.2016.4952

213. Gilhooly D, Green S, McCann C, Black N, Moonesinghe S. Barriers and facilitators to the successful development, implementation and evaluation of care bundles in the acute care setting: a systematic review. *Implementation Science*. 2019;14(47):1-12.

214. Substantial amendment to PQIP study protocol and documents. 2019. Accessed May 26, 2022. https://pqip.org.uk/pages/study_documents

215. QSR International. NVivo 20 qualitative data analysis software.

216. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*. 2013;13(117). doi:10.1186/1471-2288-13-117

217. Mascha EJ, Sessler DI. Segmented Regression and Difference-in-Difference Methods: Assessing the Impact of Systemic Changes in Health Care. *Anesth Analg*. 2019;129(2):618-633. doi:10.1213/ANE.0000000000004153

218. Caswell J. Interrupted Time Series Analysis for Single Series and Comparative Designs: Using Administrative Data for Healthcare Impact

Assessment. Accessed May 19, 2022. https://www.sas.com/content/dam/SAS/en_ca/User Group Presentations/Health-User-Groups/ITS_SAS.pdf

219. Fawcett WJ, Ljungqvist O. Starvation, carbohydrate loading, and outcome after major surgery. *BJA Educ*. 2017;17(9):312-316. doi:10.1093/bjaed/mkx015

220. NHS England. Bed Availability and Occupancy Data: 2017-18 Quarter 1. Accessed May 11, 2022. https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/

221. Moonesinghe SR, Wong DJN, Farmer L, Shawyer R, Myles PS, Harris SK. SNAP-2 EPICCS: The second Sprint National Anaesthesia Project - EPIdemiology of Critical Care after Surgery: Protocol for an international observational cohort study. *BMJ Open*. 2017;7(9):1-8. doi:10.1136/bmjopen-2017-017690

222. Wong DJN, Popham S, Wilson AM, et al. Postoperative critical care and high-acuity care provision in the United Kingdom, Australia, and New Zealand. *Br J Anaesth*. 2019;122(4):460-469. doi:10.1016/j.bja.2018.12.026

223. PQIP project team. *Perioperative Quality Improvement Programme: Annual Report 2018-19.*; 2019. https://pqip.org.uk/FilesUploaded/PQIP Annual Report 2018-19.pdf

224. Bauer M, Böhrer H, Aichele G, Bach A, Martin E. Measuring patient satisfaction with anaesthesia: Perioperative questionnaire versus standardised face-to-face interview. *Acta Anaesthesiol Scand*. 2001;45(1):65-72. doi:10.1034/j.1399-6576.2001.450111.x

225. Naci H, Soumerai SB. History Bias, Study Design, and the Unfulfilled Promise of Pay-for-Performance Policies in Health Care. *Prev Chronic Dis*. 2016;13:E82. doi:10.5888/pcd13.160133

Appendices

Appendix A

Appendix A-1 : Procedures eligible for recruitment to PQIP study

Patients undergoing the following procedures were eligible for recruitment to PQIP at: (a) the time of model derivation and internal validation (black, non italic); (b) procedures added after model derivation and prior to temporal validation *(italic);* (c) procedures discontinued after model derivation and before temporal validation (red)

All major Head and Neck resections including:

- Craniofacial resection (+/- reconstruction)
- Extensive excision of mandible (+/- reconstruction)
- Glossectomy (total)
- Laryngectomy (partial)
- Laryngectomy (subtotal)
- Laryngectomy (total)
- Laryngotracheal reconstruction following PLO
- Maxillectomy (partial/hemi for malignancy +/- reconstruction)
- Mediastinal thyroidectomy/parathyroidectomy with sternotomy
- VATS excision of mediastinal tumour (including thymectomy)
- Partial excision of trachea and reconstruction
- Pharyngectomy (partial)
- Pharyngectomy (total)
- Pharyngolaryngo-oesophagectomy (PLO)
- Radical dissection of cervical lymph nodes +/-:
  - local myocutaneous flap reconstruction
  - distant myocutaneous flap reconstruction
- Selective dissection of cervical lymph nodes +/-:
  - local myocutaneous flap reconstruction
  - distant myocutaneous flap reconstruction
- *Prosthetic replacement of temporomandibular joint*
- *Reconstruction of jaw (non-vascularised reconstruction)*
- *Tongue flap - first stage and second stage*
- *Selective dissection of cervical lymph nodes, levels 1 to 5 (+/- 6)*

Thoracics:

- VATS bullectomy (unilateral/bilateral)
- VATS excision of mediastinal tumour (including thymectomy)
- VATS lobectomy
- VATS lung volume reduction (unilateral/bilateral)
- VATS metastasectomy
- VATS pleurodesis/pleurectomy
- VATS pneumonectomy
- VATS wedge resection of lung
- Carinal resection +/- pneumonectomy
- Decortication of pleura of the lung
- Excision of chest wall tumour (+/- reconstruction)
- Lung resection with excision of chest wall
- Open excision of lesion of lung

- Open pneumonectomy
- Open resection of mediastinal tumour
- Partial excision of trachea and reconstruction
- Pulmonary lobectomy including segmental resection
- Thoracotomy and closure of bronchopleural fistula
- Thoracotomy and bullectomy (unilateral/bilateral)
- Thoracotomy and lung volume reduction
- Thoracotomy and pleurectomy/pleurodesis (+/- ligation of bullae)
- *Exploratory thoracotomy*
- *Plication of paralysed diaphragm*
- *Repair of rupture of diaphragm*
- *Secondary correction of scolios-related chest wall deformity (posterior costoplasty) (as sole procedure)*
- *Partial excision of trachea with reconstruction*
- *Tracheoplasty*
- *Correction of pectus deformity of chest wall*
- *Open pleural biopsy as sole procedure*
- *Percutaneous radiofrequency ablation of malignant neoplasm of lung*
- *Sleeve resection of bronchus or pulmonary artery with pulmonary resection*
- *Mediastinal parathyroidectomy with sternotomy*
- *VATS debridement of empyema*
- *VATS excision lesion of oesophagus*
- *VATS oesophageal / oesophagogastric myotomy*
- *VATS sympathectomy – bilateral*
- *VATS sympathectomy - unilateral*

Abdominal - Hepatobiliary:

- Exploratory laparotomy
- Hemihepatectomy (left/right)
- Partial hepatectomy (+/- ablation)
- Pancreatectomy (partial/distal)
- Pancreatectomy (total)
- Pancreaticoduodenectomy (Whipple procedure)
- Radiofrequency thermocoagulation of the liver with scapel resection
- Resection of lesion(s) of liver
- Splenectomy (partial/total)
- *Exploratory laparotomy*
- *Creation of portocaval shunt*
- *Frey's procedure*
- *Associating Liver Partition and Portal vein Ligation for Staged hepatectomy (ALPPS)*
- *Partial excision of bile duct and anastamosis of bile duct to duodenum/jejunum*
- *Anastomosis of gall bladder (to another viscus)*
- *Anastomosis of hepatic duct*
- *Anastomosis of pancreatic duct (to another viscus)*
- *Excision of lesion of pancreas*
- *Hepatectomy including partial / hemi*
- *Repair of bile duct*

Abdominal - Upper GI:

- Exploratory laparotomy
- Partial gastrectomy (+/- excision of surrounding tissue)
- Total gastrectomy (+/- excision of surrounding tissue)
- Bypass of oesophagus
- Complex restoration of intestinal continuity
- Ileo-caecal resection (with anastamosis or ileostomy formation)
- Intestinal failure reconstruction
- Laparotomy for enterocutaneous fistula
- Oesophagectomy (partial)

- Oesophagectomy (total)/Oesophagogastrectomy
- Pancreatectomy (partial/distal)
- Pancreatectomy (total)
- Pancreaticoduodenectomy (Whipple procedure)
- Partial excision of bile duct and anastamosis of bile duct to duodenum/jejunum
- Resection of duodenal tumour
- Resection of small bowel (+/- tumour)
- Splenectomy (partial/total)
- *Exploratory laparotomy*
- *Complex restoration of intestinal continuity*
- *Partial excision of bile duct and anastamosis of bile duct to duodenum/jejunum*
- *Ileo-caecal resection (with anastamosis or ileostomy formation)*
- *Closure of bypass of oesophagus*
- *Open excision of lesion of oesophagus*
- *Revision of anti-reflux procedures*
- *Transabdominal anti-reflux operations*
- *Transthoracic repair of diaphragmatic hernia (acquired)*
- *Transthoracic repair of paraoesophageal hiatus hernia*
- *Anastomosis of gall bladder (to another viscus)*
- *Anastomosis of hepatic duct*
- *Anastomosis of pancreatic duct (to another viscus)*
- *Excision of lesion of bile duct*
- *Excision of lesion of pancreas*
- *Hepatectomy including partial / hemi*
- *Radiofrequency thermocoagulation of liver with scalpel liver resection*
- *Repair of bile duct*
- *Resection of liver tumour(s)*
- *Hemihepatectomy (right)*
- *Hemihepatectomy (left)*
- *Laparoscopic biliary gastric bypass*
- *Revision of gastro-jejunostomy*
- *Total or Partial gastrectomy and excision of surrounding tissue*
- *Vagotomy and pyloroplasty*
- *Endoscopically assisted oesophagectomy*
- *Repair of ruptured oesophagus*
- *Transabdominal repair of diagphragmatic hernia*
- *Transthoracic fundoplication*

Abdominal - Lower GI:

- Abdominal revision of restorative proctocolectomy
- Abdominoperineal (AP) resection with anastamosis (+/- pouch)
- Abdominoperineal (AP) resection with end colostomy
- Anterior resection
- Colectomy and colostomy with preservation of rectum
- Colectomy (total) and ileorectal anastamosis
- Completion proctectomy
- Complex restoration of intestinal continuity
- Excision of retroperitoneal tumour (+/- ureterolysis)
- Exploratory laparotomy
- Intestinal failure reconstruction
- Ileo-caecal resection (with anastamosis or ileostomy formation)
- Hartmann's procedure
- Restoration of intestinal continuity
- Laparotomy for enterocutaneous fistula
- Left hemicolectomy (with anastomosis/colostomy)
- Panproctocolectomy and ileostomy
- Resection of duodenal tumour
- Resection of small bowel (+/- tumour)

- Reversal of Hartmann's procedure
- Right hemicolectomy (with anastomosis/ileostomy)
- Sigmoid colectomy
- *Ileo-caecal resection (with anastamosis or ileostomy formation)*
- *Open excision of lesion of duodenum*
- *Resection of small intestine*
- *Ileoanal anastomosis and creation of pouch*
- *Redo operations on ileum/colon*
- *Abdominal operation for Hirschprung's disease*
- *Excision of transverse colon*
- *Partial excision of rectum and sigmoid colon for prolapse*
- *Total mesorectal excision (TME) including Trans-Anal / TATME*
- *Transanal endoscopic mircosurgery*

Abdominal - other:

- Abdominal wall reconstruction
- Adrenalectomy (unilateral/bilateral)
- Excision of retroperitoneal tumour (+/- ureterolysis)
- Exploratory laparotomy
- Laparotomy + excision of sarcoma tumour
- Retroperitoneal lymph node dissection
- Pelvic exenteration
- *Complex restoration of intestinal continuity*
- *Intestinal failure reconstruction*
- *Restoration of intestinal continuity*

Urology

- Pelvic exenteration
- Radical prostatectomy
- Retroperitoneal lymph node dissection
- Total cystectomy (with construction of intestinal conduit or neobladder)
- Total nephrectomy (non-transplant)
- *Total exenteration of pelvis*
- *Appendiciovesicostomy / Mitrofanoff procedure*
- *Nephrectomy and excision of perirenal tissue*
- *Nephroureterectomy*
- *Percutaneous nephrolithotomy (including cystoscopy and retrograde catheterisation)*
- *Bilateral replantation of ureter into bladder*
- *Construction of ileal conduit*
- *Excision of ureterocele (with or without ureteric reimplantation) - bilateral*
- *Ileal or colonic replacement of ureter*
- *Open correction vesicoureteric reflux-bilateral*
- *Replantation of ureter into bowel (including bilateral)*
- *Bilateral Ureterolysis*
- *Enlargement of bladder*
- *Enterocystoplasty*
- *Laparoscopic pyeloplasty*
- *Laparoscopic upper or lower pole heminephrectomy*
- *Repair of bladder exstrophy*
- *Repair of vesicocolic fistula*
- *Complex secondary repair of hypospadias*
- *Repair of epispadias*
- *Prostatic cryotherapy*

Appendix B

Appendix B-1 Data extraction fields and categories

Data field	Data extracted (free text/ categorical/ numerical)	Categories used	Notes
Title	Free text	NA	
Authors	Free text	Cardiology/General Medicine	
Year	Numerical	Paediatric ENT surgery	
Language	Free text		
Type of publication	Categorical	Peer-reviewed, organisational report, other	
Source of funding	Free text		
Country of study	Free text		
Setting of study	Categorical	Primary care, secondary care, tertiary care, state-wide/regional, national, international	
Number of institutions involved	Numerical		Converted to single centre or multicentre when reported in review
Study design			
Aims and objectives	Free text		
Risk of selection bias	Categorical	Yes, No	Free text explanation added as to what risk involved
Inclusion criteria	Free text		
Exclusion criteria	Free text		
Time period reported	Date / free text		The time period the studies reports on
Specialty	Categorical	Cardiac surgery, thoracic surgery, Upper gastrointestinal, lower gastrointestinal, urology, gynaecology, intensive care, cardiology, paediatrics, other	Free text if other
Population group	Categorical	Adult, paediatric, both, not stated	
Description provided of study population demographics	Categorical	Yes, no	
Detail provided about monitoring intervention	Categorical	None, limited, satisfactory	
Detail provided about methods used to embed VLADs	Categorical	None, limited, satisfactory	

Data field	Data extracted (free text/ categorical/ numerical)	Categories used	Notes
Original (new) qualitative data presented in study	Categorical	Yes, no	Answered yes if there was any new qualitative information regarding the implementation included. This did not need to be a formal qualitative evaluation.
Number of patients included	Numerical		
Outcome reported	Categorical	Mortality, morbidity, other	If other, free text allowed
If outcome other than morbidity, how was it defined?	Free text		
Main study finding/outcomes	Free text		
Secondary study findings/outcomes	Free text		
Name given to chart in study	Categorical	VLAD, RA-CUSUM, CUSUM, O-E, E-O, other	
Method of risk adjustment	Categorical	Derived for purposes of monitoring, published risk model, other	If existing risk model used, name/reference of published model extracted
Rationale for implementing monitoring	Categorical	Quality assurance, quality improvement, both	
Context of implementation	Categorical	Local, regional, national, international, other	
Level of outcome reporting	Categorical	Individual clinician/surgeon, departmental/hospital, regional,	
Method of data collection	Categorical	Clinical database, administrative database, national registry, other	Free text clarification of method also extracted
Method of VLAD dissemination	Free text		
Use of online dashboard	Categorical	Yes, no	If yes, additional free text information extracted about: software used; primary users; any pilot period for software; how dashboard was accessed
Methods to embed VLADs into routine clinical practice	Free text		

Data field	Data extracted (free text/ categorical/ numerical)	Categories used	Notes
Alternative methods used to report findings alongside VLADs	Free text		
Use of flags to signal variation	Categorical	Yes, no	If yes, additional data extracted about how flags were calculated
Action taken on results of VLAD monitoring	Free text		
Costs/resource use	Free text		
Facilitators of VLAD monitoring	Free text		Data extracted as full paragraphs to enable thematic analysis
Barriers to VLAD monitoring	Free text		Data extracted as full paragraphs to enable thematic analysis

Appendix B-2 : Mixed Methods Appraisal Tool (MMAT)[113] assessment of included studies

The quality assessment of papers was carried out using MMAT.[113] All included studies were deemed to be either qualitative, quantitative descriptive studies or mixed methods studies. The questions used from the MMAT tool for each type of study are shown in Box 6-1. Appendix Table B-1 shows the results of the quality assessment.

All included studies were classified as either qualitative, quantitative descriptive or mixed methods studies. Box 6-1 shows an abridged version of the MMAT tool including only the questions relevant to these study types.

Box 6-1: Abridged MMAT quality assessment criteria

Screening questions:

S1. Are there clear qualitative and quantitative research questions (or objectives*), or a clear mixed methods question (or objective*)?

S2. Do the collected data allow address the research question (objective)? E.g., consider whether the follow-up period is long enough for the outcome to occur (for longitudinal studies or study components).

1. Qualitative studies

1.1. Are the sources of qualitative data (archives, documents, informants, observations) relevant to address the research question (objective)?

1.2. Is the process for analysing qualitative data relevant to address the research question (objective)?

1.3. Is appropriate consideration given to how findings relate to the context, e.g., the setting, in which the data were collected?

1.4. Is appropriate consideration given to how findings relate to researchers' influence, e.g., through their interactions with participants?

2. Quantitative descriptive studies

2.1. Is the sampling strategy relevant to address the quantitative research question (quantitative aspect of the mixed methods question)?

2.2. Is the sample representative of the population understudy?

2.3. Are measurements appropriate (clear origin, or validity known, or standard instrument)?

2.4. Is there an acceptable response rate (60% or above)? The response rate is not pertinent for case series and case report. E.g., there is no expectation that a case series would include all patients in a similar situation.

3. Mixed methods studies

3.1. Is the mixed methods research design relevant to address the qualitative and quantitative research questions (or objectives), or the qualitative and quantitative aspects of the mixed methods question (or objective)?

3.2. Is the integration of qualitative and quantitative data (or results*) relevant to address the research question (objective)?

3.3. Is appropriate consideration given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results*) in a triangulation design?

*These two items are not considered as double-barrelled items since in mixed methods research, (1) there may be research questions (quantitative research) or research objectives (qualitative research), and (2) data may be integrated, and/or qualitative findings and quantitative results can be integrated.

