

SIQINS: **STRENGTHENING **Q**UALITY IMPROVEMENT **I**NTervention REPORTING IN **S**URGERY**

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

By Emma Leanne Jones, BSc (Hons), MSc, MCSP HCPC registered

Department of Health Sciences

March 2017

ABSTRACT

SIQINS - **S**trengthen**I**ng **Q**uality Improvement **I**ntervention reporting in **S**urgery

By Emma Jones

Background

Surgical quality improvement (QI) research has been conducted with promising results, but translating learning into practice is complicated by incomplete reporting. This research aimed to identify which reporting items are most frequently incomplete, and why incomplete reporting occurs.

Methods

A systematic review aimed to identify the current standard of reporting in the surgical QI literature. MEDLINE, Scopus and the Cochrane Central Register of Controlled Trials (CENTRAL) were searched. Articles were scored against a modified Template for Intervention Description and Replication (TIDieR) checklist.

A qualitative interview study was conducted. QI authors, consumers (those who apply QI research in practice), editors and reporting guideline authors were interviewed to identify why reporting is hard. An author checking exercise involved asking the interview participants to describe how they would replicate an intervention described in a QI article. The article's authors checked whether their intervention matched the participants' interpretation. Data was analysed using the constant comparative method.

Results

The systematic review identified 100 relevant articles. Reporting of fidelity was absent in 74% of articles; and modifications were absent in 73%.

Participants (42) in the qualitative study included: 15 authors, 12 consumers, 11 journal editors and four reporting guideline authors. Of these, 28 were clinicians. Reporting QI is hard because: QI is an emerging field; features of hospitals and journals create tensions for reporting; context is hard to describe; publications are not always intended to be used for exact replication.

Discussion and conclusion

QI is a youthful field but stakeholders have well developed aspirations for QI reporting. As the field matures, those involved in QI reporting place value on describing context, identifying active ingredients, and recognising how QI publications are to be used in practice. This research is relevant to an international audience and could help galvanise a renewed sense of importance for reporting.

ACKNOWLEDGEMENTS

“We can all profit from sharing experience. For everything great that is done, there is a process. Lessons must be learned.” (Gill Penny, SIQINS patient advisor)

This thesis has come from my own intellect, effort, emotion and personality traits, but it would not have taken the shape it has today without help from so many others.

I am inspired by all those who participated in this study. They work tirelessly, passionately, and with so much attention and care towards ensuring that when each person is admitted to hospital, for what can often be a frightening and uncertain experience, they are kept safe and feel well looked after. Evaluating performance, celebrating success, and continuing to improve upon it can be academic and scientific, yet it is also emotional, and can feel personal. I am grateful that 42 people, who are committed to face the challenges of QI, were able to share their experiences and opinions with me.

I thank my supervisors Professor Mary Dixon-Woods and Professor Graham Martin, and The Health Foundation who funded this study for their commitment to supporting me. I thank my colleagues and advisory panel who inspired me and nudged me towards the right path. My local collaborators at each recruitment site helped enormously with encouraging recruitment for this study and progress would have been much slower without them. I thank my family, my friends from far away, and my friends at home in Leicester: at St James the Greater, at the University Hospitals of Leicester NHS Trust and at the University of Leicester. Finally, I thank John for all his gifts, but especially of belief and rest.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
INTRODUCTION	1
1.1 The emergence of quality improvement in healthcare	3
1.2 The case for QI in healthcare	7
1.3 QI, research and QI research	9
1.4 Why is perioperative care an area in which the challenges of QI reporting should be explored?	11
1.5 What is reporting?	12
2 BACKGROUND TO THE CHALLENGES OF QI REPORTING	15
2.1 The problem of reporting in QI: a review of the literature	15
2.2 What reporting tools exist to improve reporting in QI?	17
2.3 Why is reporting problematic?	22
2.3.1 Challenges of reporting guidelines	22
2.3.2 Challenges of the writing and publication process	24
2.4 Why is reporting QI likely to be particularly problematic?	26
2.4.1 The evolution of QI as a scientific field – type of reporting	26
2.4.2 The evolution of QI as a scientific field – depth of reporting	28
2.4.3 Understanding how to report QI in perioperative care	29
2.5 What are the consequences of poor reporting?	29
2.5.1 Threats to the conduct of good systematic review	30
2.5.2 Threats to successful replication	32

2.5.3	Previous examples of replication in QI	33
2.5.4	Research waste.....	35
2.6	Researchable questions, yet to be answered	37
3	METHODS	40
3.1	Introduction to the methods chapter.....	40
3.2	Personal motives	40
3.3	Systematic review	42
3.3.1	Study selection: Overcoming conceptual and terminological confusion.....	43
3.3.2	Study selection: Search strategy	50
3.3.3	Study selection: Eligibility criteria.....	55
3.3.4	Data extraction	56
3.3.5	Data analysis.....	61
3.4	Qualitative work	61
3.4.1	Qualitative study objectives	64
3.4.2	Advisory panel.....	64
3.4.3	Telephone interviews.....	66
3.4.4	Inclusion criteria.....	66
3.4.5	Definition of recruitment groups.....	67
3.4.6	Participant identification, eligibility assessment and recruitment ..	70
3.4.7	Number of participants	72
3.4.8	Participant interview schedule	73
3.4.9	Author-checking exercise	75
3.4.10	Data analysis.....	76
3.5	Qualitative work: Study operating procedures.....	81
3.5.1	Informed consent.....	81
3.5.2	Confidentiality and data handling	82