Appendix Table B-1: Results of the quality assessment performed using the Mixed Methods Appraisal Tool (MMAT)

Author and year	Classification of study	SCREENING QUESTIONS		1. QUALITATIVE STUDIES				2. QUANTITATIVE DESCRIPTIVE STUDIES				3. MIXED METHODS STUDIES			COMMENTS	Overall score
		S1	S2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3		
Albert 2004[119]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				No patient demographics presented. No rationale for hospitals included or excluded from study.	**
Albert 2004[120]	Qualitative descriptive - narrative	Yes	Yes	Can't tell	Can't tell	Yes	No									**
Arrowsmith 2006[121]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Can't tell	N/A					*
Belliveau 2012[122]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A					**
Borracci 2007[123]	Quantitative descriptive	Yes	Yes					Yes	Yes	Can't tell	N/A					**
Brunelli 2011[124]	Quantitative descriptive	Yes	Yes					Yes	Yes	Yes	N/A					***
Clarke 2010[125]	Qualitative descriptive - narrative	Yes	Can't tell	Yes	Can't tell	Can't tell	No									*
Collett 2009[126]	Quantitative descriptive	Yes	Yes					Yes	Yes	Yes	N/A					***

Author and year	Classification of study	SCREENING QUESTIONS		1. QUALITATIVE STUDIES				2. QUANTITATIVE DESCRIPTIVE STUDIES				3. MIXED METHODS STUDIES			COMMENTS	Overall score
		S1	S2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3		
Driessen 2016[127]	Mixed methods	Yes	Yes	Yes	Can't tell	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Limited description of survey process, and qualitative data analysis	*
Duckett 2007[98]	Quantitative descriptive	Yes	Yes					Yes	Yes	Yes	N/A					***
Fusco 2012[128]	Quantitative descriptive	Yes	Yes					Yes	Yes	Yes	N/A					***
Kuhan 2018[129]	Quantitative descriptive	Yes	Yes					Yes	Yes	Can't tell	N/A					**
Lovegrove 1999[130]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	No	N/A				No demographic or patient characteristics included in paper. Single centre cohort.	*
Lovegrove 1997[91]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	No	N/A				Unclear why the 3 groups for VLAD were chosen.	*
Morton 2008[131]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				No context provided for hospitals providing data versus those in the state who do not take part.	**

Author and year	Classification of study	SCREENING QUESTIONS		1. QUALITATIVE STUDIES				2. QUANTITATIVE DESCRIPTIVE STUDIES				3. MIXED METHODS STUDIES			COMMENTS	Overall score
		S1	S2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3		
Pagel 2013[95]	Mixed methods	Yes	Yes	Yes	Can't tell	Yes	No	Yes	Yes	Yes	N/A	Yes	Yes	No	Limited description of process of collecting qualitative data despite qualitative information being presented.	**
Patella 2016[132]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Can't tell	N/A					*
Roberts 2013[133]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				No presentation of patient demographics, comorbidities etc. Single centre study.	**
Sketcher-Baker 2010[134]	Qualitative descriptive - narrative	No	No					Yes	Can't tell	Yes	N/A				Narrative manuscript on experience within Queensland Health. Clear objectives for paper not set out.	**
Snyder 2014[135]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				No demographic or patient characteristics included in paper.	**

Author and year	Classification of study	SCREENING QUESTIONS		1. QUALITATIVE STUDIES				2. QUANTITATIVE DESCRIPTIVE STUDIES				3. MIXED METHODS STUDIES			COMMENTS	Overall score
		S1	S2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3		
Vasilakis 2011[136]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				Uses UK and international definitions of infection. No patient demographics for cohort included, single centre study.	**
Williams 2015[137]	Quantitative descriptive	Yes	Yes					Yes	Can't tell	Yes	N/A				Single centre, no patient demographics presented. Data independently check by two authors.	**

Appendix C

Appendix C-1 Postoperative Morbidity Survey (POMS) organ domains and criteria.

Criteria highlighted in bold are included in our composite POMSmajor outcome (Clavien-Dindo grade II and above). Excluded criteria are highlighted in italic (Clavien-Dindo grade I).

POMS organ system	POMS domain criteria	Assigned Clavien-Dindo grade
Pulmonary	**New requirement for oxygen**	**2**
Pulmonary	**New requirement for respiratory support**	**2**
Infectious	**Currently on antibiotics**	**2**
Infectious	*Temperature >38°C in the last 24hr*	*1*
Renal	*Urinary catheter in situ*	*1*
Renal	**Increased serum creatinine (>30% from preoperative level)**	**2**
Renal	**Presence of oliguria <500 mL/24hr**	**2**
Gastrointestinal	*Unable to tolerate an enteral diet for any reason*	*1*
Gastrointestinal	*Vomiting or abdominal distension, or use of antiemetics*	*1*
Cardiovascular	**Thrombotic event requiring anticoagulation (new)**	**2**
Cardiovascular	**Atrial or ventricular arrhythmias (new)**	**2**
Cardiovascular	**Hypotension (requiring pharmacological or fluid therapy >200 mL/hr)**	**2**
Cardiovascular	**New myocardial infarction or ischaemia**	**2**
Cardiovascular	**Cardiogenic pulmonary oedema**	**2**
Neurological	**New coma**	**3**
Neurological	**New confusion or delirium**	**2**
Neurological	**New focal neurological deficit**	**2**
Haematological	**Platelet, fresh-frozen plasma, or cryoprecipitate transfusion in last 24hrs**	**2**
Haematological	**Packed erythrocyte transfusion in the last 24hrs**	**2**
Wound	**Wound dehiscence requiring surgical exploration or drainage of pus from the operation wound with or without isolation of organisms**	**2**
Pain	*New pain significant enough to require parenteral opioids*	*1*
Pain	**New pain significant enough to require regional analgesia**	**2**

Appendix C-2 Frequency each candidate variable was selected into each backwards step-wise model across the 1000 bootstrap samples. Variables selected into >80% of bootstrap models were selected into the final model.

Variable	Frequency selected into bootstrap models (%)
Surgical specialty	100
Severity of surgery	100
Sex	99.5
American Society of Anaesthesiologists - Physiological Status (ASA-PS)	97.7
Body mass index	96.1
Heart rate (preoperative)	95.1
Systolic blood pressure	94.7
Age (years)	92.1
Number of operations in past 30 days	91.3
Respiratory history findings (POSSUM variable)	88.3
Smoking history	75.9
White cell count	75.3
Haemoglobin	70.0
Diagnosis of cancer in last 5 years	62.1
Oxygen saturations	58.6
New York Health Association (NYHA) classification	55.3
ECG findings (POSSUM variable)	46.3
Serum Urea	45.9
Cardiac history findings (POSSUM variable)	45.0
Urgengy of surgery (NCEPOD classification)	38.1
History of cerebrovascular disease	37.1
Current alcohol consumption	33.9
Serum creatinine	29.3
Respiratory infection in past month	27.1
History of diabetes	22.0
History of liver disease	20.3
Serum potassium	19.3
Serum sodium	18.9
Interaction term: Age (years) x Systolic blood pressure	74.0

Appendix C-3 Predicted and observed risk of postoperative morbidity predicted by PQIP model in the temporal validation cohort.

	n	Mean predicted risk	Mean observed risk	n with morbidity (predicted)	n with morbidity (observed)
[0.0219,0.0817)	826	0.060	0.053	49.90	44
[0.0817,0.1088)	825	0.096	0.085	79.35	70
[0.1088,0.1319)	825	0.120	0.124	99.31	102
[0.1319,0.1523)	825	0.142	0.126	117.19	104
[0.1523,0.1734)	825	0.163	0.147	134.30	121
[0.1734,0.1972)	825	0.185	0.135	152.68	111
[0.1972,0.2258)	825	0.211	0.170	174.14	140
[0.2258,0.2629)	825	0.243	0.234	200.73	193
[0.2629,0.3251)	825	0.291	0.299	240.48	247
[0.3251,0.6471]	825	0.391	0.379	322.72	313

The model is seen to overestimate risk in the 17.3-22.6% risk bands. Data from this table is used to produce the calibration plot shown in Figure 3-10

Appendix C-4 Recalibration of the model with estimation of shrinkage factors

As discussed in the limitations section (see 3.5.3), shrinkage factors for each surgical specialty can be used to improve calibration of a model and improve its accuracy. In order to estimate these shrinkage factors an intercept and slope value for each specialty's calibration plot is calculated. The estimates of the intercept and slope values for this type of recalibration are shown in Appendix Table C-1.

Recalibration of the model through the application of linear shrinkage factors improved calibration within specialties (see Appendix Figure C-1 in comparison to Figure 3-11). The intercept and slope estimates shown in Appendix Table C-1 reflect the relative mis-calibration of the PQIP model within each specialty when applied to the validation cohort. A perfect calibration would be represented by an intercept value of zero and a slope gradient of one.

An example of how shrinkage factors are used to produce the recalibrated risk estimates:

If the PQIP morbidity formula yields a probability (p) of having POMSmajor-defined morbidity on postoperative day 7 of 25%, and the surgical specialty is Lower Gastrointestinal then:

$$logit(p) = -1.10$$

and

$$logit[P(Y)] = -0.27 + 0.864 * (-1.10),$$

where P(Y) is the probability of having POMSmajor morbidity estimated by the calibrated model. P(Y) in this example would therefore be 22.8%.

In a second example, if the PQIP morbidity formula yields the same probability (p) of having POMSmajor-defined morbidity on postoperative day 7 of 25%, but the surgical specialty is Upper Gastrointestinal then:

$$logit(p) = -1.10$$

and

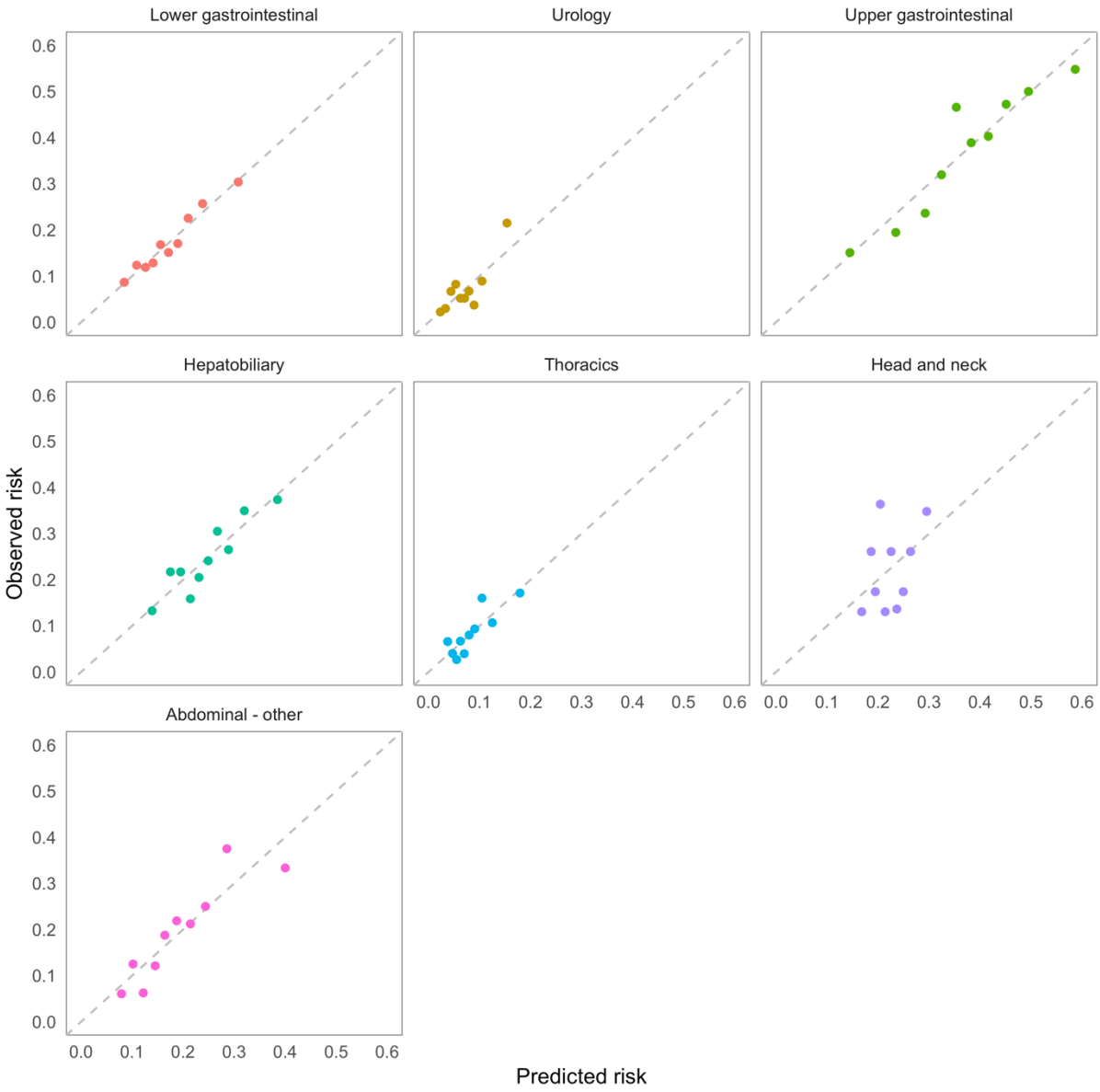
$$logit[P(Z)] = 0.345 + 1.22 * (-1.10),$$

where P(Z) is the probability of having POMSmajor morbidity estimated by the calibrated model. P(Z) in this example would therefore be 27%.

This is one method that can be used to improve calibration at the surgical specialty level. Other options include the development of an individual risk-prediction model for each specialty, or the development of a multilevel model that allows for the intercept and/or the slope of predicted log odds to vary for each surgical specialty.

Appendix Table C-1: Estimated shrinkage factors for the PQIP-morbidity adjustment model. The factors are used to optimise the performance of the model for each surgical specialty. PQIP, Perioperative Quality Improvement Programme

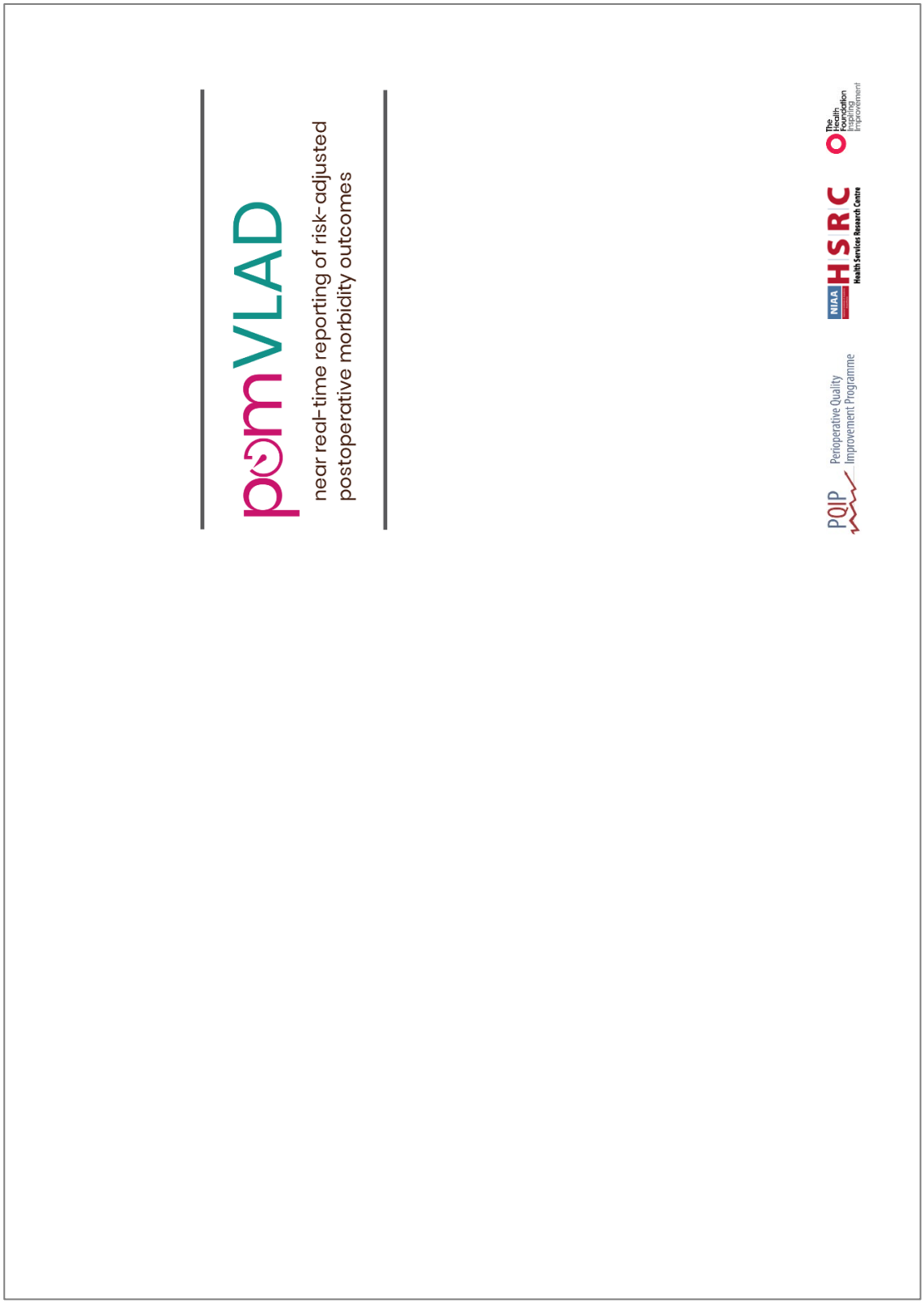
Surgical specialty	Intercept	Slope
Lower gastrointestinal	-0.27	0.864
Urology	-0.983	0.945
Upper gastrointestinal	0.345	1.22
Hepatobiliary	-0.127	0.824
Thoracics	-0.799	0.731
Head and neck	-0.672	0.373
Abdominal - other	-0.011	1.036



Appendix Figure C-1: Calibration plot for the PQIP model after recalibration using the shrinkage factors in Appendix Table C-1

Appendix D

Appendix D-1 Introductory booklet provided to intervention sites



The pomVLAD project aims to support local quality improvement through rapid feedback of risk-adjusted morbidity outcomes and targeted process measures via an online dashboard.

pomVLAD is nested within the Perioperative Quality Improvement Programme (PQIP) and will be launched in 10 hospitals in late April 2018. There is no additional burden of data collection as the information used is routinely collected by PQIP. Outcome data will be available via the PQIP webtool immediately when patient records are locked.

The dashboard will be available for all specialties recruiting to PQIP, but the targeted recommendations are aimed at abdominal surgical specialties.

pomVLAD is part of the Health Foundation's Innovating for Improvement programme. The Health Foundation is an independent charity committed to bringing better health and health care for people in the UK.

Defining postoperative morbidity and POMS major

Within the pomVLAD project morbidity is defined using the Postoperative Morbidity Survey (POMS)[1] on postoperative day 7. POMS is a valid and reliable measure of short term postoperative morbidity[2,3]. We will use 'POMS major' as our outcome measure which includes POMS criteria equivalent to a Clavien-Dindo grade II complication and above[4,5].

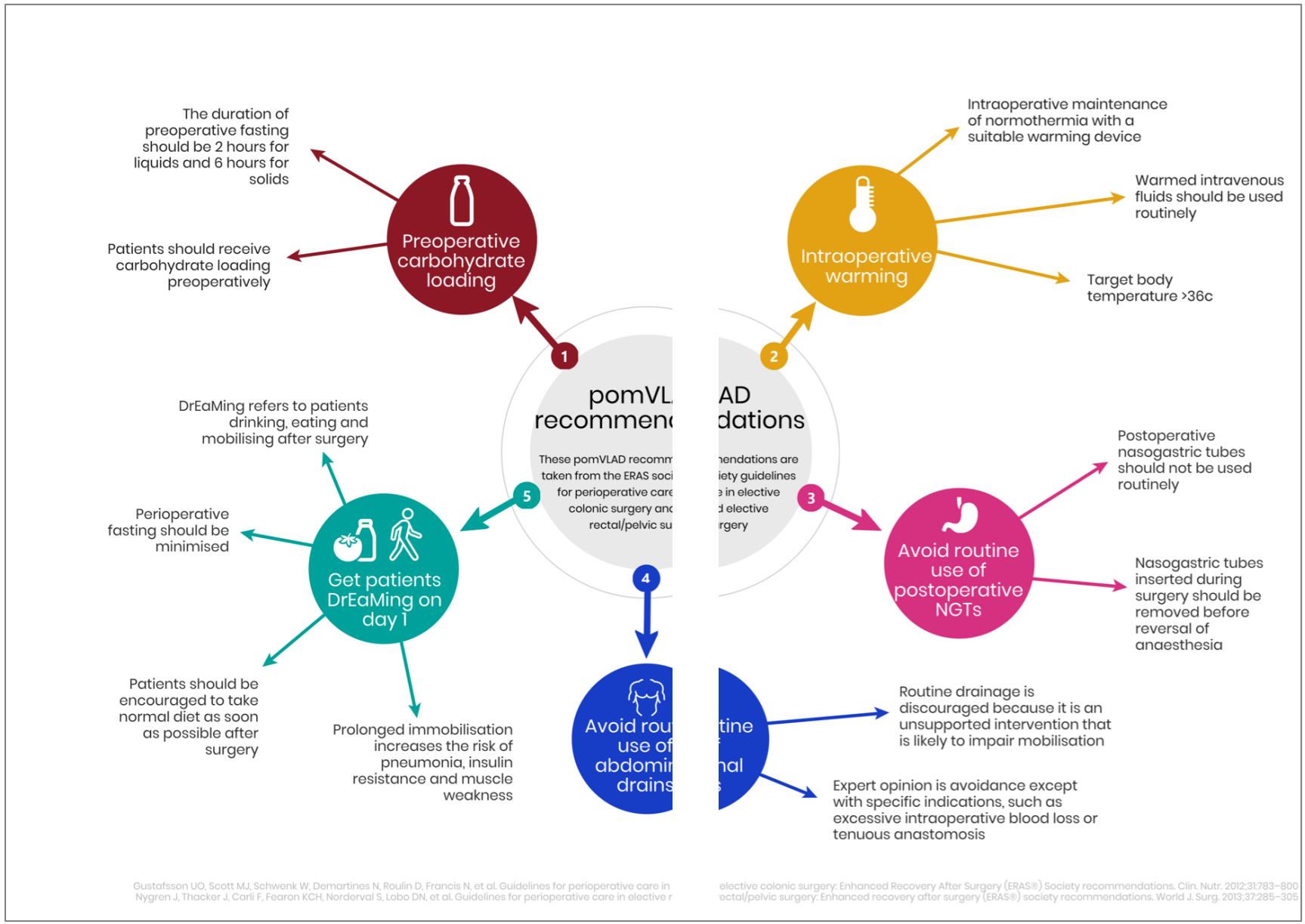
The following criteria are classified as POMS major:

Pulmonary
- New requirement for oxygen
- New requirement for respiratory support

Infectious
- Currently on IV antibiotics

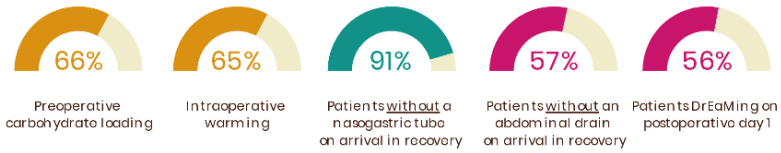
Renal
- Increased serum creatinine (>30% from preoperative level)
- Presence of oliguria <500ml/24 hours

(continued...)



Gustafsson UO, Scott MJ, Schwenk W, Demartines N, Roulin D, Francis N, et al. Guidelines for perioperative care in
Nygren J, Thacker J, Carli F, Fearon KCH, Norderval S, Lobo DN, et al. Guidelines for perioperative care in elective r
elective colonic surgery: Enhanced Recovery After Surgery (ERAS®) Society recommendations. Clin. Nutr. 2012;31:783–800
ectal/pelvic surgery: Enhanced recovery after surgery (ERAS®) society recommendations. World J. Surg. 2013;37:285–305

17% of patients received all five recommendations*



The figures above show national compliance with each of the pomVLAD recommendations for patients undergoing lower gastrointestinal surgery recruited to PQIP*.

The five pomVLAD recommendations have been developed based on guidelines from the Enhanced Recovery After Surgery (ERAS) Society and data from the first 12 months of PQIP. To develop a succinct list of recommendations we looked at the association between enrolment on an enhanced recovery pathway and each of the ERAS society recommendations for routinely collected information within PQIP.

Enrolment of patients onto an enhanced recovery pathway was associated with improved compliance with each of the five recommendations. There was insufficient data available to assess the direct association between each recommendation and postoperative morbidity.

*Data from the Perioperative Quality Improvement Programme as of 14th March 2018, including patients undergoing lower gastrointestinal surgery up to 28th February 2018

Measuring and reporting the pomVLAD recommendations

Preoperative carbohydrate loading -
This data is collected in Q2.36 of the PQIP webtool (Were preoperative carbohydrates given on the day of surgery?). The percentage of patients recorded as 'yes' will be shown on the dashboard.

Intraoperative warming -
This data is collected in Q3.6 of the webtool. If a patient receives forced air warming **and** IV fluid warming they are classed as compliant with this recommendation.

Avoid routine use of postoperative nasogastric tubes -
Collected in Q4.3 'Was a nasogastric tube present on arrival from theatres?'. Compliance with this process measure is if the patient does **not** have a nasogastric tube present on arrival in recovery.

Avoid routine use of abdominal drains -
Collected in Q4.2 'Was an abdominal drain present on arrival from theatres?'. Compliance with this process measure is if the patient does **not** have an abdominal drain present on arrival in recovery.

Get patients DrEaMing on postoperative day 1 -
This recommendation aims to get patients drinking, eating, and mobilising quickly after surgery. The data for this measure is collected in Q5.2, Q5.3, Q5.4 of the PQIP webtool. Drinking is defined as 'free fluids'; eating as 'soft diet'; and mobilising as sitting out of bed with maximal assistance of one person, all within 24 hours of surgery ending.

We are encouraging sites to aim for >80% compliance with each of the recommendations.

Next steps

The pomVLAD dashboard is currently under development with a plan to launch at the end of April 2018. The project will run for an initial 12 months.

Local PQIP teams will be notified when the dashboard is live.

PQIP collaborators in sites receiving the dashboard will be invited to take part in telephone interviews to explore their views of the pomVLAD dashboard and how they are using it.

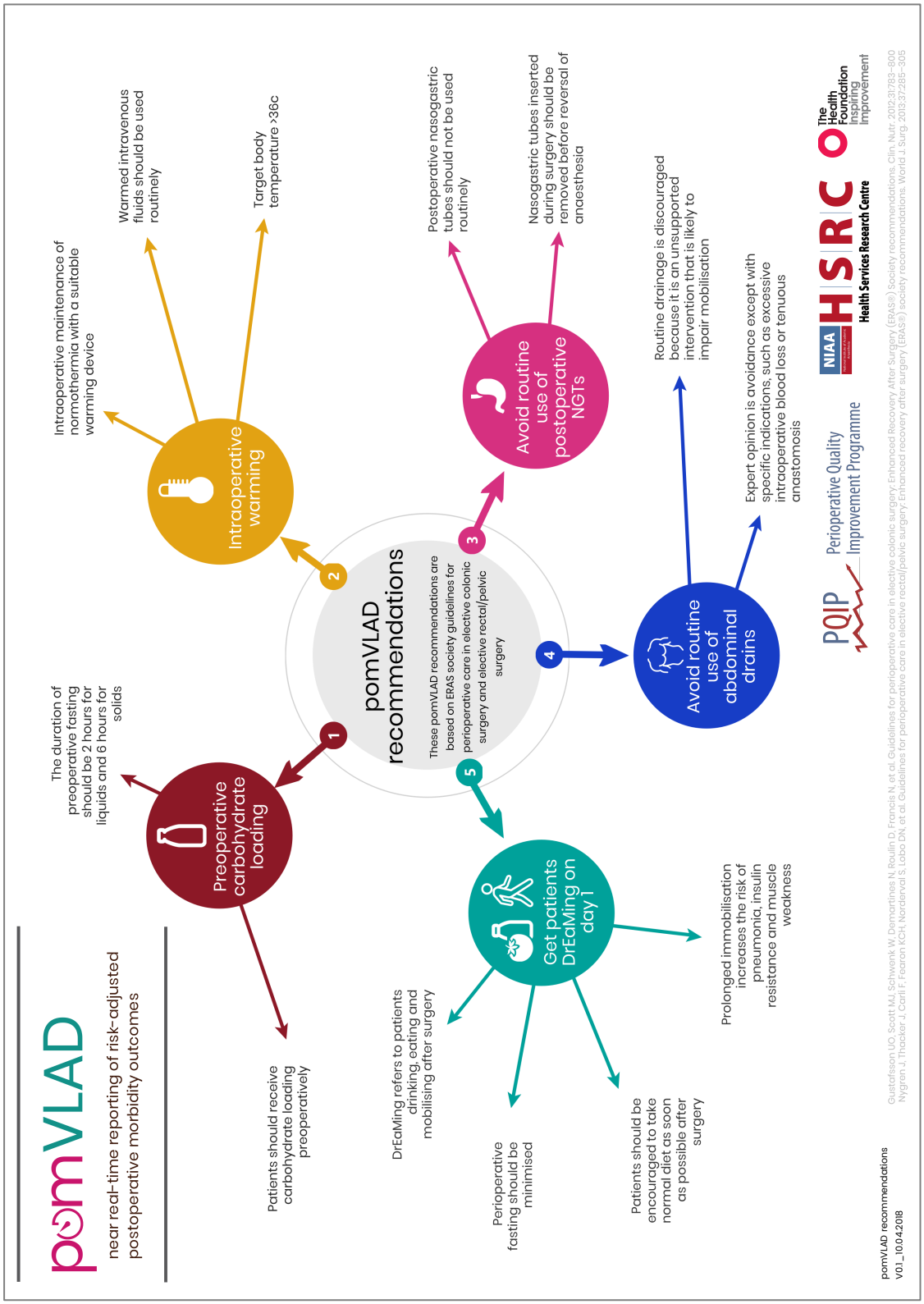
Example pomVLAD chart

The expected risk of postoperative morbidity will be calculated for each patient using a risk-adjustment model developed in over 4000 patients recruited to the PQIP study. This model includes 10 patient and surgical variables and full details of the model will be published in the medical literature.

References

1. Bennett-Guerrero E, Welsby I, Dunn TJ et al. The use of a postoperative morbidity survey to evaluate patients with prolonged hospitalization after routine, moderate-risk, elective surgery. Anesthesia and analgesia 1999; 89: 514–9.
2. Grocott MPW, Browne JP, Meulen JVD, Matejowsky C, Mutch M. The Postoperative Morbidity Survey was validated and used to describe morbidity after major surgery. Journal of Clinical Epidemiology 2007; 60: 919–28.
3. Davies SJ, Francis J, Dilley J, Wilson RJT, Howell SJ, Allgar V. Measuring outcomes after major abdominal surgery during hospitalization : reliability and validity of the Postoperative Morbidity Survey. Perioperative Medicine 2013; 2: 1–9.
4. Wong DJN, Oliver CM, Moonesinghe SR. Predicting postoperative morbidity in adult elective surgical patients using the Surgical Outcome Risk Tool (SORT). British Journal of Anaesthesia 2017; 119: 95–105.
5. Dindo D, Demartines N, Clavien P-A. Classification of Surgical Complications. Annals of Surgery 2004; 240: 205–13.
6. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. Lancet. 1997;350:1128–30.

Appendix D-2 Enhanced recovery recommendations



Appendix D-3 Presentation slides used at introductory events

pomVLAD

near real-time reporting of risk-adjusted postoperative morbidity outcomes

Dr James Bedford
ST5 Anaesthesia/Intensive Care Medicine
PQIP fellow

@jbedford84

PQIP Perioperative Quality Improvement Programme

NIAA HSRC Health Services Research Centre

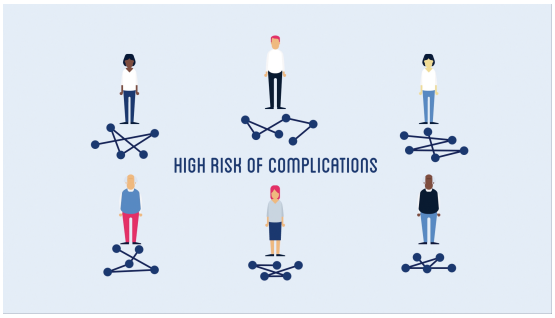
The Health Foundation Inspiring Improvement

Project overview

- Enhanced recovery recommendations; five process measures
- Dashboard display; reporting ER recommendations and risk-adjusted morbidity outcomes
- Top 20 recruiting colorectal sites
- Part of PQIP - no additional data collection

pomVLAD

Reporting outcome data



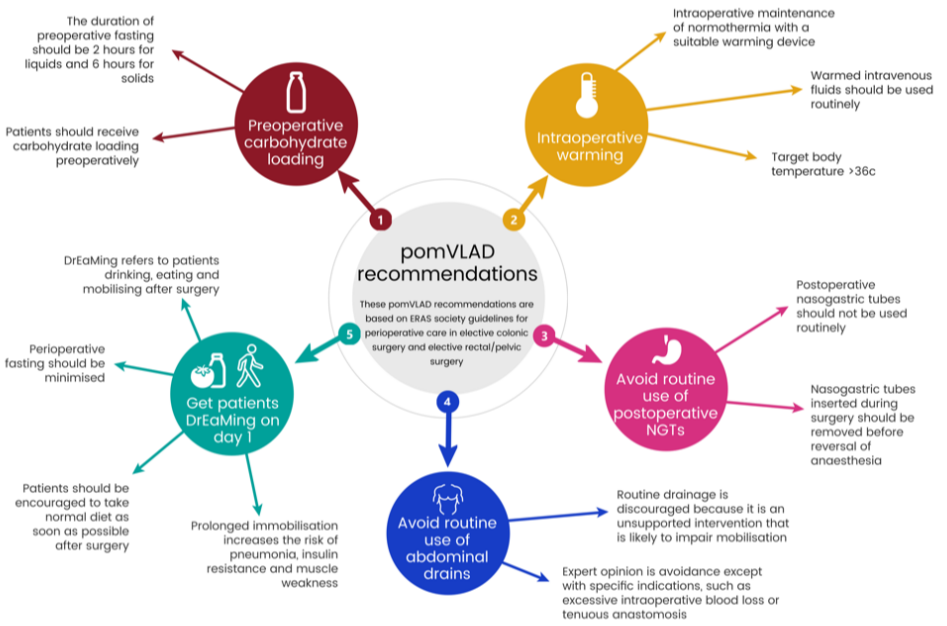
pomVLAD

Defining postoperative complications

- Using the Postoperative Morbidity Survey[1] (POMS) on postoperative day 7
- POMS major[2]
- Major morbidity (e.g. requirement for a new therapy or a diagnosis of a complication) not a minor deviation from usual course (e.g. urinary catheter)

1. Bennett-Guerrero E, Welsby I, Dunn TJ, Young LR, Wahl T a, Diers TL, et al. The use of a postoperative morbidity survey to evaluate patients with prolonged hospitalization after routine, moderate-risk, elective surgery. Anesth. Analg. 1999;89:514–9.
2. Wong DJN, Oliver CM, Moonesinghe SR. Predicting postoperative morbidity in adult elective surgical patients using the Surgical Outcome Risk Tool (SORT).

pomVLAD



Gustafsson UO, Scott MJ, Schwenk W, Demartines N, Roulin D, Francis N, et al. Guidelines for perioperative care in elective colonic surgery: Enhanced Recovery After Surgery (ERAS®) Society recommendations. Clin. Nutr. 2012;31:783–800
Nygren J, Thacker J, Carli F, Fearon KCH, Norderval S, Lobo DN, et al. Guidelines for perioperative care in elective rectal/pelvic surgery: Enhanced recovery after surgery (ERAS®) society recommendations. World J. Surg. 2013;37:285–305

pomVLAD: Variable life-adjusted display

Risk-adjustment model is used to estimate the risk of postoperative morbidity for each patient

The outcome is either: 1 = morbidity or 0 = no morbidity

The chart is then created as a continuous plot using expected minus observed outcome for each subsequent patient

So, if risk of morbidity for a patient is 15%, and that patient has morbidity, then: 0.15 - 1 = -0.85

if they do not have morbidity: 0.15 - 0 = +0.15

pomVLAD

pomVLAD

pomVLAD risk-adjustment model

- Produced using data from 14 months of PQIP, 4466 patients
- Final model includes 10 preoperative patient and surgical variables
- Performance favourable compared to POSSUM and SORT morbidity

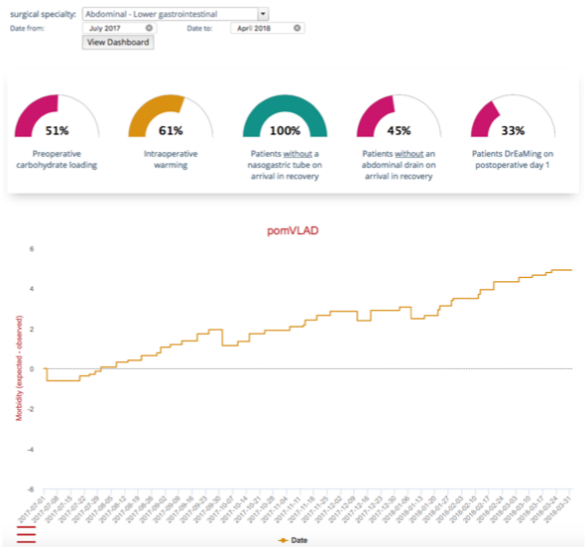
pomVLAD

Risk modelling

Surgical specialty
Severity of surgery
Gender
ASA grade
BMI
Heart rate
Systolic BP
Age (years)
Number of operations in last 30 days
Respiratory history findings

pomVLAD

Dashboard display



pomVLAD

Implementation

Now live!

- 10 sites will receive recommendations and dashboard
- Recommendations accessible to all other sites
- Available through PQIP webtool

pomVLAD

Project evaluation

- Project planned to run for initial 12 months
- Evaluation based on quantitative and qualitative outcomes
- Telephone interviews with teams at some sites receiving the dashboard

pomVLAD

QUESTIONS

twitter: @jbedford84

www.pqip.org.uk
twitter: @PQIPNews

Appendix D-4 Participant information sheet for qualitative interviews

pomVLAD : A rapid feedback evaluation of the pomVLAD project

Participant information sheet

The Perioperative Quality Improvement Programme (PQIP) was launched in 2016 by the National Institute of Academic Anaesthesia Health Services Research Centre, working on behalf of the Royal College of Anaesthetists. The aim of the programme is to investigate perioperative care of patients undergoing major non-cardiac surgery and measure complication rates, failure to rescue and patient reported outcomes. As part of PQIP, in May 2018 the pomVLAD project launched to provide 10 hospital sites with rapid feedback of their risk-adjusted morbidity outcome data using an online dashboard combined with regular reports.