3.5.3	Discontinuation/withdrawal of participants	83
3.5.4	Safety monitoring.....	83
3.5.5	Quality control, quality assurance procedures and approvals	83
3.6	Qualitative work: Reflective practice	85
4	SYSTEMATIC REVIEW FINDINGS.....	88
4.1	Article identification	88
4.2	Study characteristics	89
4.3	Completeness of reporting: Quality interventions and QI techniques..	93
4.3.1	Completeness of reporting: Quality interventions (TIDieR).....	93
4.3.2	Completeness of reporting: QI technique.	99
4.3.3	Additional study features	100
4.4	Discussion: Systematic review	101
4.4.1	Limitations of the systematic review	101
4.4.2	Explanation of the review findings	107
5	DEFINING QI AND FINDING A SCIENTIFIC FRAME FOR REPORTING.....	111
5.1	Introduction to the findings chapters for the qualitative work.....	111
5.2	Sample size	112
5.3	Participant characteristics	113
5.4	Defining QI and finding a scientific frame for reporting.....	115
5.5	The vocabulary of QI.....	115
5.5.1	What is quality improvement?	116
5.5.2	The terminology of QI	121
5.6	QI reports and QI research	125
5.6.1	Type of journal.....	127
5.6.2	The use of narrative data in QI	127
5.6.3	Generalisability and using QI reports in a new setting.....	128

5.6.4	Bias in QI	130
5.6.5	Ethical approval	132
5.7	Reporting QI interventions	132
5.7.1	Complexity	133
5.7.2	Rationale	134
5.7.3	Materials and procedures	134
5.7.4	Fidelity and modifications	136
5.7.5	Outcome measurement	138
5.7.6	Active ingredients	139
5.8	Discussion: Defining QI and finding a scientific frame for reporting ..	142
6	INFLUENCES ON COMPLETENESS OF QI REPORTING	145
6.1	Motives for good reporting	145
6.1.1	Academic/career reward	145
6.1.2	Patient benefit	149
6.1.3	Reduction of waste	152
6.2	Influences on reporting associated with organisations	153
6.2.1	Influences on reporting associated with hospitals	153
6.3	Influences on reporting associated with journals	158
6.3.1	Rules and regulations	158
6.3.2	Perceptions and expectations	163
6.3.3	Locating QI literature	168
6.4	Discussion: Influences on reporting	171
7	THE PROBLEM OF HOW TO REPORT CONTEXT	174
7.1	Introduction	174
7.2	Defining context and understanding what to report	174
7.3	Noticing contextual features	178
7.4	Reluctance to write candidly about context	179

7.5	A failure to value qualitative approaches.....	180
7.6	Encouraging scientific writing about context	183
7.6.1	Generalisability, replication and scaling up.....	184
7.6.2	Establishing causality	188
7.7	Discussion: The problem of how to report context	191
8	QI AUTHORS AND THEIR AUDIENCES.....	194
8.1	Introduction	194
8.2	Does an audience of a QI article understand the authors' description of an intervention?	203
8.3	Authors and their audience – disagreements arising from the author checking exercise	207
8.3.1	A blueprint for replication?	207
8.3.2	The purpose and type of QI publication.....	208
8.4	Participants' views of how reporting can be improved	209
8.5	Discussion: Authors and their audience	213
9	DISCUSSION.....	216
9.1	Personal reflection	216
9.2	Youth, aspiration and maturity	217
9.2.1	The youthful field of QI	218
9.2.2	The aspiring field of QI	226
9.2.3	The maturing field of QI	232
9.2.4	Conclusion – Youth, aspiration and maturity	244
9.3	Peculiarities unique to the field of perioperative QI	245
9.4	Limitations of the qualitative study	247
9.5	Recommendations for practice	248
9.6	Recommendations for further research.....	253
	APPENDICES.....	255

Appendix 1: Chapter 2: Poor reporting in QI - a list of 11 systematic reviews	255
Appendix 2: Standard letter of invitation	257
Appendix 3: Participant information sheet	258
Appendix 4: Participant interview schedule	261
Appendix 5: The highlighted and anonymised Lean article used during participant interviews	267
Appendix 6: Lean authors participant information sheet: Author checking exercise	268
Appendix 7: Lean authors' interview schedule.....	271
Appendix 8: Consent form	279
Appendix 9: The references of 100 perioperative QI papers included in the systematic review *	280
Appendix 10: Milestones of SIQINS: Qualitative work	281
Appendix 11: Publications arising from this PhD	283
REFERENCES.....	286

LIST OF TABLES

Table 1 Intervention reporting tools.....	19
Table 2 Types of poor reporting	39
Table 3 Quality Improvement Taxonomy†	47
Table 4 Search Strategy	52
Table 5 Data extraction template items.....	58
Table 6 Advisory panel members.....	65
Table 7 Study design and country of publication frequency for 100 perioperative QI papers	90
Table 8 100 perioperative QI publications categorised according to targeted clinical issue	92
Table 9 Example of fully complete TIDieR items.....	96
Table 10 Participants by professional category and recruitment group.....	114
Table 11 Patient and public involvement in QI	166
Table 12 Excerpts from interview transcripts	196
Table 13 How reporting in QI could be improved	200
Table 14 Recommendations for practice.....	250