To understand the impact of pomVLAD at local sites, and to inform a wider implementation after the initial pilot we will need to carry out an evaluation to capture staff perceptions of the project, dashboard and reports to help us maximise engagement and use of the data for local quality improvement. The information we capture through the evaluation will directly inform future developments in the dashboard and reports to improve their delivery.

We will carry out a rapid feedback evaluation to understand the factors acting as barriers and facilitators in engagement with the pomVLAD dashboard and reports, and how these differ from existing PQIP reporting mechanisms. We will interview PQIP collaborators based at five sites receiving the pomVLAD intervention and five comparison sites who are participating in PQIP but not receiving the pomVLAD intervention. The study is being carried out by a team from UCL. You can contact us using the details printed at the back of this form.

The aim of this information sheet is to help you understand why we are carrying out this evaluation and what would be required of you if you decide to take part in the study.

1. Who has given ethical approval for the study?

The study has been confirmed as a service evaluation by the Joint Research Office at the University College London Hospitals NHS Trust and University College London. Ethical approval is therefore not required. The study has been registered with University College London Hospitals NHS Trust as a service evaluation.

2. Why have I been asked to take part?

You have been asked to take part because you are currently involved in the delivery of PQIP at your local site. We are interviewing representatives from five sites receiving the pomVLAD intervention and five comparison sites who are not receiving the intervention. We wish to speak with you because we believe you have a valuable perspective on how local sites are using the reports and dashboards, and the potential benefits and limitations of these.

3. What does taking part involve?

If you decide to take part, the researcher will ask you to sign a consent form. After the consent form has been signed, he will liaise with you to arrange the time for an interview. The interview will take place over the telephone at a time that suits you and will last approximately 30 minutes. We would like to carry out two interviews in total with you over the one-year duration of the project.

The interview will include questions on your current role, your hospital, and how you and your site use PQIP data and/or the pomVLAD dashboard and reports, and what your views of them are. The researcher will ask to audio record the interviews. You can ask the researcher not to record the interview or to stop the recorder at any time and they will take notes instead.

4. Do I have to take part?

No, it is up to you to decide whether or not to take part. If you decide to take part, we will ask you to sign a consent form and email it back to us before the interview takes place. You can hold a copy of this consent form. Whether or not you decide to take part in the interview, your employment status or relations will not be affected in any way.

5. Is what I say confidential?

Yes, we will not inform anyone outside the research team that you have participated in the study. Your personal information will not be attached to any information you provide. All information will be stored securely and will only be accessed by members of the research team. We will not identify you by name in any reports or publications. Your personal data will be destroyed within a year of the study's completion, while research data will be archived securely for 20 years after the study's completion, before its eventual destruction.

If you disclose information that the researcher feels has implications for professional practice, we may report these concerns to the head of service or other managers. Any information passed on will be anonymised, ensuring you cannot be identified.

6. What if I change my mind?

You are free to withdraw from the study at any time. You do not have to give a reason for withdrawing. Even if you start an interview, you can stop it at any point if you want to. If you wish to withdraw, please contact us using the details at the end of this sheet.

If you withdraw, we will hold onto the information you provided before withdrawing. If you lose capacity to participate, we will withdraw you from the study automatically. In this case we will also keep the information you provided.

7. What are the risks of taking part?

Helping us with this study will take up a little of your time, but we will do our best to minimise any inconvenience to you by arranging to a time that suits you. If you feel uncomfortable discussing any aspect of this study, you can withdraw from the interview and observations at any time. You can also contact the study team to discuss any concerns you have before and after agreeing to take part.

The researcher who conducts the interview will abide by a professional code of conduct.

8. What are the benefits of taking part?

There may be limited personal benefits emerging from the study, but the evaluation will help inform the implementation of PQIP and the pomVLAD project. The findings from this study will be presented to the staff leading the project on an on-going basis, pointing out any problems with the design and implementation of pomVLAD. Our documentation of how PQIP reports/dashboards and the pomVLAD dashboard are being used and may be improved will provide valuable lessons on the development of similar web-based dashboards in healthcare settings.

9. What will happen to the results of the research study?

We will report our findings to pomVLAD and PQIP staff and other people interested in the project after each round of interviews, using reports and presentations. At the end of our study, we will produce a full report. We will publish our findings in scientific journals and present them at national and international scientific meetings and conferences. Your name will not be used at any time. Our final report will also be disseminated to staff and we will be happy to send you a copy. You can also let the researcher know if you are interested in receiving updates on the study.

10. What happens if something goes wrong?

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated by members of staff you have met through your participation in the research, National Health Service complaints mechanisms are available to you. For further information, please see the Sponsor's website:
http://www.uclh.nhs.uk/PandV/Helpandsupport/Commentssuggestionsandcomplaints/

Alternatively, you may wish to contact the Principal Investigator (details below). Please note that NHS Indemnity does not offer no-fault compensation i.e. for non-negligent harm, and NHS bodies are unable to agree in advance to pay compensation for non-negligent harm.

11. Where can I find out more about the research?

Further information can be found by contacting the study team:

Principal Investigator
Dr James Bedford - [redacted]

Co-Investigators
Dr Cecilia Vindrola - [redacted]
Dr Duncan Wagstaff - [redacted]
Dr Sonya Crowe - [redacted]
Prof Ramani Moonesinghe - [redacted]

Should you require independent advice, you may wish to contact the UCL/UCLH Joint Research Office on 020 3447 5557.
You may also wish to look up the INVOLVE website at: www.invo.org.uk.

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION AND FOR CONSIDERING HELPING WITH OUR STUDY

Appendix D-5 Consent form for qualitative interviews

pomVLAD : A rapid feedback evaluation of the pomVLAD project

Consent form

Name of researcher:

Please read the following statements and mark the boxes to show you agree

1. I confirm that I have read and understand the information sheet (Version 0.2, 19/07/2018) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my legal rights or employment status being affected. ☐
3. I understand that my participation in these discussions will be audio recorded and stored anonymously. ☐
4. I understand that relevant sections of my anonymised data collected during the study may be looked at by members of the research team. I give permission for these individuals to have access to this data. I understand that this information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. ☐
5. I understand that data and quotations I provide may be used (anonymised fully) in future publication of this study. ☐
6. I understand that, in the event of my withdrawal from the study, data I provide prior to my withdrawal will be retained (anonymised fully) for analysis and publication. ☐
7. I agree to take part in the above study. ☐

Name of participant	*Date*	*Signature*
Name of person receiving consent form	*Date*	*Signature*

Appendix D-6 Interview topic guide

pomVLAD : A rapid feedback evaluation of the pomVLAD project

Interview topic guide: Round 1 interviews

Thank you for agreeing to take part in this interview.

Interviewee role

1. It would be good to begin with a description of your current role within your hospital/trust
2. How long have you been in this role?
3. How you held any previous roles within your trust?

Exploring hospital context

1. Does your hospital have electronic health records?
2. Do you have designated office space with access to a computer?
3. Do you have access to any non-PQIP dashboards to monitor performance in your hospital?
4. Do you feel your team has adequate financial support, resources and time to allow QI activity within your trust?

Interviewee interactions with PQIP

1. What is your current role for PQIP within your hospital?
2. How much time do you think you spend each week on PQIP related activities?
3. Do you have time allocated to your PQIP role to support QI activity?
4. Are your aware of the PQIP online dashboards? If yes, how often do you access them?
5. How/where do you access the PQIP reports and dashboards?
6. Do you feel there are any benefits of using online dashboards over the quarterly reports?
7. Are there any barriers to you accessing and using the PQIP dashboard or quarterly reports?
8. How does your trust share information from the dashboards and quarterly reports?
9. Have you or your team identified any improvement opportunities from your PQIP quarterly reports, existing PQIP dashboards (not pomVLAD) or based on the top 5 improvement opportunities from the annual PQIP report?
10. If yes, how are you planning to address these opportunities, and how will you measure any change?
11. Are there ways that we could improve the delivery of the PQIP quarterly reports or dashboards?

Views on pomVLAD intervention

1. What are the aims of the pomVLAD project?
2. Do you feel the pomVLAD project offers any additional benefit to the dashboard and reports already produced by PQIP?
3. What are your views of the enhanced recovery recommendations made as part of the project?
4. Have you or your team identified any improvement opportunities based on the pomVLAD dashboard?
5. If yes, how do you plan to address these opportunities and how will you measure any subsequent change?
6. Have you accessed the pomVLAD dashboard? If so, have you shared the dashboard information with colleagues? How?
7. How frequently do you access the pomVLAD dashboard? If/when you do access it, how was your experience of using it?
8. Are there any barriers to you accessing and using the dashboard data?
9. Are there any factors that you feel support interaction and engagement with your local data?
10. How do you feel we could improve the pomVLAD dashboard to better support you?

Interview topic guide: Round 2 interviews

This topic guide may be modified based on information gained during the first round of interviews. We will explore in more detail any emerging themes.

Thank you for agreeing to take part in this interview.

General views on data feedback within PQIP and pomVLAD

1. To start with a general question, can you tell me how you're finding being involved in PQIP?
2. How do you feel PQIP is running in your hospital?
3. How do you feel about the way data is fedback within PQIP?

Interviewee interactions with PQIP

1. Do you feel access to PQIP data has changed your understanding care delivery in your hospital? How has it changed your understanding?
2. Can you compare the use of the PQIP quarterly reports to the dashboards in your hospital?
3. How do you share information in your trust from the dashboards and quarterly reports? Are the dashboards or reports used more to share information? Why do you think that is?

Use of data for quality improvement

1. Have you or your team identified any opportunities for improvement based on PQIP (or pomVLAD) data?
2. How are you planning on address these opportunities, and how will you measure any changes?
3. Do you feel dashboards or reports will adequately support you to monitor any change?

Triggers for accessing data

1. Are there any triggers that increase your likelihood of looking at local data? Is there anything that makes you sit down and review local data?
2. Do these triggers apply to the reports and dashboard data, or one more than the other? Why do you think that is?
3. Is there anything you feel inhibits or stands in the way of you looking at and using your local data? Are these the same for the dashboards and reports? If it impacts on one more than the other, why do you think that is?

Views on pomVLAD intervention

1. Do you feel having access to the pomVLAD dashboard has changed your understanding of local outcomes? For better or worse?
2. Do you feel your engagement with the dashboard (pomVLAD or PQIP) has changed over the 12-month pilot project? Why? How?
3. How frequently do you access the pomVLAD dashboard? If/when you do access it, how was your experience of using it?
4. How have you and colleagues found interpreting your local VLAD?
5. Are there any improvements you feel could be made to the dashboard? Is there anything that you feel should be displayed or analysed differently?
6. Have you discussed the VLAD dashboard with colleagues? How did they feel about it?

Potential updates to pomVLAD dashboard – control limits and support for interpretation

1. Do you feel the addition of control limits or confidence intervals would help or hinder you in interpreting your local data?
2. Are there any concerns you might have with the addition of control limits to the VLAD?
3. Have you managed to review the draft interpretation and investigation documents sent via email? If yes, do you feel the publication of these documents will change understanding of the VLAD dashboard or how it is used locally? Will this change be beneficial or a negative change?

Interviewee role

1. Since our previous interview has your role within your hospital or PQIP changed?

Exploring hospital context

1. Have there been any changes in your trust that your feel has improved your ability to deliver QI work, for example having time allocated in your job plan, increased financial support, or improved staffing?
2. Have there been any changes in your trust that you feel have been detrimental to your ability to deliver QI work?

Appendix E

Appendix E-1 Example quarterly report provided by PQIP study to local site investigators

Table of Contents

PQIP Perioperative Quality Improvement Programme

Sample Hospital report
July 2018

This report includes patients who underwent surgery during the period **01 December 2017** to **27 July 2018**. The patient recruitment information includes all patients recruited to PQIP (locked and unlocked case records). **The remainder of the report is based only on records that were locked at the time of the report production (27 July 2018 13:53).**

RCoA Royal College of Anaesthetists · NIAA HSRC Health Services Research Centre · The Health Foundation Inspiring Improvement · source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

PQIP's Top 5 Improvement Opportunities for 2018-19

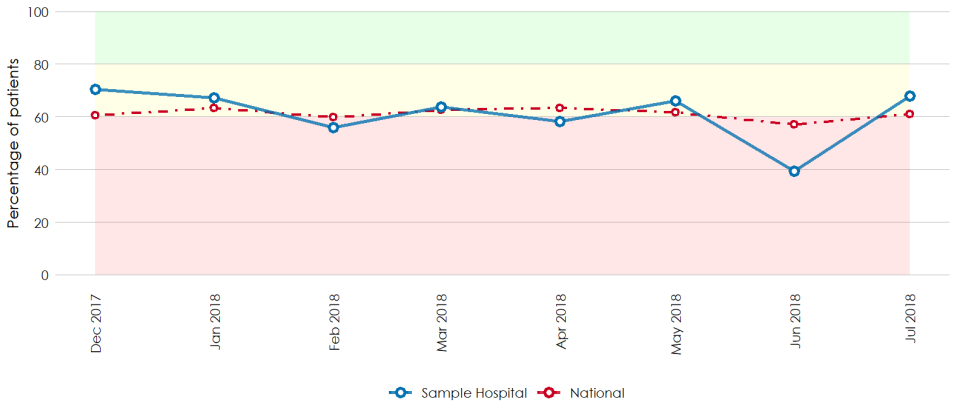
PQIP released the first annual report in April 2018 which is available to download and view on the PQIP Website. As part of the annual report the top 5 national improvement opportunities have been highlighted in section 1 of the report.

Anaemia & Diabetes

Anaemia

New guidelines suggest that men and women should be considered anaemic if their haemoglobin is less than 13g/dL. Preoperative anaemia is associated with higher morbidity , length of stay and mortality in major non-cardiac surgery. The 2017 consensus statement on the peri-operative management of anaemia and iron deficiency can be adapted to your local context, it can be found here. Figure 1 below shows the percentage of patients who had a recorded preoperative haemoglobin that was above 13g/dL. Between 01 December 2017 and 27 July 2018 152 patients were anaemic. Of these 129 (84.9%) were having elective operations rather than expedited or urgent operations.

Figure 1: Percentage of patients who had a haemoglobin above 13g/dl preoperatively by month of surgery



RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

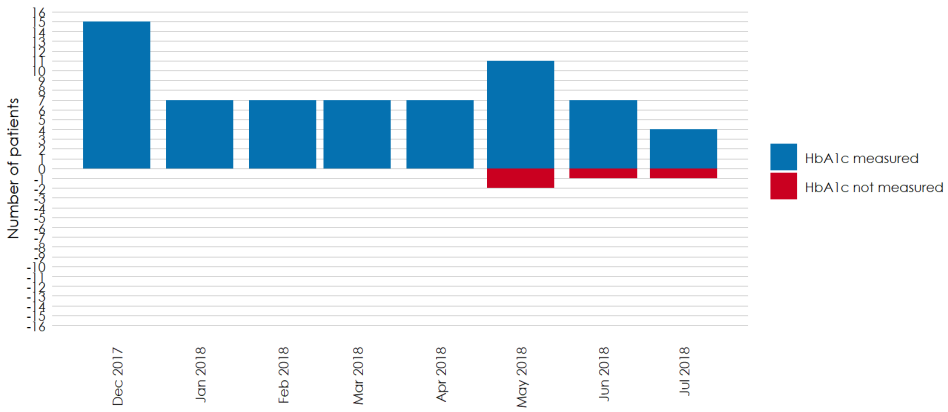
PQIP Perioperative Quality Improvement Programme

HbA1c testing

National Guidelines state that all diabetic patients should have a HbA1c measured before elective surgery. At Sample Hospital 17.1% of patients recruited to PQIP were recorded as being diabetic.

Figure 2 below shows the number of diabetic patients who did and did not have a recorded preoperative HbA1c test

Figure 2: Number of diabetic patients who had a preoperative HbA1c test preoperatively by month of surgery.



The recommended upper threshold for preoperative HbA1c is 8.5 mmol/mol . If higher than this consideration should be made to postponing the surgery if possible. Between 01 December 2017 and 27 July 2018 65 HbA1c tests were performed, of which 12 % were above 8.5 mmol/mol.

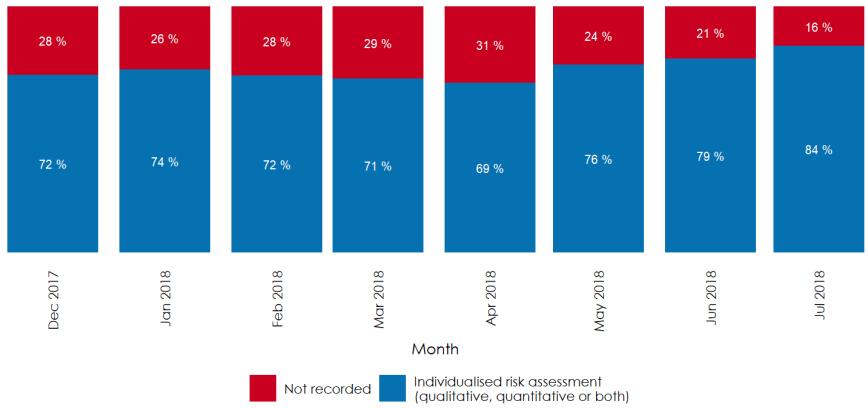
RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Individualised Risk Assessment

Preoperative risk assessment can be used to plan appropriate perioperative care and to inform the process of shared decision-making. The 2011 NCEPOD report 'Knowing the Risk' identified that formally documented assessment of perioperative risk is infrequently and often inadequately performed. The Montgomery v Lanarkshire Health Board case of 2015 brought the law in line with previous GMC professional guidance, drawing attention to the responsibility of a doctor to provide all material risks to a patient. The 2017 AAGBI Consent for Anaesthesia guidelines recommend that anaesthetists should record their discussion in the patient record, noting any individualised risks, benefits and alternatives that were explained. Figure 3 below shows the percentage of patients at Sample Hospitalthat have a documented individualised risk assessment by month of surgery.

Figure 3: Percentage of patients who had an individualised risk assesment preoperatively by month of surgery.



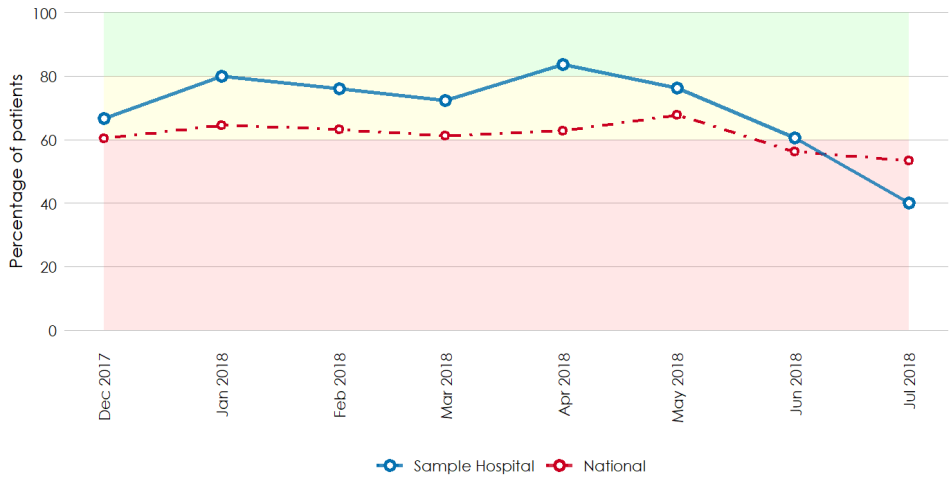
Enhanced Recovery

Enhanced recovery pathways provide individualised protocolised care to reduce complications and length of stay. There is variation between specialties and hospitals in terms of enhanced recovery pathway adoption and compliance. Enhanced recovery after surgery (ERAS) pathway data was available for 404 patients, of which 293 (72.5 %) were enrolled on an ERAS pathway from 01 December 2017 to 27 July 2018 .

RCoA Royal College of Anaesthetists · NIAA HSRC Health Services Research Centre · The Health Foundation Inspiring Improvement · source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Figure 4: Patients enrolled on an ERAS pathway by month of surgery



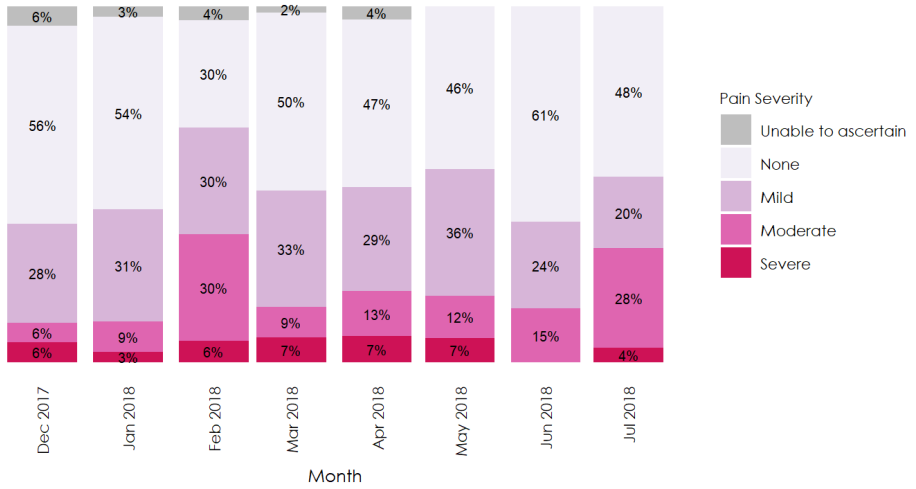
RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Individualised Pain Management

Poorly controlled pain after surgery can contribute to the stress response if not adequately treated. Optimal analgesia can aid the return to normal function and minimise complications after surgery. It is an RCoA-recommended quality measure that no patient should be discharged from recovery with uncontrolled pain. PQIP collects information on the maximum pain score recorded in recovery. Pain score data was complete for 404 out of a possible 404 patients (100%) between 01 December 2017 and 27 July 2018. Figure 5 below shows maximum pain scores in recovery for Sample Hospital by month of surgery.

Figure 5: Maximum pain scores recorded in recovery



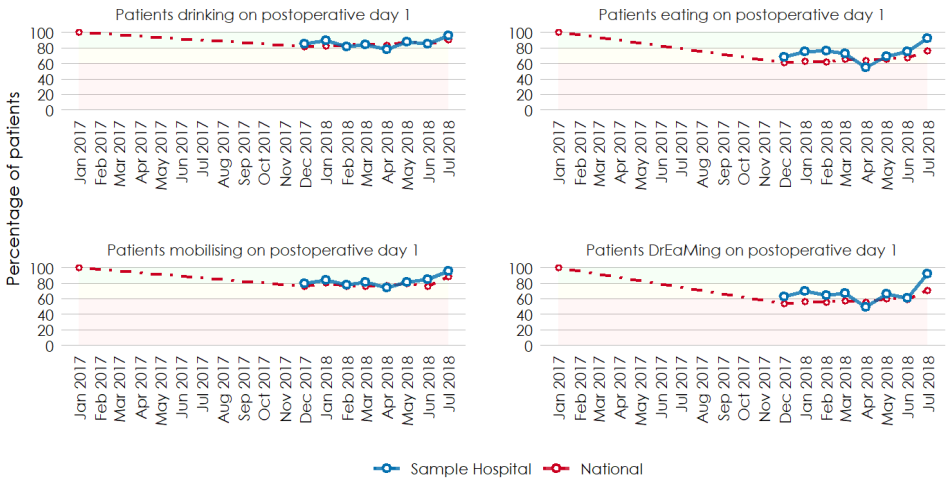
PQIP Perioperative Quality Improvement Programme

Drinking, Eating & Mobilising - DrEaMing

DrEaM (drinking, eating and mobilising) is an international initiative supported by anaesthetists and surgeons working together with patients. PQIP measures the proportion of patients who are drinking (at least free fluids), eating (at least soft diet), and mobilising (sitting out of bed, with maximum assistance of one person) by the end of day 1 postoperatively.

Figure 6 shows the percentage of patients who were drinking, eating and mobilising by the end of day 1. The final graph (bottom right of the four) shows the percentage of patients who had completed all three of those measures and were 'DrEaMing'. DrEaM data was available for every patient recruited to PQIP between 01 December 2017 and 27 July 2018

Figure 6: Patients drinking, eating and mobilising on postoperative day 1



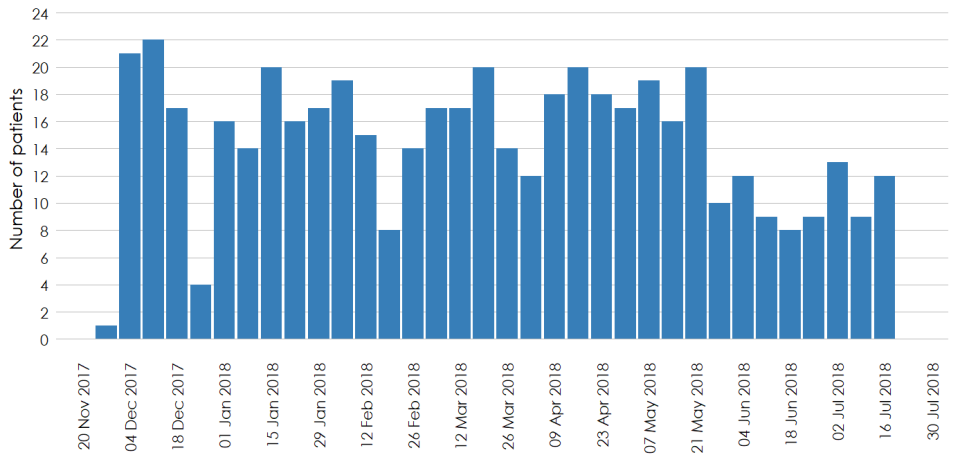
PQIP Perioperative Quality Improvement Programme

Recruitment

Patient recruitment

From 01 December 2017 to 27 July 2018 Sample Hospital recruited a total of 494 patients to PQIP. Figure 7 below shows recruitment to PQIP by week. It is not possible display a recruitment target as one is not recorded. If you would like to update your recruitment target please contact the PQIP team .

Figure 7: Recruitment of patients to PQIP by week
(locked and unlocked case records)



Of the 494 patients recruited, there were 404 case records locked. The remainder of this report is based on locked records. The number of patients recruited by each surgical speciality is shown in table 1.

Table 1: Number of patients recruited per surgical specialty

Surgical specialty	Number of patients
Urology	101
Abdominal - Upper gastrointestinal	51
Abdominal - Lower gastrointestinal	104
Abdominal - Hepatobiliary	1
Thoracics	117
Head and neck	1
Abdominal - other	29
Orthopaedics	0
Spinal	0
Burns & Plastics	0
Not recorded	0

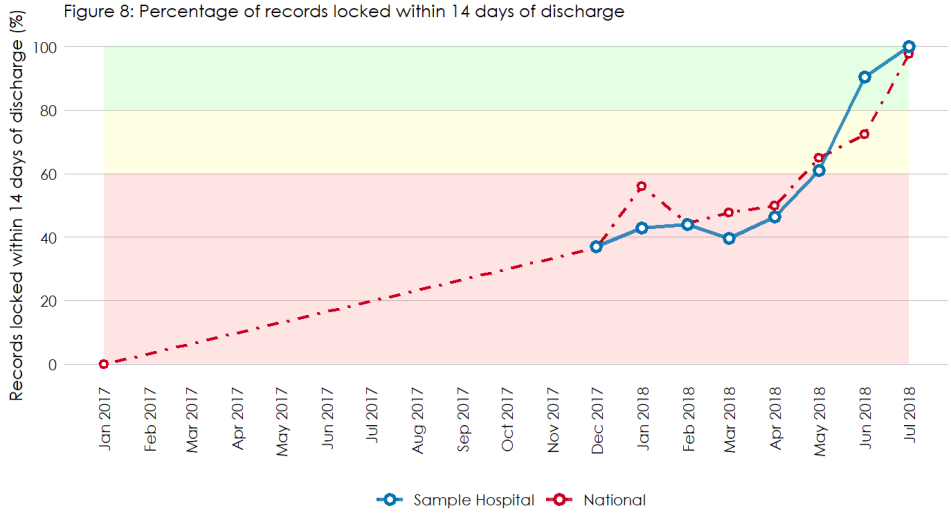
RCOA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Records locked within 14 days of discharge

The percentage of records locked within 14 days of discharge is used by PQIP as a measure of hospital engagement with the study. Figure 8 shows the percentage of records included within this reports that were locked within 14 days of discharge by month.

Figure 8: Percentage of records locked within 14 days of discharge



The overall proportion of records locked within 14 days of discharge or death was 52.1 %, (monthly range 37 % to 100 %). between 01 December 2017 and 27 July 2018.

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

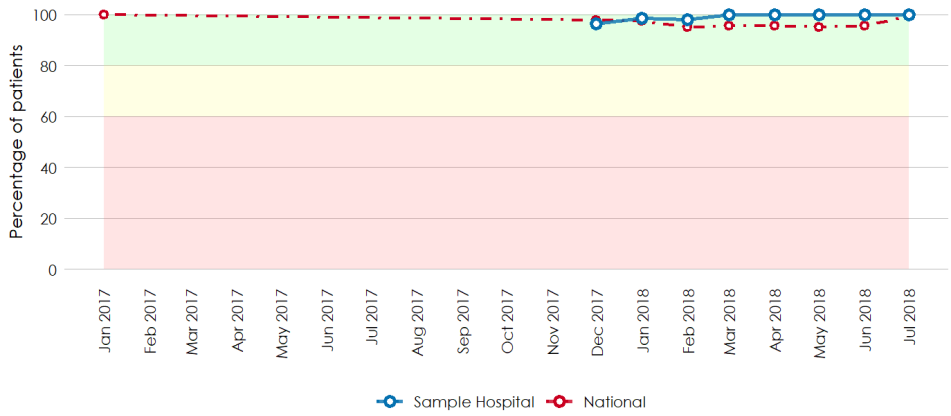
Preoperative

Preoperative assessment

The 2011 NCEPOD report Knowing the Risk recommended that all elective high-risk patients should be seen and fully investigated in pre-assessment clinics in order to improve patient care, improve theatre utilisation, reduce bed occupancy, and reduce errors due to inadequate communication.

A total of 400 out of 404 patients underwent face to face preoperative assessment between 01 December 2017 and 27 July 2018.

Figure 9: Patients undergoing face-to-face preoperative assessment



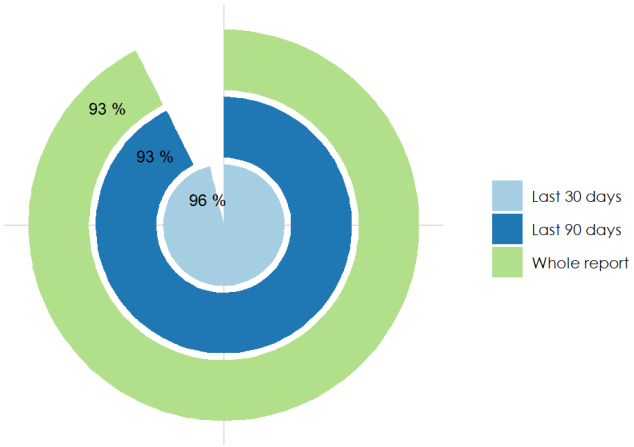
RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Admission to hospital on the day of surgery

Admission on the day of surgery is a key aspect of enhanced recovery and has been shown to have an effect on length of stay by Simpson et al. Between 01 December 2017 and 27 July 2018 374 out of 404 patients (92.6%) were admitted on the day of surgery. Figure 10 shows the percentage of patients admitted on the day of surgery over the last 30 days, 90 days and throughout the period of this report.

Figure 10: Percentage of patients admitted on the day of surgery



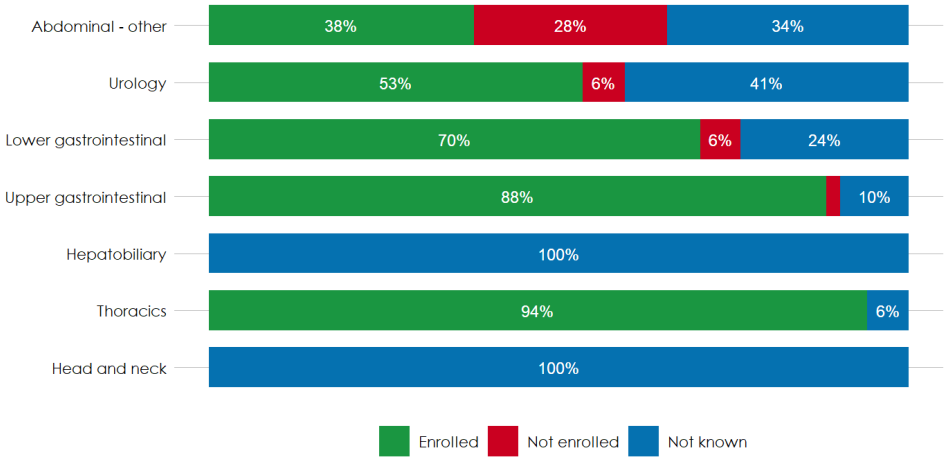
RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Enhanced Recovery by specialty

At Sample Hospital 72.5 % of patients were enrolled on an ERAS pathway from 01 December 2017 to 27 July 2018 . Figure 11 below demonstrates how each of the specialties at Sample Hospital is adopting enhanced recovery

Figure 11: Percentage of patients enrolled on a enhanced recovery pathway by specialty



RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

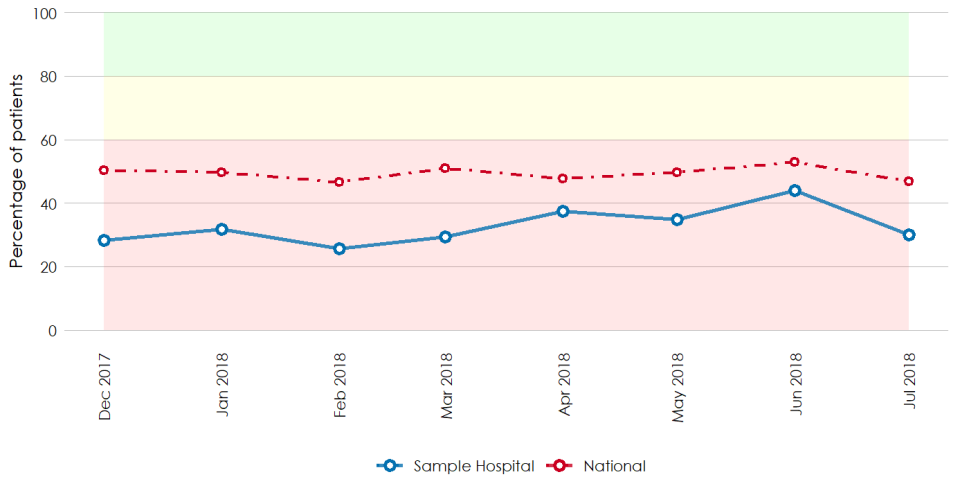
PQIP Perioperative Quality Improvement Programme

Preoperative carbohydrate loading

The ASER/POQI joint consensus statement (2016) recommends that, with the exception of type I diabetics, pre-operative complex carbohydrate (e.g. maltodextrin) be used when available to improve insulin sensitivity and reduce protein catabolism. Alternatively, clear fluid used to maintain oral hydration pre-operatively should contain at least 45g of carbohydrate. This evidence relates to colorectal surgery, the use of preoperative carbohydrates may not be appropriate in every setting.

Between 01 December 2017 and 27 July 2018 of the 335 non-diabetic patients attending Sample Hospital , 108 patients received preoperative carbohydrate drinks. Figure 12 shows the percentage of patients who received preoperative (approximately 2 hours) carbohydrate drinks by month of surgery.

Figure 12: Percentage of patients receiving preoperative carbohydrate drinks



RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

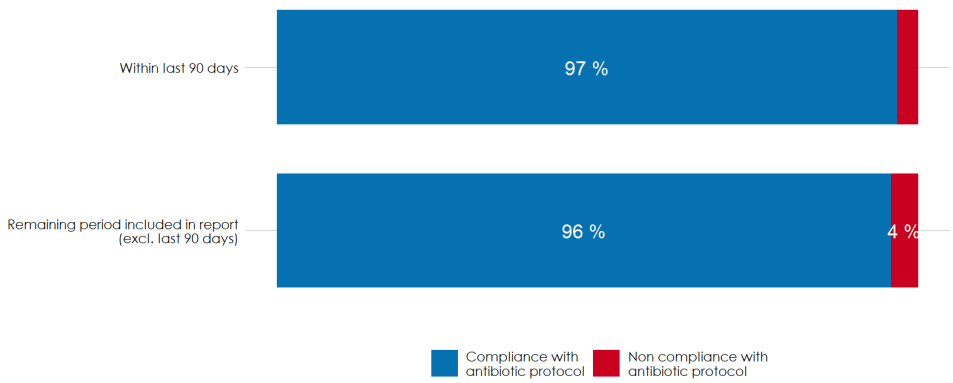
PQIP Perioperative Quality Improvement Programme

Intraoperative & Recovery

Administration of prophylactic antibiotics

Antibiotic prophylaxis is recommended by NICE Guidelines for preventing surgical site infections in certain procedures. Between 01 December 2017 and 27 July 2018 388 out of 404 patients (96%) received antibiotics according to local protocols within 60 minutes before knife to skin. Figure 13 shows the adherence with local antibiotic protocols over the last 90 days and the remaining period included within this report. It is expected that 100% of patients follow local protocols for prophylactic antibiotics.

Figure 13: Patients receiving antibiotics within 60 minutes before knife to skin according to local protocol (%)



RCoA Royal College of Anaesthetists · NIAA HSRC Health Services Research Centre · The Health Foundation Inspiring Improvement · source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Patients arriving in recovery with a body temperature below 36ºc

NICE guidelines define hypothermia as a core body temperature below 36ºC and recommend that body temperature is measured in the perioperative period. Of the 404 recruited patients between 01 December 2017 and 27 July 2018 at Sample Hospital 29 (7%) had a temperature below 36ºc in recovery. Figure 14 below demonstrates the proportion of patients that had a temperature above and below 36ºc in recovery by month of surgery

Figure 14: The percentage of patients that were normothermic in recovery by month of surgery.