LIST OF FIGURES

Figure 1 PRISMA Flow Diagram	89
Figure 2 Range of TIDieR scores 0-11.....	95
Figure 3 TIDieR scores for perioperative Quality Improvement papers.....	95
Figure 4 QI technique scores for 100 perioperative QI papers.....	99
Figure 5 From McCulloch et al: Table of interventions (p.1044).....	202
Figure 6 Types of QI article	236

LIST OF ABBREVIATIONS

BMJ	British Medical Journal
CENTRAL	Cochrane Central Register of Controlled Trials
CLAHRC	Collaborations for Leadership in Applied Health Research and Care
CONSORT	Consolidated Standards of Reporting Trials
CTIMP	Clinical Trial of an Investigational Medicinal Product
CQI	Continuous quality improvement
CQM	Continuous quality management
CVC-BSI	Blood Stream Infections from Central Venous Catheters
DH	Department of Health
EPOC	Cochrane Effective Practice and Organisation of Care
ERAS	Enhanced Recovery After Surgery
EQUATOR	Enhancing the QUALity and Transparency Of health Research
EU	European Union
GCP	Good Clinical Practice
GPM	Graham Paul Martin
HF	Human Factors
HRA	Health Research Authority
IHI	Institute for Healthcare Improvement
IMRaD	introduction, methods, results and discussion
IP	intellectual property
IS	Implementation Science

ISI	International Secondary Information service
IT	Information technology
ITAX	Intervention Taxonomy
JARS	Journal Article Reporting Standards
LIIPs	Leicestershire Improvement, Innovation and Patient Safety Unit
LOS	Length of stay
LSE	London School of Economics and Political Science
MeSH	Medical Subject Headings
MDW	Mary Dixon-Woods
MEWS	Modified Early Warning System
MUSIQ	Model for Understanding Success in Quality
NIHR	National Institute for Health Research
NHS	National Health Service
NJL	Nicholas James Lees
NSQIP	National Surgical Quality Improvement Program
PDCA	Plan-Do-Check-Act
PDSA	Plan-Do-Study-Act
PICO	population, intervention, comparison and outcome
PPI	Patient and public involvement
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PROM	Patient Reported Outcome Measure
QI	Quality Improvement

QI-MQCS	QI Minimum Quality Criteria Set
RCS	Royal College of Surgeons
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
REF	Research Excellence Framework
REWARD	Reduce Research Waste and Reward Diligence
RIAT	Restoring Invisible and Abandoned Trials
R&D	Research and Development
SAPPHIRE	Social Science Applied to Healthcare Improvement Research
SI	Statutory Instrument
SIQINS	Strengthening quality improvement intervention reporting in surgery
SORT	Standards of Reporting Trials
SPC	Statistical Process Control
SQUIRE	Standards for Quality Improvement Reporting Excellence
STARD	Standards for the Reporting of Diagnostic accuracy
StaRI	Standards for Reporting Implementation studies of complex interventions
STRICTA	Standards for Reporting Interventions in Controlled Trials of Acupuncture
STROBE	Strengthening Reporting of Observational Studies in Epidemiology
TIDieR	Template for Intervention Description and Replication

TPS	Toyota Production System
TQM	Total Quality Management
UK	United Kingdom
UoL	University of Leicester
US	United States
VTE	Venous Thromboembolism
WHO	World Health Organization
WMA	World Medical Association

INTRODUCTION

Quality improvement (QI) in healthcare can be understood as the combined efforts of healthcare staff, patients and carers, and researchers “to make the changes that will lead to better outcomes (health), better system performance (care), and better professional development (learning)” (p.1).¹ The spread and adoption of QI has occurred across many clinical areas and settings, from diabetes management in primary care² to perioperative care delivered in hospitals.³ In surgery, QI programmes have shown promising results across the entire perioperative journey,^{4,5} which is defined as care received before, during and after surgery in periods known as pre-operative, intra-operative and post-operative, respectively.⁶ However, published accounts of QI in healthcare are frequently incomplete,^{7,8} which can produce important unwanted consequences. This thesis is the first piece of research to offer a careful examination of the problem of QI reporting in the perioperative academic literature, with the aim of understanding what needs to be done to improve reporting for QI stakeholders who are actively working in the field, such as authors, journal editors, and the consumers of QI work.