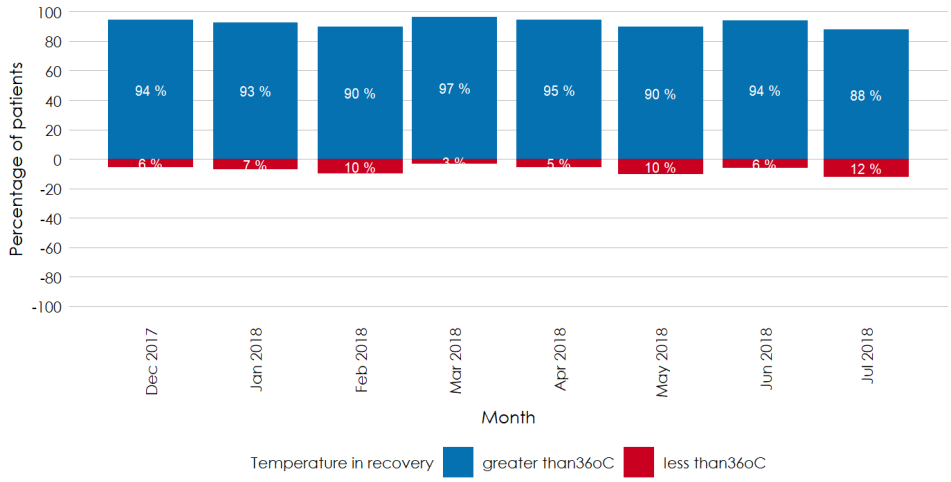


Table 2 below demonstrates the number of warming devices that were applied in theatre to those patients that had a recovery temperature below 36ºC.

Table 2: Number of warming devices applied in theatre to those patients with a recovery temperature below 36ºC

Number of warming devices used	Number of patients	Percent (%)
0	1	3
1	5	17
2	16	55
3	7	24

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Critical care admission when risk of death is greater than or equal to 5%

The 2011 NCEPOD report 'Knowing the Risk' identified that postoperative care of the high risk surgical patient needed to be improved. Any patient with a predicted mortality greater than 5% should be considered high risk and considered for critical care admission postoperatively.

Mortality risk within this report is calculated using the Surgical Outcome Risk Tool (SORT). SORT is a preoperative risk prediction tool for death within 30 days of surgery that has been developed and validated for use in inpatient non-neurological, non-cardiac surgery in adults (aged 16 or over).

Table 3: Patients admitted to critical care where risk of death is calculated to be greater than or equal to 5%

Date range	Patients with risk of death greater or equal to 5%	Number admitted to HDU/ICU
Last 30 days	0	0
Last 90 days	7	7
During time of this report	37	33

A SORT mortality prediction could be calculated for all 404 patients between 01 December 2017 and 27 July 2018. Where a patient underwent a different procedure to the one planned, the SORT mortality risk has been calculated for the actual procedure performed.

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Postoperative

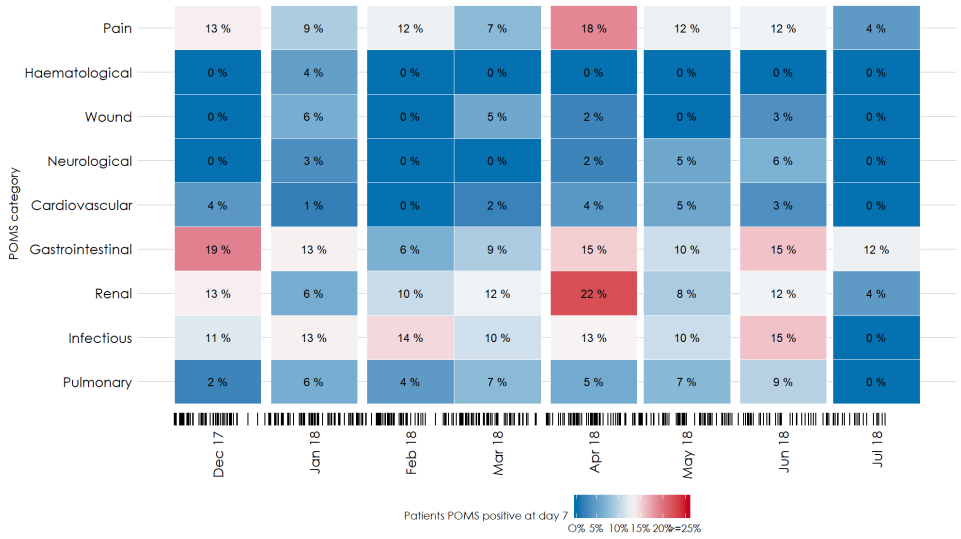
These measures of postoperative recovery and morbidity can be used to assess how well patients recover following surgery against expected or predicted morbidity, and act as a focus for quality improvement. They can be used to:

- Measure the benefits and harms of treatments
- Benchmark services
- Identify potential outliers
- Support your case for change
- Inform the consent process and aid shared decision-making.

Postoperative morbidity survey at day 7

The Postoperative Morbidity Survey (POMS) was first published in 1999 by Bennett-Guerrero et al. and has been validated to describe morbidity following major surgery of a nature that would prolong length of stay. It consists of an 18-point survey which characterises morbidity into nine organ systems. For patients who are recruited into the PQIP study POMS is calculated at day 7. Patients who have been discharged from hospital by day 7 are assumed to have a POMS of zero.

Figure 15: Percentage of patients POMS positive at day 7 by domain



Between 01 December 2017 and 27 July 2018 there were 140 out of 404 patients still in hospital on postoperative day 7 (34.7%). Of those remaining in hospital 96 (68.6%) were Postoperative Morbidity Survey positive. POMS data was complete for all 404 patients recruited.

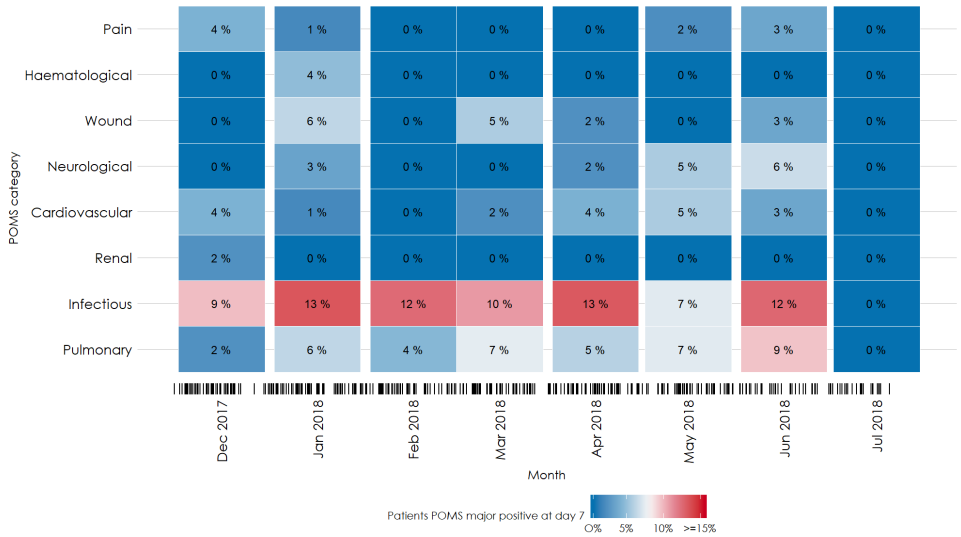
RCoA Royal College of Anaesthetists · NIAA HSRC Health Services Research Centre · The Health Foundation Inspiring Improvement · source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Major postoperative morbidity at day 7

Of those patients who were overall POMS positive between 01 December 2017 and 27 July 2018, there were 54 patients who were POMS major positive. The categorisation of POMS criteria into minor and major was first proposed by Wong et al. and further details of mapping of POMS sub-domains to the Clavien-Dindo classification of surgical complications can be found in appendix 7.1.

Figure 16: Percentage of patients POMS major positive at day 7 by domain



Complications experienced by patients recruited to PQIP

PQIP records the grade of complication of experienced by patients. Table 4 shows the maximum grade of complication experienced by patients at Sample Hospital according to the Clavien-Dindo classification over the last 30 days, 90 days, and throughout the whole period 01 December 2017 to 27 July 2018. Clavien Dindo-complication data was available for 404 out of 404 patients recruited to PQIP.

Table 4: Complications experienced by patients (raw number and (% of patients with data available))

Maximum grade of complication	Last 30 days	Last 90 days	Period of report
Grade II or below	27 (100%)	110 (90%)	331 (82%)
Grade IIIA - Intervention not under general anaesthesia	0	3 (2%)	46 (11%)
Grade IIIB - Intervention under general anaesthesia	0	3 (2%)	12 (3%)
Grade IVA - Single organ dysfunction (including dialysis)	0	5 (4%)	12 (3%)
Grade IVB - Multi-organ dysfunction	0	0	2 (0%)
Grade V - Death	0	1 (1%)	1 (0%)

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

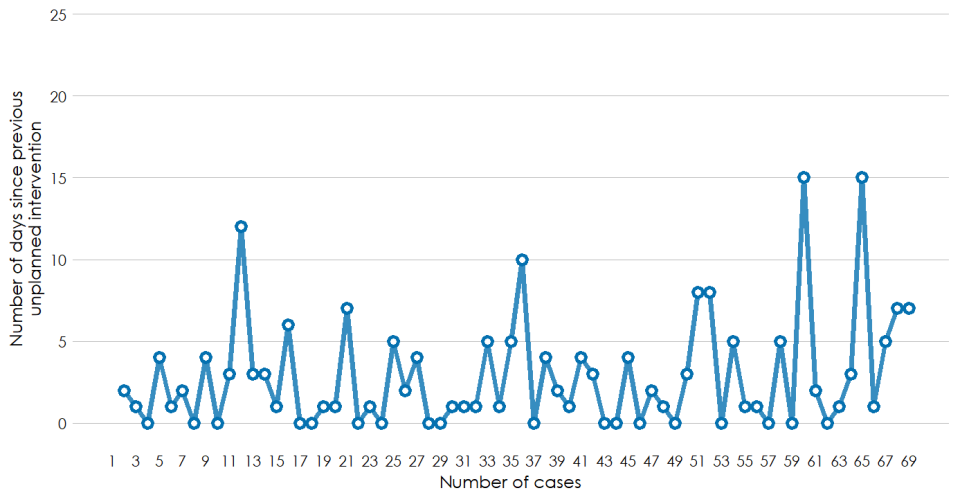
Need for unplanned intervention postoperatively

Between 01 December 2017 and 27 July 2018 69 patients experienced **at least** a grade III Clavien-Dindo complication (see section: "Complications experienced by patients recruited to PQIP" for more detail on the classification of postoperative complications). This includes surgical, endoscopic or radiological intervention (whether under general anaesthesia or not), unplanned HDU/ICU admission and death. Figure 17 shows the time between these events, with a mean duration of 2.9 days.

For this graph the x-axis is the number of cases who have had an unplanned admission to HDU. The y-axis is the number of days from the date of surgery of case 1 to the date of surgery of case 2, and then from the date of surgery of case 2 to case 3 and so on. This is the same type of graphical representation that is used to show rare events in healthcare, such as hospital MRSA bacteraemias or C. difficile infections. **Good performance is indicated by few events with long periods between them (i.e. higher on the y-axis), poorer performance would be shown by both a higher number of events and short periods between them.**

More information on rare events charts can be found in this article by Benneyan et al. within the section 'Surgical site infections'.

Figure 17: Unplanned postoperative intervention (Clavien-Dindo grade III, IIIa, and IIIb complications)



RCOA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

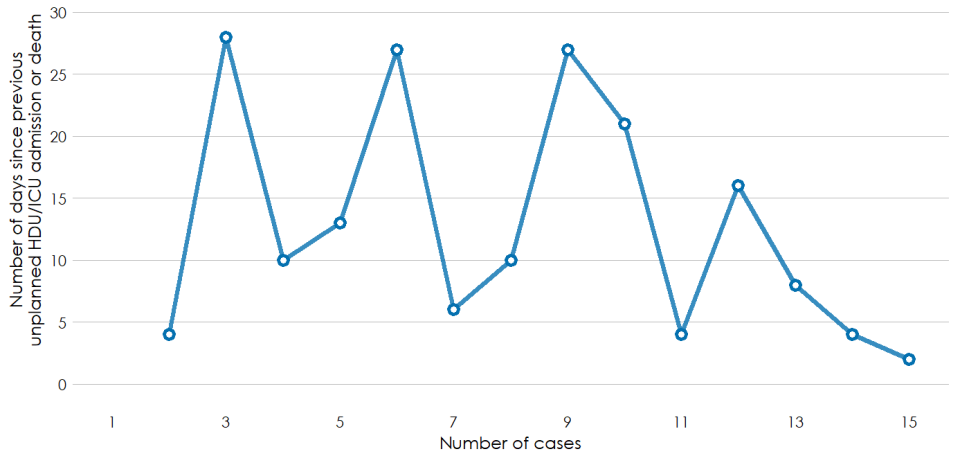
PQIP Perioperative Quality Improvement Programme

Unplanned HDU/ICU admissions

Data for unplanned HDU/ICU admissions is gathered via the Clavien-Dindo classification of surgical complications. Grade 4A, 4B and grade 5 (death) complications are included to produce a 'T chart', also known as a 'rare events chart' which plots the number of cases against time since the previous case. **Good performance is indicated by few events with long periods between them (i.e. higher on the y-axis), poorer performance would be shown by both a higher number of events and short periods between them.**

As PQIP does not collect the date of unplanned HDU/ICU admission, we use the date of surgery as a surrogate marker, and calculate the time between initial surgery for each case. This interval time is plotted on the y-axis. More information on rare events charts can be found in this article by Benneyan et al. within the section 'Surgical site infections'.

Figure 18: Unplanned admissions to HDU/ICU



Between 01 December 2017 and 27 July 2018 there were 15 unplanned HDU/ICU admissions (including deaths) during the period included within this report. The mean duration between date of surgery for these cases was 12.9 days.

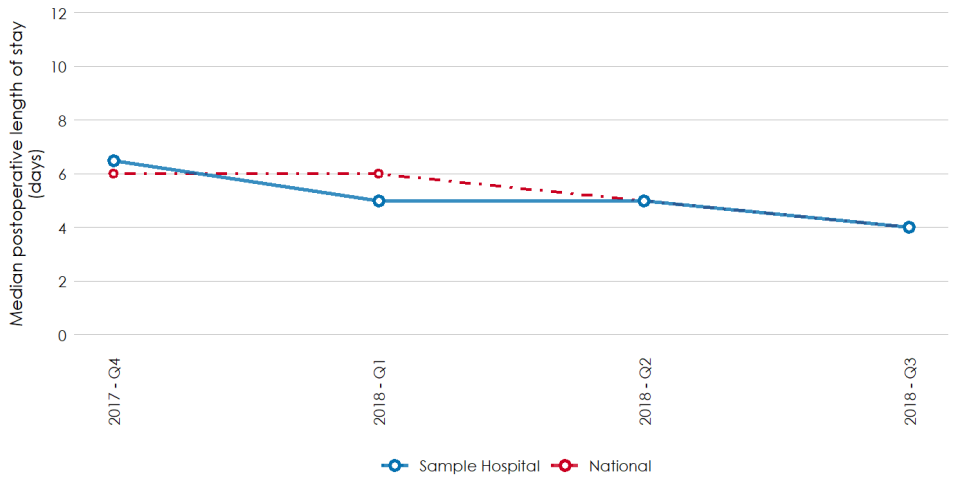
Postoperative length of stay

Figure 19 shows the median postoperative length of stay by quarter of surgery. Between 01 December 2017 and 27 July 2018 the overall median length of stay was 5 days for this report (IQR 3 - 9, mean LOS = 7.9 days).

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Figure 19: Postoperative length of stay by quarter of surgery



RCOA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

Patient reported outcome measures

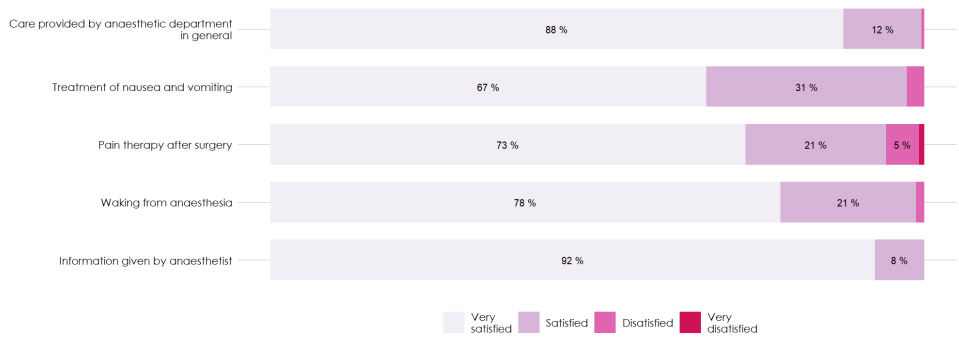
Patient reported outcome measures can be used to assess patient experience and recovery following surgery, and act as a focus for quality improvement. They can be used to:

- Measure the benefits and harms of treatments.
- Benchmark services
- Identify potential outliers
- Support your case for change
- Inform the consent process and aid shared decision-making

Bauer patient satisfaction with anaesthesia care survey

The figure below shows the satisfaction of patients recruited to PQIP with the perioperative anaesthesia care received at Sample Hospital. Between 01 December 2017 and 27 July 2018 a total of 229 patients completed the Bauer patient satisfaction survey, of which 229 (100%) would recommend the anaesthetic service to friends and family.

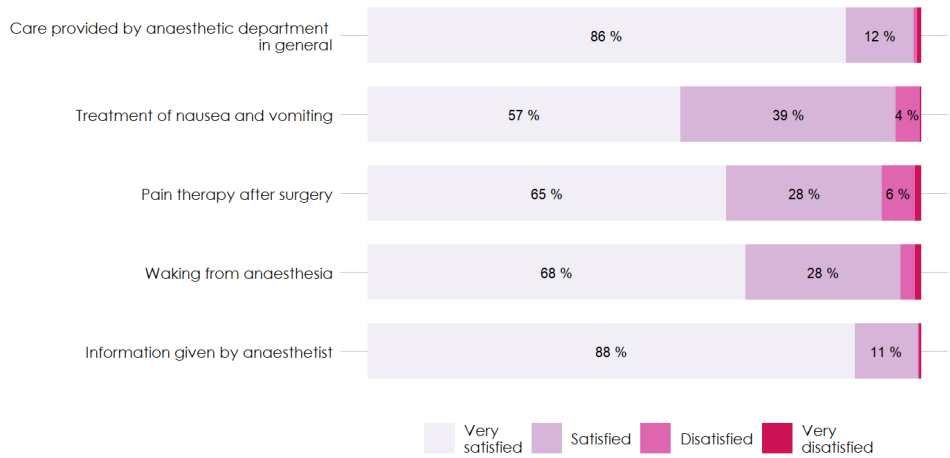
Figure 20: Patient satisfaction with anaesthesia care (Bauer survey)



RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

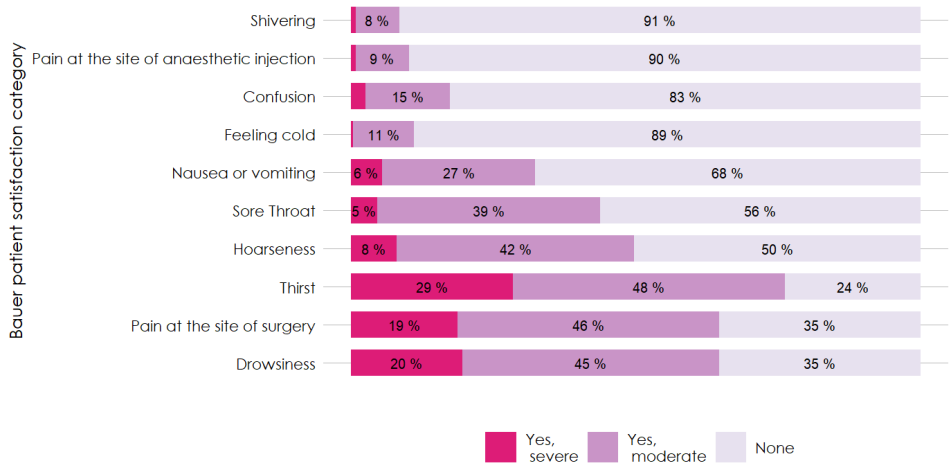
Figure 20a: Patient satisfaction with anaesthesia care (National data)



Bauer patient satisfaction survey - day 1

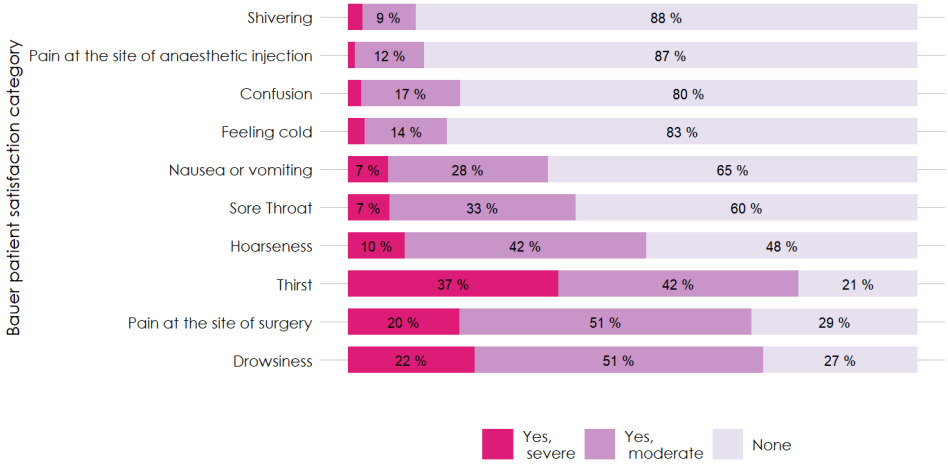
The Bauer patient satisfaction score assesses patients' experience of anaesthesia related discomfort within the first 24-hours following surgery, and their satisfaction with anaesthesia care. Survey results were available for 235 out of the 404 patients recruited to PQIP between 01 December 2017 and 27 July 2018 .

Figure 21: Patient experience of anaesthetic related discomfort (Bauer patient satisfaction score)



RCoA Royal College of Anaesthetists · NIAA HSRC Health Services Research Centre · The Health Foundation Inspiring Improvement · source surgical outcomes research centre

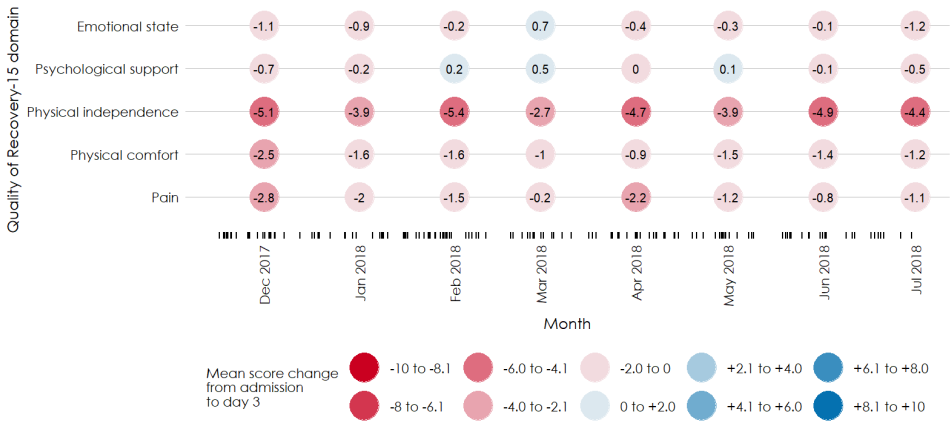
Figure 21a: Patient experience of anaesthetic related discomfort (National data)



Day 3 quality of recovery score

Quality of recovery on day 3 is assessed using the QOR-15. This is a shortened version of the QOR-40 which was first developed and validated in 2000 by Myles et al. The QOR-15 score was developed by Stark et al. and first published in 2013. It has been shown to be a valid, extensive, and yet efficient evaluation of postoperative quality of recovery.

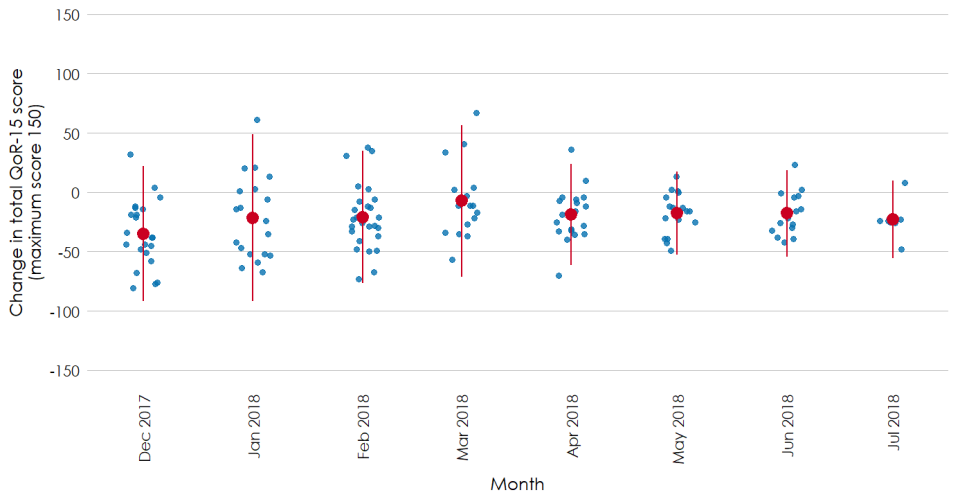
Figure 22: Mean QoR-15 domain score change by month (preop to day 3 postop)



Between 01 December 2017 and 27 July 2018 quality of recovery data was available for 143 patients. Figure 22 shows the mean monthly score by QOR-15 domain. The maximum score in each category is 10. Figure 23 below shows the change in QOR-15 score from admission to day 3 where data is available. The maximum total QOR-15 score is 150. Blue points represent the change for individual

patients, the red points are mean change for that month, with one standard deviation of the mean shown by the red lines extending from the mean. Outliers in a particular domain may draw attention to areas where quality of care could be improved.

Figure 23: QoR-15 score change from admission to day 3 by month



The data from the QOR-15 scores are presented for information. There are a wide range of factors which may influence the admission and day 3 scores that are not recorded as part of PQIP.

Health status at six months

The following information is based on all locked records atSample Hospital, not just those from the last 13 months. Measurement of generic health status is performed pre operatively and again at 6 & 12 months post operatively using EQ-5D-5L. This was developed in 2011 by Herdman et al. In each domain patients grade the severity from no problem to unable. Figure 24 shows change in severity in each domain at 6 months post operatively. Improvement is a decrease in severity by one or more levels. Deterioration is an increase in severity by one or more levels.

Figure 24: Change in level of severity at 6 months post operatively

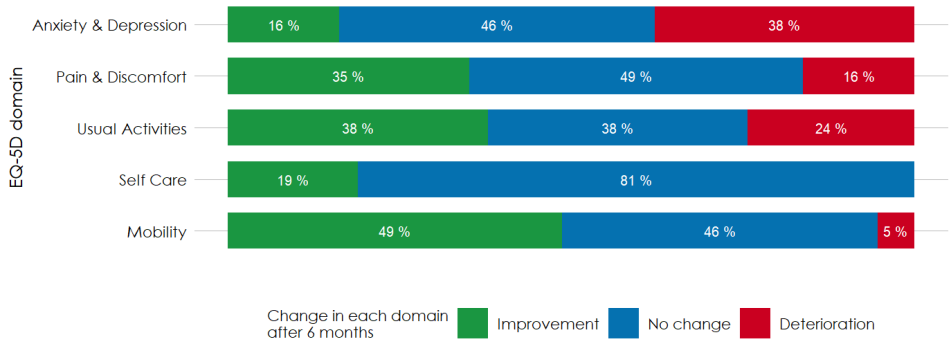
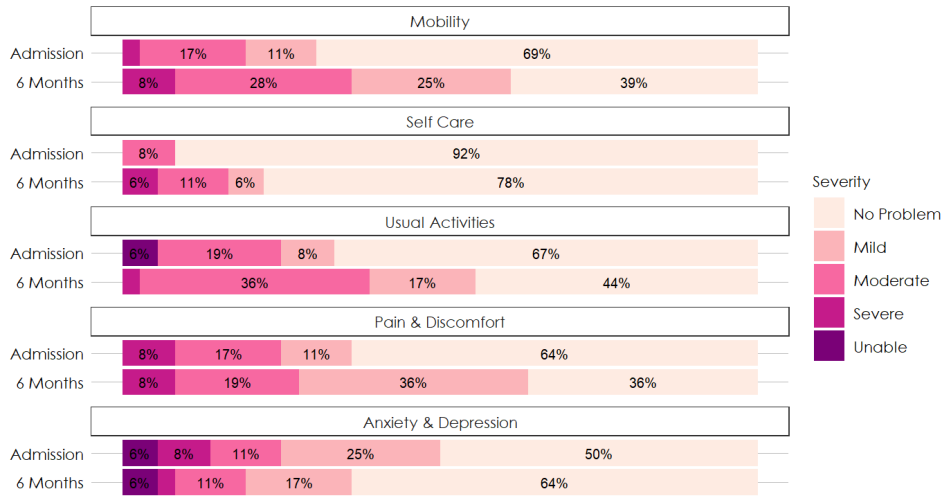


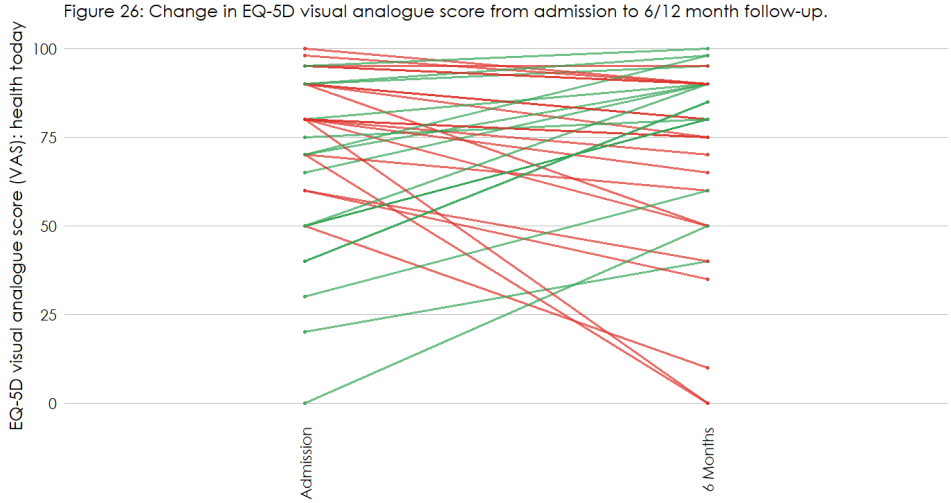
Figure 25 demonstartes change in proportion of patients in each severtity level by domain of EQ-5D at Sample Hospital

Figure 25: Change in proportion of patients in each severtity level by domain of EQ-5D.



Between 01 December 2017 and 27 July 2018 37 patients out of 52 had postoperative EQ-5D available at postoperative month 6. As part of EQ-5D each patient rates how good or bad their health is, on a visual analogue scale (0 - 100). In PQIP this is recorded preoperatively, six and twelve months post operatively. Figure 26 shows the change in each visual analogue score from admission to 12 months postoperatively. Each line represents a unique patient.

Figure 26: Change in EQ-5D visual analogue score from admission to 6/12 month follow-up.



Between 01 December 2017 and 27 July 2018 there were 37 patients out of 52 who had postoperative EQ-5D data available at month 6. Of the 0 patients who had their operation more than 1 year ago, 0 patients had completed their 12 month questionnaire.

Report feedback and contact details

The PQIP team want to make the way we present data as useful as possible for local quality improvement. We have created a short survey which gathers some information on your views on the PQIP dashboards and reports, and how you plan to use the data locally. The survey can be accessed here.

If you have any other queries or suggestions please contact us via options below:

- PQIP website: http://pqip.org.uk
- email: PQIP@rcoa.ac.uk
- Telephone: 020 7092 1678
- @PQIPnews

PQIP Perioperative Quality Improvement Programme

Appendix

Postoperative Morbidity Score mapping to Clavien-Dindo grades

POMS major is defined as those sub-domains that are assigned Clavien-Dindo grade 2 or above according to the table below. The mapping of POMS sub-domains to Clavien-Dindo graded complications was proposed by Wong et al.

Details of how the POMS organ-system sub-domains were mapped against Clavien-Dindo grades (reproduced with permission from Wong et al.) ## Postoperative Morbidity Score colour mapping

POMS organ system	POMS sub-domain	Assigned Clavien-Dindo grade
Pulmonary	New requirement for oxygen	2
Pulmonary	New requirement for respiratory support	2
Infectious	Currently on antibiotics	2
Infectious	Temperature >38Ã‚Â°C in the last 24hr	1
Renal	Urinary catheter in situ	1
Renal	Increased serum creatinine (>30% from preoperative level)	2
Renal	Presence of oliguria <500 mL/24hr	2
Gastrointestinal	Unable to tolerate an enteral diet for any reason	1
Gastrointestinal	Vomiting or abdominal distension, or use of antiemetics	1
Cardiovascular	Thrombotic event requiring anticoagulation (new)	2
Cardiovascular	Atrial or ventricular arrhythmias (new)	2
Cardiovascular	Hypotension (requiring pharmacological or fluid therapy >200 mL/hr)	2
Cardiovascular	New myocardial infarction or ischaemia	2
Cardiovascular	Cardiogenic pulmonary oedema	2
Neurological	New coma	3
Neurological	New confusion or delirium	2
Neurological	New focal neurological deficit	2
Haematological	Platelet, fresh-frozen plasma, or cryoprecipitate transfusion in last 24hrs	2
Haematological	Packed erythrocyte transfusion in the last 24hrs	2
Wound	Wound dehiscence requiring surgical exploration or drainage of pus from the operation wound with or without isolation of organisms	2
Pain	New pain significant enough to require parenteral opioids	1
Pain	New pain significant enough to require regional analgesia	2

The colour coding for the POMS displays (Postoperative morbidity survey at day 7, figures 14 and 15) was calculated using data from the first 772 patients recruited to the PQIP study with complete data.

The proportion of patients who were POMS positive at day 7 was calculated for each hospital. Summary statistics were then calculated. The proportions were not normally distributed, showing a right skew so classification of outliers was calculated using the interquartile range (IQR) and third quartile (Q3), with the formula:

$$UpperRegionOutlier = Q3 + (1.5 * IQR)$$

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

PQIP Perioperative Quality Improvement Programme

R Markdown

This PQIP report was produced using R Markdown. Markdown is a simple formatting syntax for authoring HTML, PDF, and Powerpoint documents. For more details on using R Markdown see http://rmarkdown.rstudio.com. In order to produce the 400 reports that we have released this quarter we have automated production. This means there is the potential for some small formatting errors to have occurred without being picked up during the production process. If you do notice anything that you think may be an error please contact is via the routes above.