Surgery is an important focus for quality improvement. Major surgery is invasive, involving incision, excision or manipulation⁹ to investigate or treat a pathological condition which requires profound sedation or anaesthesia to control pain.¹⁰ A full explanation of why surgery is selected to be the focus of this thesis is provided in section 1.4, but reasons can be briefly summarised as follows: surgery is a high-risk specialty of healthcare where the impact of complications can include reduced survival.^{11,12} Global estimates of complication rates provided by the World Health Organization (WHO) range from 3% to 22%.¹³ Fourteen record-review studies (from Europe, North America, Africa and Australia) together indicated that around 14% of patients undergoing surgery will experience an adverse event, of which 3.6% can be fatal and 10% can be severe.¹⁴ Figures also suggest that in 5.2% of patients who experience an adverse event, the incident is potentially avoidable.¹⁴

In surgery, QI has become an important mechanism by which a range of problems, such as avoidable complications, can be examined, understood and

addressed to reduce their future recurrence.³ In addition to patient safety, changes implemented during QI work can be made with the aim of improving other dimensions of quality,¹⁵ which include: effectiveness, efficiency, patient-centredness, timeliness and equity.¹⁶ This thesis will focus on how QI is reported across the whole patient journey which includes the pre-operative, intra-operative and post-operative stages of care (encapsulated by the term perioperative).⁶ The terms surgery and perioperative have different meanings, but the phrase 'surgical research' is well recognised¹⁷ and it will be used throughout this thesis to relate to QI research which spans the perioperative journey.

QI work can involve a variety of approaches to achieve its aims, including methodologies adopted from the manufacturing industry, such as Total Quality Management (TQM), Plan-Do-Study-Act (PDSA) cycles, Lean and Six Sigma.³ Each approach will not be defined in detail here, but they can be defined in brief by exploring their common principles – they usually involve iterative measurement, with many cycles of data collection; an examination of capacity, flow or reliability in a system; and engagement of the workforce in the improvement work.¹⁸ QI often uses a suite of approaches (such as combining PDSA with Lean),¹⁹ rather than one standalone method.²⁰ Systems to support people who deploy and make sense of QI work now operate across the world. For example, in the United Kingdom (UK), a national perioperative quality improvement programme is being piloted with the aim of ensuring QI is an objective from the outset of perioperative outcome measure collection.²¹ In the United States (US), the National Surgical Quality Improvement Program (NSQIP) has reported promising effects on patient care.²²

There are many reasons why reporting of QI in the academic literature might be particularly problematic. QI is a new science^{23,24,25} where a consensus on terminology²⁴ and methodological direction²⁶ is still in development. QI work can also involve a large amount of variability in how interventions are designed and implemented, and how the study itself is designed and conducted.^{3,4,5,26,27,28,29,30} One systematic review suggests that a “virtually infinite number of combinations of features and local environmental circumstances” (p.244)²⁶ can make study design, and the reporting and synthesis of QI evidence, extremely challenging.

The available literature provides some examples of how challenges associated with reporting might be tackled, but practical strategies which move beyond the reporting checklist are still very much in their infancy.

This thesis seeks to answer two questions.

1. What is the completeness of reporting in the perioperative literature on QI methods and interventions and which elements are most frequently missing?
2. Why is the reporting of QI in perioperative care so hard?

This chapter will explain why it is now recognised that it is important to monitor, measure and improve quality in healthcare, and how the field that has become known as QI has developed. It will then explore the notion that surgery is an especially interesting area to investigate the challenges of QI reporting in more depth. A definition of what is meant by the term 'reporting' is provided. A second introductory chapter will explore the background literature in more detail to fully explain what has motivated the completion of this thesis and why it provides a compelling and helpful addition to the field of QI research. Each chapter in this thesis feeds into the next, and they will build a series of core arguments to fully explain why reporting in QI can be hard.

1.1 The emergence of quality improvement in healthcare

This section will explain how QI has evolved over many decades into a set of terminologies and approaches which are recognised by clinicians and healthcare researchers today. However, while QI approaches do have some common features, many discrepancies remain in how it is described, and work to improve healthcare does not necessarily always align with neatly delineated approaches.

In healthcare, the evolution of 'quality improvement' – often shortened to QI – emanated from a growing acknowledgement of the importance of a systematic approach to examining the quality of care.³¹ Throughout history, many individuals have raised awareness of quality. The work of Florence Nightingale and Avedis Donabedian will be used here to highlight two approaches to

tackling quality-related problems which have been particularly influential for modern-day healthcare practitioners and researchers alike.^{32,33}

Nightingale, a British nurse practising during the nineteenth century, raised issues of poor quality in her 1858 publication *'Notes on hospitals'*³⁴ where she eloquently provides baseline data, rationale and a clear description of planned interventions to tackle problems relating to infection control and staffing. She sets out a strong case for the need for improvement, organising data and measurement in a similar way in which a QI project might be set up today. Sheingold and Hahn³⁵ suggest that Nightingale's work can be aligned with QI through examining accounts of her improvement efforts during the Crimean war of 1854.

Attempts to systematically improve quality continued into twentieth-century healthcare and in 1966, Avedis Donabedian (a physician regarded as the 'father of quality assurance' p.472)³² proposed the structure-process-outcome model. This model evaluates the quality of healthcare by examining relationships between structures of care, processes of care and health outcomes through measurement.^{13,36} Since 1966, many other models have been developed to measure and improve quality, and such models (including Donabedian's) often recognise that the behaviour and values of the people working in healthcare systems should be studied³⁷ because a system can influence³⁸ (and be influenced by)¹⁵ these behaviours. The examination of behaviour is also recommended during QI work, to fully understand how an intervention works.³⁹ Thus, knowledge generated during an earlier movement to improve quality of care helped to build a platform from which QI could emerge as a separate discipline,³¹ and QI is still being helpfully supported by models which were established at an earlier time.^{37,40,41}

Despite a wealth of literature which documents an early commitment to improving quality of healthcare, quality did not emerge as a general theme in the medical literature until much later. A search of the International Secondary Information (ISI) service's database Journal Citation Reports suggests that the first journals indexed to emphasise the quality of healthcare in their titles were the *American Journal of Medical Quality* and *The Journal of Nursing Care*

Quality, which were both first printed in 1986. The prominence of efforts to improve quality in healthcare literature stimulated debate about how quality-orientated literature should be used as part of the broader evidence based practice movement.⁴² Thus, through the cumulative efforts of many in the fields of healthcare research and practice, a movement to study quality of care grew, and QI eventually came to be known as an important vehicle through which care could be improved.

Many ideas which began in industry have penetrated into healthcare to shape what has become recognised as QI today. In industry, as early as 1914 Frank and Lillian Gilbreth published on their techniques of 'scientific management' which were designed to optimise efficiency (originally in bricklaying) by reducing waste.^{43,44} Later, in 1924 while working at the Hawthorne telephone manufacturing plant in Chicago, Walter Shewhart described the first statistical process control (SPC) chart, a tool for identifying variance which is 'special cause' (data which vary in an unpredictable manner) and 'common cause' (variation which is inherent) in a system.^{45,46} Shewhart met William Edwards Deming and Joseph Juran who also made an important contribution to the development of QI. Deming promoted Shewhart's work on iterative measurement and feedback mechanisms, which became known as Deming cycles, or PDCA cycles,⁴⁵ and he developed a 'system of profound knowledge'.⁴⁷ Deming's system combined Shewhart's views on variation and quality control with elements of psychology and theory of knowledge, and this has since been used to inform the field of improvement science.¹⁵ Juran, who was also influenced by Shewhart, developed the 'Juran trilogy' which suggests that continuous improvement requires equal attention to be given to planning, control and improvement, three inter-related functions of industry.⁴⁸

The science of improving quality was then studied by others, including Taiichi Ohno and Kiichiro Toyoda who developed the Toyota Production System (TPS) during the 1940s and 1950s.⁴⁹ Vilfredo Pareto, Peter Drucker, Kaoru Ishikawa and Henry Ford^{50,51} have also all made important contributions to the growth of a movement towards improving quality throughout the twentieth century. Acquiring an appreciation of this early work can facilitate understanding of what QI actually is,¹⁵ and how it is described, which includes concepts such as

reducing variation and waste, improving work flow and system performance, and adapting the environment.⁵²

The factory floor of the 1920s was a world away from a Crimean field hospital of the 1850s, yet Shewhart, Deming, Juran, and Nightingale were asking similar questions about quality. Thus, in today's healthcare arena, various efforts to improve quality across fields as disparate as industry and healthcare have evolved to become known as quality improvement. From these disparate fields, QI has emerged as a relatively new field in healthcare. Unsurprisingly, it suffers from poor consistency in terminology and a lack of consensus on how it should be defined. There is no single accepted definition of QI,¹⁸ and phrases such as 'improvement science', 'science of improvement', and 'implementation science' have all been used to relate to QI.^{15,53,54} Walshe⁵⁵ studied the frequency of QI terminology in the scientific literature between 1988 and 2007, and found that some terms including Total Quality Management (TQM), audit, and patient safety have been used consistently in healthcare for the past 28 years. Other QI terms have come in and out of use over time, with QI approaches 're-invented' rather than 'innovative', which Walshe terms 'psuedoinnovation.' Powell and colleagues⁴ also reviewed QI terminology and found that terms such as Business Process Re-engineering, TQM and Continuous Quality Management (CQM) have been used in healthcare since the 1990s.

Despite the presence of many different terms and methodologies, each QI approach has certain characteristics in common.⁵⁵ QI approaches in healthcare tend to be systematic and cyclical, using a deliberate structured process involving measurement and data collection.^{1,55} They also tend to use a set of distinctive tools such as: Plan-Do-Study-Act (PDSA)/Deming cycles, Statistical Process Control (SPC), Lean, Six Sigma, TQM, and Continuous Quality Management (CQM), which are easily recognised as originating in the manufacturing industry,⁵³ or tools such as brainstorming and flow charting, which can be associated with many other fields.⁵⁵ Finally, QI approaches tend to be focused on a purposeful attempt to change frontline care in order to improve outcomes and quality of care.^{1,55,56} These changes are sometimes directed at the improvement of healthcare systems, but they can also focus on

specific healthcare procedures (such as the prescription of antibiotics,⁷ for example).⁵³

The next section will explain why QI has been positioned as important to healthcare. Only when it is clear why widespread and sustained attention has been given to the subject of QI in healthcare, can a case for its complete and accurate reporting be made. Conversely, it might also be argued that until complete and accurate reporting of QI becomes better established, the importance of QI to healthcare cannot be properly assessed.