RCoA Royal College of Anaesthetists NIAA HSRC Health Services Research Centre The Health Foundation Inspiring Improvement source surgical outcomes research centre

Appendix F

Appendix F-1 Framework matrix of pomVLAD theory of change

	A : Building engagement (central to local level)	B : Identification of QI opportunities through pomVLAD dashboard	C : Monitoring of QI implementation	D : Perceptions of pomVLAD project	E : Perceptions of process measure recommendations	F : QI activity related to pomVLAD dashboard	G : Understanding of data quality and data completion
1 : Participant #11		Clear QI opportunities for surgical teams		Tailormade for surgeons to do QI work Opportunity to engage surgeons			Process measure performance (poor) stimulated discussion
2 : Participant #12	Setup phone call stimulated engagement VLAD may need specific explanation	Time-lag from implementation to starting QI work Identification of QI opportunities may be supported by addtion of control limits/reference points	Current utility limited by lack of QI work	Morbidity a suitable outcome given low mortality incidence	Suitable measures Definition of eating may need revision - outcome vs. process DrEaMing excellent measure		

	A : Building engagement (central to local level)	B : Identification of QI opportunities through pomVLAD dashboard	C : Monitoring of QI implementation	D : Perceptions of pomVLAD project	E : Perceptions of process measure recommendations	F : QI activity related to pomVLAD dashboard	G : Understanding of data quality and data completion
3 : Participant #14	Inclusion in pomVLAD project promoted engagement Interpretation of VLAD prompted engagement with central team			Potential need for 'bedding in' period prior to ensure understanding Need for accurate data New style of data presentation (VLAD) stimulated interest May support surgical engagement VLAD dashboard simplifies complex morbidity data Range of morbidity/complication definitions can be confusing			Focus on accuracy of morbidity data Increased understanding of morbidity data Varying morbidity/complication measures can be confusing Confidence intervals or control limits may help interpretation
4 : Participant #2	PQIP national event stimulated engagement		Local data collection around local QI work - seperate to PQIP/pomVLAD				PQIP measures - apparent poor performance identified as recording error
5 : Participant #3	Targeting senior management may support engagement						PQIP measure - pain, poor performance stimulated more detailed assessment of data
6 : Participant #6	PQIP national event stimulated engagement	Identification of performance and improvement work identified	Useful to monitor QI work but limited improvement activity happening	pomVLAD may support early identification of problems			Apparent poor performance investigated and felt to be documentation issue Poor performance stimulated further investigation into data and processes

	A : Building engagement (central to local level)	B : Identification of QI opportunities through pomVLAD dashboard	C : Monitoring of QI implementation	D : Perceptions of pomVLAD project	E : Perceptions of process measure recommendations	F : QI activity related to pomVLAD dashboard	G : Understanding of data quality and data completion
7 : Participant #7	PQIP 'experts' presenting to local sites incl. management	pomVLAD dashboard stimulated deeper investigation of data Supported ongoing local QI work (intraop warming)	Useful to monitor improvement work	Real-time feedback supports early intervention Process measures not applicable to all specialties Timescale of morbidity measure may not be suitable for all surgery types	ERAS package vs. importance of individual elements May not be suitable for all specialties	Intraoperative warming	Apparent poor performance stimulated investigation into data Morbidity data has prompted local work outside pomVLAD/PQIP project
8 : Participant #8							
9 : Participant #1	Site initiation calls supported initial engagement	Focus on DrEaMing QI work	Dashboard useful to monitor ongoing QI work DrEaMing QI work	Suitability of risk-adjustment model important Morbidity appropriate outcome measure	Local context can impact on delivery of processes Process recommendations can focus delivery of care DrEaMing key focus	DrEaMing	Apparent poor performance led to investigation of data and planned improvement VLAD will improve understanding of quality of care and feed into M&M meetings
10 : Participant #15							PQIP data adn apparent poor performance (ERAS) has led to discussion around data quality Good quality data is needed - and dedication to deliver it
11 : Participant #10	Awareness, but limited involvement in project						Improved documentation implemented after apparent poor performance identified
12 : Participant #13							

	A : Building engagement (central to local level)	B : Identification of QI opportunities through pomVLAD dashboard	C : Monitoring of QI implementation	D : Perceptions of pomVLAD project	E : Perceptions of process measure recommendations	F : QI activity related to pomVLAD dashboard	G : Understanding of data quality and data completion
13 : Participant #4	Selection into pomVLAD project positive and good opportunity Setup phone call increased engagement Increasing management engagement might help clinical teams	Focus on DrEaMing - data quality Identification of QI opportunities based on dashboard dials	Up to date nature of dashboard better for monitoring changes in care	Presents opportunity not to be squandered VLAD easy to interpret Real time data useful			Apparent poor performance has led to local discussion around documentation and interpretation of data points Apparent poor performance led to local investigation of data (process measures)
14 : Participant #5	Feedback to patients about study outcomes important	Easy identification of areas for improvement Interest in understanding data quality Improvement work on carbohydrate loading		Real value in project is when data is acted upon		Preop carbohydrate loading	Apparent poor performance led to local discussion about data collection and interpretation
15 : Participant #9	Engagement not viewed as relevant to role						

Appendix F-2 Framework matrix of strategies to establish and maintain engagement

	A : Triggers to engage	B : Strategies to maintain local engagement across specialties
1 : Participant #11		Regular meetings to review data and progress Departmental audit meetings - incl. cross specialty
2 : Participant #12	Data entry as trigger to look at dashboard Report release acts as trigger POM ward rounds link data to clinical work Regular PQIP meetings to review data/performance	Developing PQIP 'team' locally Involvement in inputting data and final checks Circulation of reports Informal communication with MDT Sharing of anonymised individual clinician process/outcomes Local area networks Presentation at clinical effectiveness meetings
3 : Participant #14	Report release acts as trigger Departmental meetings between report cycles trigger dashboard engagement	Integration with perioperative medicine ward rounds Presentation at departmental meetings Involvement in data entry/checking Circulation of reports
4 : Participant #2	Combined clinical governance meeting National annual report highlighted some poor performance Monthly review meetings with research nurses to ensure data quality Lack of alarming performance reduces need to engage	Presentation at clinical governance meetings Trying to develop shared targets for improvement Circulation of reports by email Use of PQIP data at local QI poster competition
5 : Participant #3	Management team meeting Local QI work (pain/anaemia) stimulates engagement with data Departmental audit meetings Report release PQIP newsletter stimulates local engagement	Presentation at regular local meetings Involvement of trainees Circulation of reports incl. summary Feedback of positive performance

	A : Triggers to engage	B : Strategies to maintain local engagement across specialties
6 : Participant #6	Annual report resulted in local work presenting data QI work based on report data stimulates engagement GIRFT visit Reports and contact from PQIP team trigger comparison to local data	Circulation of reports by email Presentation at departmental audit meetings
7 : Participant #7		Build trainee involvement
8 : Participant #8	Time allocated in job plan Departmental audit meeting presentation Local QI activity Identification of poor performance	Presentation at departmental meetings Individual presentation to surgical colleagues Trainee involvement Use of PQIP data towards appraisal Notice boards displaying information Circulation of reports by email
9 : Participant #1		Presentation of results at operational governance meetings Informal communication
10 : Participant #15	Annual report circulated Regular audit meeting presentation may support engagement	Presentation at audit meetings Development of local networks
11 : Participant #10	Report release Regular PQIP meeting to review data and any issues highlighted	Presentations to department Circulation of reports Build trainee engagement Regular PQIP 'team' meetings to review data and results
12 : Participant #13	Medical students joining for research unit Annual report and quarterly report release Local QI work (anaemia pathway)	

	A : Triggers to engage	B : Strategies to maintain local engagement across specialties
13 : Participant #4	Report release MDT communication based on report release and QI work	Presentation at departmental meetings Build trainee engagement Feedback of positive results
14 : Participant #5	Report release or data input GIRFT visit Significant change in performance	Presentation at departmental meetings incl. cross specialty Informal communication Trainee engagement QI work based on PQIP data
15 : Participant #9		Involvement in study recruitment

Appendix F-3 Framework matrix of dashboard feedback

	A : Experience of using the dashboard	B : Participants perceptions of dashboard vs. report feedback	C : Recommendations for improving the dashboard	D : Statistical considerations
Participant #1		Waiting for reports to engage Acknowedges not using dashboards effectively Dashboards more useful for monitoring planned improvement work Dashboards useful for frequent monitoring - monthly or more frequently Responsibility on clinicians to use data effectively		Concern re: adequate risk adjustment Model performance compared to existing well known/used models
Participant #10		Release of reports leads to engagement with data Reads reports for own interest Reports well laid out, easy to read Lack of time prevents reading reports Job role not QI therefore doesn't feel need to read reports		
Participant #11	Impressed with dashboard Sees potential to engage surgical team in QI VLAD interpretation straightforward	Lack of time limits engagement with dashboards Reports easily accessible and easy to disseminate Lack of QI interventions limits utility of dashboard and reduces frequency of access Dashboards are user friendly way to interogate data a bit more - data query system a problem Dashboard may support surgical engagement	Currently difficult to access data to perform own analyses Being able to seperate results by specialty increases relevance	VLAD simple to interpret and use

	A : Experience of using the dashboard	B : Participants perceptions of dashboard vs. report feedback	C : Recommendations for improving the dashboard	D : Statistical considerations
Participant #12	Useful for quick overview of data Looks at dashboard when inputting data Morbidity/complication outcome more useful than mortality	Dashboards more useful for frequent access Reports a trigger to engage Dashboard utility linked to monitoring QI interventions Understanding which patients are included in report is difficult Reports quick/convenient to look at	Include N as well as % Allow comparison with local sites - driven by sites not central Consider adding length of stay dashboard	Notification of dip in outcomes might be worth highlighting to trusts Trialling control limits may be helpful
Participant #13	Lack of dashboard use - report provides relevant information	Only uses report - includes everything feels needed		VLAD easy to interpret Percentages can vary significantly because of small numbers of patients recruited
Participant #14	Uses on adhoc basis pomVLAD data helps to simplify morbidity/complication data	Dashboard access useful for frequent and quick review Reports evolving, covering varying timescales No real requirement to review data very frequently Complexity of data in reports can be overwhelming National agenda not neccesarily local agenda Monthly data inadequate to monitor QI work Large swings in % due to low number of patients recruited	Customisability - national agenda not necessarily key local issues Range of morbidities reported (POMS vs. Clavien) can be confusing Improve information about risk-model and how to interpret VLAD Tools to support interpretation - what is a significant trend? Add comparison to other local sites or confidence intervals	Recruitment affects speed of changes observed Desire to understand risk-adjustment and ensure morbidity data accurate Confusing range of morbidity/complication measures in programme
Participant #15		Real time data has potential for greater QI impact that intermittent reports	Data should be relevant to important bodies (CQC/RCS)	
Participant #2	Useful to have live, up to date reporting via dashboards Dashboard useful to monitor QI work Navigation straight forward	Dashboards more useful for continuous feedback of performance Dashboards useful to monitor QI work	Support identification of high risk patients Improve ability to identify patients within dashboard - to support follow up	

	A : Experience of using the dashboard	B : Participants perceptions of dashboard vs. report feedback	C : Recommendations for improving the dashboard	D : Statistical considerations
Participant #3	Limited experience of dashboards - tends to use reports	Reports provide premade resource Dashboards require time input for initial navigation/understanding Dashboards may be more useful than reports for monitoring QI work Reports provide overview before more detailed analysis of data	Moving averages may help interpretation Issues with small numbers causing large changes in %	
Participant #4	Easy to understand after initially appearing overwhelming Straightfoward to interpret process measures Limited use after initial access	Lack of time reduces frequency of accessing dashboard Dashboard more up to date Dashboard offers opportunity to monitor change faster than reports	Provide national or local comparisons Include breakdown of constituent DrEaMing parts	Short period of time to gain understanding of VLAD How do process measures link to morbidity?
Participant #5	Easy to visualise areas for improvement Infrequent access to dashboard Unsure of colleagues use of dashboard	Report format useful for quick overview of data Both report and dashboard format useful to support understanding		

	A : Experience of using the dashboard	B : Participants perceptions of dashboard vs. report feedback	C : Recommendations for improving the dashboard	D : Statistical considerations
Participant #6	Limted access to dashboard Limited engagement within department and time to act on results reduces utility of dashboard	Dashboards provide opportunity to act early on concerning issues Tendency to rely on reports Acknowledges room to improve own use of dashboards Familiarity with reports eases understanding. Dashboards less familiar Reports easy to read Dashboards require some time input to get used to Lack of QI activity reduces accessing of data Dashboard and reports provide useful information for external bodies Limited engagement within department and time to act on results reduces utility of dashboard		
Participant #7	Lack of previous dashboard use might reduce current utility Dashboards useful to a degree - limited by local relevance of measures	Reports offer more detailed information Previous lack of dashboard use reduces utility Dashboard (real time) makes it easier to track change in performance	Customisability to focus on measures important to local site	
Participant #8	Limited use of dashboards Impressed when has accessed	Report provides overview in once place - easily accessible, broad overview Dashboard may be more useful if QI work is being done		
Participant #9	Accessible, easy to access			

Appendix F-4 Framework matrix of participants perceptions of their role in QI

	A : Participants perceptions of their role in quality improvement
1 : Participant #11	Need for baseline data prior to starting QI work Build enthusiasm with team to stimulate them to do QI Current QI is anaesthesia related - in area participant interested in QI work not sole responsibility - needs to engage surgical colleagues
2 : Participant #12	Frequent involvement in QI Involvement at national level in perioperative medicine field Feels need to demonstrate impact priot to being given time/financial support Non-clinical work greater than time allocated within job plan Improvement requires MDT input Difficult to engage surgical colleagues in QI work related to PQIP
3 : Participant #14	QI work dependent on clinicians to deliver Lack of support structure in place to facilitate QI work Difficulty engaging colleagues in QI Need for baseline data prior to implementing change Siloed working - difficulty engaging surgical colleagues
4 : Participant #2	Acknowledgement of MDT nature - developing combined goals for QI Responsibility for leading QI on participant Research nurse role to flag up potential issues - externalises Identification of varying engagement in department

	A : Participants perceptions of their role in quality improvement
5 : Participant #3	Involvment in previous successful improvement work Need for management to be engaged in QI work Data analysis should not be responsibility of clinicians Clinician responsibility to drive change Need for MDT involvement in QI Coordinating role - overview of pathway important Needs support in delivering QI but own priorities may not match trust ones
6 : Participant #6	Externalises QI work to other parts of MDT Requests external support in change management Acknowledges lots to improve Sees need to engage colleagues but feels unable to
7 : Participant #7	Feels need to engage people in research and project
8 : Participant #8	Needs support from colleagues to deliver QI Feels responsible for not pushing project harder with colleagues Lack of time to take ownership of all QI work related to study
9 : Participant #1	Overview of data and ensure QI work happening - but not direct involvement in QI Responsibility on others to lead QI work
10 : Participant #15	Limited involvement in project and QI work Responsibility of others with QI experience to deliver improvement Improvement focussed on areas of poor performance QI work not structured - needs robust structures in place to be consistent Limited time to fully engage in QI work

	A : Participants perceptions of their role in quality improvement
11 : Participant #10	Role to collect data - not deliver QI QI clinician responsibility but limited by time QI needs engaged team Interest in delivering QI but not possible currently due to time/job role
12 : Participant #13	Role to run study, collect data QI responsibility of clinicians and nurses involved in direct patient care Wants to see improvement from work Leading medical students in QI projects Interest in improvement work but lack of time
13 : Participant #4	Informal involvement in data feedback Understaffed and under supported so unable to deliver any QI work Clinicians responsible for delivering QI Role is to ensure running of study, patient recruitment etc.
14 : Participant #5	Limited time and resource QI responsibility of clinicians Role to collect data and run study - not QI Feels need for data to be acted on
15 : Participant #9	Role is data collection not QI Feels QI is important QI is clinicians responsibility

Appendix F-5 Framework matrix of barriers and facilitators of QI

	A : Barriers to local QI activity	B : Facilitators to local QI activity
1 : Participant #11	Perceived need for baseline data Lack of time to undertake QI work Difficulty engaging colleagues Difficulty accessing and analysing data outside that provided by PQIP team	Research team and trainee involvement to support data collection Regular (two-monthly) PQIP meetings to review data and get together as PQIP team generates enthusiasm Push from RCoA to encourage engagement - especially trainees
2 : Participant #12	Existing performance monitoring dashboards not clinically relevant Lack of support from management for QI work No formal time allocated to QI work in job plans Difficulty engaging surgical colleagues	Communication from central team to senior management New information considered 'interesting' from clinical perspective Sharing anonymised clinician level process/outcomes Sharing 'good news' stories from external sites of where QI work has been implemented Informal local networks can drive improvement Comparison with other similar sites/control limits/confidence intervals may support improvement of poor performance
3 : Participant #14	Existing dashboards not clinically relevant or widely used Lack of formal time recognition in job plan Lack of support from management Lack of QI structures in hospitals Local agenda not fitting national improvement agenda Difficulty interpreting local data in context of small numbers recruited Other national projects compete for importance	Feeling of QI being important within the trust - being supported Experience within department of QI methodology Time allocation to support QI work Engagement of trainees Research nurse allocation to support data collection

	A : Barriers to local QI activity	B : Facilitators to local QI activity
4 : Participant #2	Difficulty engaging surgical colleagues Clinical targets for improvement do not match organisational targets QI work dependant on clinicians Lack of designated time to focus on non-clinical/QI work Lack of drive to improve unless results are alarming Poor managerial support unless financial incentives attached	Electronic health records should improve data collection and usability of data Provision of dedicated 'QI days' within trust raises profile or QI work Feeling supported by divisional leads and colleagues Allocation of time to support PQIP and QI activity Comparison of local performance with national Awareness and ability to take QI work from external sites into local practice Research nurse involvement helps keep study running on day to day basis Push of perioperative medicine agenda by Royal Colleges Financial incentives such as best practice tariffs
5 : Participant #3	Existing hospital dashboards focussed on managerial level outcomes Lack of specified time to implement and sustain projects Lack of analyst support to interrogate data Trainee rotations limit input they can have into projects Difficulty engaging colleagues Management engagement needed to deliver QI work Organisational targets do not match clinician targets for improvement	Identification of QI lead within department Previous experience of QI work Time allocated for research/QI work Push from Royal Colleges - particularly RCS could improve engagement Reseach nurse involvement to maintain recruitment and data collection Relevance of data to clinical practice Trainee engagement with development of specific projects Developing group of colleagues to develop QI work Positive engagement from senior management to help deliver QI work Making link from just having data to real patient outcomes Understanding patient pathway and key stakeholders for each part - building support

	A : Barriers to local QI activity	B : Facilitators to local QI activity
6 : Participant #6	Difficulty transforming discussion into action Lack of specified time for project work Difficulty engaging colleagues Lack of management support for clinical improvement aims Limited trainee interest in research	Comparison with similar external sites and shared learning from experience of implementation Support and engagement from colleagues Engagement from management and feeling of importance given to project by them External bodies visiting (GIRFT)
7 : Participant #7	Lack of existing clinically relevant dashboards Lack of designated time Perceived inability to deliver processes Local relevance of measures reported Inability to access existing data sources to use for QI Lack of managerial support for non-clinical activities	Time allocated to QI work Engagement from central project team to local site to raise profile QI work Research nurse involvement to keep study running
8 : Participant #8	Lack of support for delivering QI Difficulty engaging surgical colleagues Good or satisfactory performance inhibits desire to improve Lack of designated time for QI	Research culture within hospital Time allocation for PQIP/QI work Push from Royal Colleges to promote study and QI work Trainee engagement Examples from external trusts about QI work that has worked/why trusts are positive deviants
9 : Participant #1	Local context - staff turnover Lack of cross-specialty improvement culture Recording data in usable format not an organisational priority Organisational data capture priorities differ from clinical priorities	Support from management/senior management in trust Importance given to project on national scale Providing evidence for links between process and outcome

	A : Barriers to local QI activity	B : Facilitators to local QI activity
10 : Participant #15	Inadequate resources for QI Limited surgical engagement in project Satisfactory performance or above average performance removes drive to improve Lack of formal QI structures and skills in place to deliver QI Need for good quality data and recruitment Lack of time for QI work - whose responsibility should it be?	Perception of relevance of data feedback to local performance Knowledge and capability within organisations to deliver change projects Regional QI networks can facilitate knowledge exchange Contemporaneous data feedback supports QI activity
11 : Participant #10	Poor engagement from senior colleagues/management Lack of time to deliver QI work Local context - staff changes affecting recruitment to study Difficulty accessing data for local projects not included in dashboards/reports Lack of surgical engagement	Examples of successful QI work carried out elsewhere Funding and support for QI activity Meetings to review data collection and quality
12 : Participant #13	Turnover of staff affects day to day running of project Lack of time to undertake QI work - not perceived to be part of role Small numbers recruited can impact utility of data	Time available and dedicated for QI Case studies of successful QI work
13 : Participant #4	Poor trust level support of research Lack of formal structures to feedback data to ward staff Lack of time and not part of role deliver QI Low banding of study from NIHR Positive results reduce desire to implement change	Areas of poor performance identified in reports National or local comparisons of performance Informal communication from interested individuals Stable team supports running of project Engagement of senior management by central PQIP team may support local teams
14 : Participant #5	Lack of support and funding for research activity from senior management Local staffing affecting recruitment Lack of time to review data NIHR banding of study - band 1 Poor engagement from management	Trainee engagement with specific project allocation Informal communication and sharing of data between individuals

	A : Barriers to local QI activity	B : Facilitators to local QI activity
15 : Participant #9	Limited cross MDT working QI not perceived as part of role	Feeling of support and drive within the institution to deliver QI