1.2 The case for QI in healthcare

In healthcare, the movement to improve quality was brought to the public's attention in 1999, 2000 and 2001 by the publication of three reports: *To Err is Human: Building a Safer Health System*,⁵⁷ the *World Health Organization (WHO) Report on Improving Performance*,⁵⁸ and *Crossing the Quality Chasm: A New Health System for the 21st Century*.¹⁶ These reports drew attention to failures and, in particular, to preventable deaths caused by medical error. Since these reports were published, the volume of research publications about patient safety has increased by 72%.⁵⁹ Failures in safeguarding safety also formed part of the catalyst for QI work in the UK. Several public inquiries were important, but the Bristol Inquiry⁶⁰ in particular, which investigated deficiencies in paediatric heart surgery at the Bristol Royal Infirmary between 1984 and 1995, brought the public's attention to medical harm. The findings that unintended harm might be reduced by improving team work^{61,62} and communication,⁶³ and by adopting a systematic approach to monitoring care quality,⁶² were central to the developing field of improvement. As emphasis on the quality of care grew, so too did the drive to consciously place quality at the centre of policy making.⁴⁸ The UK's Department of Health published *An Organisation with a Memory*,⁶⁴ which suggested that safety could be enhanced by developing systems for learning and improving. Data capture and measurement lies at the heart of improvement work,⁴⁰ and this was promoted by the UK's Institute for Innovation and Improvement, which was established in 2005, and replaced by NHS Improving Quality in 2013,⁶⁵ and then the 'NHS Sustainable Improvement team' in 2015. Lord Ara Darzi's 2008 report '*High quality care for all*' maintained this emphasis

on quality,⁶⁶ emphasising that QI should be carried out by local clinicians who take ownership of their own data capture and learning.⁶⁷

Although a spotlight had been placed on improving quality, in 2009, the UK's Healthcare Commission, an independent regulator of healthcare quality, published its report on failings in care quality at Mid Staffordshire NHS Foundation Trust.⁶⁸ In 2010⁶⁹ and 2013⁷⁰ two reports led by Sir Robert Francis QC explained what happened at Mid Staffordshire and how lessons could be learned from it. These reports had a strong impact on the media, politics and research. Later in 2013, Don Berwick (senior fellow with the Institute of Healthcare Improvement (IHI) in the US) and Sir Bruce Keogh (NHS England's medical director) published reports^{71,72} on improving the safety of patients in England. As founder of the Institute for Healthcare Improvement (IHI), Don Berwick has endeavoured to keep quality in the hearts and minds of clinicians and researchers, and his report placed special emphasis on the adoption of QI methods to facilitate healthcare improvement. A number of national and regional groups were developed such as the NHS Sustainable Improvement team and NHS Improvement. These teams aim to promote and co-ordinate national and local efforts to plan, conduct and disseminate QI activity.

The adoption of QI activity has also been encouraged by the adoption of mandatory as well as voluntary QI activity for clinicians. For many years clinicians across the world have been required to take part in improvement activities such as audit⁶² (for example, in the US, New Zealand, Australia⁷³ and the UK^{74,75}) not only for clinical revalidation,⁷³ but also as part of their everyday work. Though audit is certainly an important tool for improvement, the QI activities which are recommended by official clinical bodies vary. In the US clinical students are encouraged to learn about a suite of improvement tools, but in the UK training in QI tends to have a narrower focus⁷³ (typically including audit which is a focus of the NHS standard contract).⁷⁵ Improving quality to avoid sub-standard care has become the focus of many governments worldwide,⁷⁶ but the key issues which influence the delivery of QI are likely to be different in each country of the world.⁷⁷ This thesis will, in future chapters, examine QI reporting internationally.

Public inquiry and government policy are clearly important drivers for QI, but the world of research has also provided an important stimulus for increasing awareness of the role and potential value of QI. A growing number of publications have described the application of QI techniques. Systematic reviews have demonstrated improvements in outcome using QI approaches. For example, in surgery, QI interventions have benefited patients across the whole perioperative journey including preoperatively (reduction in time to surgery),^{3,4} intraoperatively (reducing closure complications)³ and postoperatively (reduction of surgical site infection,³ central line infection,³ antibiotic errors⁵ and antibiotic administration compliance).^{3,7} However, cautionary notes have been published. The efficacy and effectiveness⁷⁸ of QI studies is not always strong,⁷⁹ and this could be due to varied methodological problems such as randomisation when QI is evaluated in RCTs,⁸⁰ or the variable fidelity in the application of QI techniques.⁸¹ In light of this, the literature does not paint a picture that is too favourable, stating that QI should not be seen as a 'magic bullet'^{76,20} to improve services. Yet, enthusiasm has not been hampered and authors call for improvement in the conduct in QI so that its value as a useful approach to tackle operational defects can be harnessed.⁸¹

Examining the literature on how QI came to be an important part of the landscape of healthcare quality has revealed that QI activities can be of different types. A QI study can be conducted as part of a programme of research or as a local QI project, and this distinction will be explored in the next section.

1.3 QI, research and QI research

Improvement projects are different to clinical research:^{30,81,82,83} they are typically pragmatic and led by local clinicians,⁶⁷ often in a single setting,⁸³ and they tend to be ongoing, or continual, in contrast to research which takes place over a defined period of time.⁸⁴ Research evidence⁸⁵ is a different type of evidence to that generated by local improvement projects. Øvretveit et al⁸³ define three types of QI research: research which examines the optimal conditions for an intervention's effectiveness at the level of an individual patient; research which examines the effectiveness of QI processes at an organisational level; and research which examines how systems of care produce outcomes. These types

of QI research tend to be conducted across several sites⁸³ and often seek to generate new knowledge by testing a hypothesis.³⁰ QI research has been likened to health services research, which typically aims to examine how care can be delivered to achieve the highest quality of care,⁸⁴ and applied research, which can “bridge the gap” (p.1962)⁸⁶ between the ideal setting of clinical trials and actual routine care. Research which can be more readily applied to usual clinical environments is a mechanism by which science can be translated into action with more ease,⁸⁶ and recognition of this has prompted interest in the growing field of QI research.

In QI research, which is usually applied to the real world, interventions tend to be complex.⁸⁷ They can involve many components which can be implemented in a number of different settings or organisational levels, and each intervention component can interact in expected and unexpected ways at varying stages of the intervention-to-effect pathway.⁸⁸ Interest in researching complex interventions has occurred alongside the growth of pragmatic trials, which test an intervention under the usual conditions in which it will operate, in contrast to explanatory trials which test causal hypotheses.^{78,89} QI research thrives in the arena of applied science, and it aligns well with the view that pragmatic and explanatory RCTs are not always mutually exclusive, because QI can operate within either pragmatic or explanatory trials, or a mixture of the two.³⁰ QI research can also be grounded in a ‘broad church’³⁰ (p.2) of other types of evaluative study beyond clinical trials, including health services research and other disciplines. Thus, QI research may require the adoption of some principles which originate in clinical research, such as the explicit reporting of rationale,²³ as well as the adoption of new approaches, such as the reporting of real time data about context in complex healthcare systems.⁹⁰ The need to recognise the practical and applied nature of QI research is highlighted here as a driver for QI work, but complexity in intervention design and implementation might also be problematic, complicating attempts to provide full reports of QI work.

While some features have been proposed which can distinguish QI projects from QI research, what counts as research evidence is not always easy to delineate.⁸⁵ For example, generalisability could be used to distinguish research from non-research (NHS Health Research Authority (HRA) decision tool:

<http://hra-decisiontools.org.uk/ethics/EngresultN1.html>). However, some types of research place more value on generalisability than others⁹¹ and both improvement projects and QI research require careful attention to detail so that generalisable lessons might be gleaned from the evidence gained from individual projects.⁹² Both improvement projects and research attempt to understand the complex interactions that occur within a healthcare system, rather than control them as a traditional clinical trial might do. Portela et al³⁰ point out that debate about whether it is useful to distinguish between local practical improvement projects and scientific research, remains unsettled. When agreement is poor about whether QI is regarded as research or not, varied expectations about what is reported and how it is reported could translate into inconsistent or incomplete QI reporting. This thesis will further explore this debate about what makes QI research, research, through review of the literature and empirical work.

1.4 Why is perioperative care an area in which the challenges of QI reporting should be explored?

Although some data suggests that the effectiveness of QI is limited, support for QI is widespread⁸¹ and this heightens the need for reports that are published to be written completely and explicitly so that reasons for failures as well as successes can be fully understood. Perioperative care is an especially important area in which to explore the challenges of QI reporting in more depth because QI is a mechanism by which both efficiency and safety can potentially be improved for a large global volume of patients who receive 234 million major surgical interventions per annum.^{10,13} Surgical teams are encouraged to participate in QI activity to meet performance standards required by regulatory bodies, and to provide data which is visible in the public domain.^{93,94} Finally, both QI and perioperative care have a number of complementary features.

First, perioperative care as a scientific field is committed to encouraging innovation,⁹⁵ applied research, and the development of complex behavioural and technological interventions which can improve the quality and safety of patient care.⁹⁶ The perioperative literature recognises a need for responsible innovation, that new techniques are not immune to problems,⁹⁷ and activities to improve quality are held in high regard.⁹⁶ Second, the value of qualitative

research is growing in both the QI and perioperative literature. The use of qualitative methods was documented by perioperative researchers as early as 1995,⁹⁸ and this has continued into the 2000s.^{99,100} Surgical researchers are also encouraged to embed qualitative approaches into their research design, to learn more about the recruitment problems which can be particularly troublesome for surgical trials.¹⁰¹ Likewise, QI literature has encouraged the use of qualitative approaches for the collection of data about context, which can be narrative in nature.¹⁰² Third, in surgical research, small changes may be seen in the technique of one surgeon, or between different surgeons on a case-by-case basis.⁹⁵ This is not the same as the deliberate modifications to a QI intervention which occur through several iterative cycles, but perioperative researchers may be more accustomed to documenting complex interventions, which may not always operate in a completely standardised manner. Consequently, the field of surgery and perioperative care could offer fertile ground for the uptake of QI, and for further study to examine QI reporting and what practical solutions might be developed to improve it.

1.5 What is reporting?

In historical usage, the term 'reporting' (defined in the 1926 edition of the New Gresham English Dictionary) means 'to carry back' or 'to bring back', 'to give an answer', 'to relate', 'to tell.'¹⁰³ Although the literal meaning of this word remains the same, many mechanisms by which reporting might be conducted have become established in the nine decades since this dictionary was published. This thesis will focus on the reporting of QI in peer reviewed journal articles.

Academic journals are repositories of evidence collected in paper and/or online formats as "articles." The process by which an article in a journal becomes published is not linear and involves communication between teams of authors and editors. A manuscript (which will be called an 'article' throughout this thesis) which is submitted for publication is managed by an editorial team. This involves a group of individuals who may work in a hierarchical structure, including: senior editors (sometimes known as editors-in-chief, or chief editors); deputy editors (sometimes known as editor of scholarly content, or section editor); and associate editors. They use clear structures and rules, such as

subjecting articles to external peer review, to evaluate whether the manuscript is fit for publication.¹⁰⁴

The question: '*What motivates good reporting?*' should first include consideration of the ethical and regulatory background. In research, this has been well described. The Declaration of Helsinki, a set of ethical principles for research involving human subjects which has also been adopted by the World Medical Association (WMA),¹⁰⁵ states that researchers have an ethical duty to ensure that accurate and complete results of their research (whether positive or negative) are made publicly available. Ensuring complete study reports are publicly available is also emphasised by funders and governing bodies for clinical research such as the National Institute for Health Research,¹⁰⁶ the Health Research Agency¹⁰⁷ and the Medical Research Council¹⁰⁸ in the UK, in addition to the European Medicines Agency.¹⁰⁹ These regulations emphasise a requirement to publish, but more than this, they illuminate a need for authors to publish with integrity, which means providing honest complete accounts of what was done during any research study.¹¹⁰

The motivation to improve the quality of reporting has been promoted by the publication of a number of influential papers, in particular, Greenhalgh et al's work 'Evidence based medicine: a movement in crisis?'¹¹¹ and Ioannidis' work 'How to make more published research true'.¹¹² These articles support a movement towards ensuring research findings are not only published, but published with details which make the findings reproducible. Campaigns such as AllTrials also contribute towards this movement.¹¹¹ Although steps are being taken towards creating a scientific community in which better reporting can flourish, shared standards for data sharing practices are still an aspiration and the level of enforcement for registration of research (and making it available) varies between funders, sponsors, legislators and research ethics committees.¹¹³

Publication of research findings which are fully reported is important in enabling healthcare staff to choose interventions wisely, avoiding treatments which are unlikely to provide any benefit.^{114,115,116} Systematic review supports this by summarising the findings and methodological quality of a large number of

published articles.¹¹⁷ Meta-analysis can be used to combine data from several independent experiments, and this can provide a more accurate and transparent appraisal of literature.¹¹⁸ The decision to meta-analyse data is influenced by heterogeneity. A low degree of heterogeneity between the selected articles is preferable,¹¹⁹ as it increases the strength of any associations uncovered between groups¹¹⁷ so that pooling data from similar studies can provide a helpful summary of results. It is important therefore that complete reports are available of all parameters at population, intervention and outcome level, in order to allow systematic reviewers to accurately assess studies for similarity later on.^{116,120}

Another reason to publish fully complete reports, which is closely linked to a sense of moral duty to both patients and healthcare staff, is that research which is shared with other researchers may be less likely to be unnecessarily duplicated, reducing waste of finite healthcare and research resources.¹¹³ However, the blurred boundary between QI and research complicates the straightforward adaptation of these principles to the field of QI.

This chapter has shown that QI in healthcare has grown from efforts (which have evolved over many years) to systematically improve quality in healthcare, and from ideas which originated in industry, in fields including bricklaying, telephone engineering and car manufacturing. The case for QI in healthcare has been supported by the results of public inquiry and government policy. Perioperative researchers and clinicians are encouraged to drive forward local improvements in care and QI research, and to promote the spread and adoption of techniques which could improve care in other settings. The next chapter will build on this knowledge to explain the challenges researchers and clinicians face when attempting to describe their QI work in academic publications, and why it is so important to overcome them.

2 BACKGROUND TO THE CHALLENGES OF QI REPORTING

This chapter will present a literature review to explain which aspects of QI tend to be poorly reported and what interventions are available to improve reporting. The problem that high-quality reporting in the field of QI is likely to be a particular problem due to the manner in which it has evolved (and continues to evolve) as a scientific field will also be explored. Finally, this chapter will explain what the consequences of poor reporting are.

2.1 The problem of reporting in QI: a review of the literature

Researchers have raised concerns that harnessing learning from published work about QI interventions has been difficult.^{76,121,122} One of the barriers to learning is poor reporting,¹²³ as is clear from three systematic reviews^{7,8,124} (listed in Appendix 1) on reporting quality in QI, and eight systematic reviews^{3,4,5,26,27,28,29,125} on more general topics involving QI in perioperative care (such as the efficacy of QI). This section will describe a literature review which included these 11 articles.

The quality of reporting of QI has been directly examined in two systematic reviews. Taylor et al⁸ evaluated reporting of PDSA cycles in healthcare, and found that 81% of included articles failed to provide a full description of the number and timing of PDSA cycles. Levy et al⁷ evaluated reporting of QI interventions to increase compliance with antibiotic prophylaxis in surgery and found that 75% of included articles failed to report missing data and 67% failed to report generalisability adequately.⁷ A third systematic review by Ivers et al¹²⁴ did not evaluate the quality of reporting directly, but set out to identify the risk of bias of QI studies evaluating interventions to improve care in diabetes. Ivers et al called for “quality improvement in quality improvement” (p.1)¹²⁴ because risk of bias was frequently high, but also because reporting was inadequate. Allocation sequence and concealment was unclear in 58% and 55% of included trials respectively.

Six systematic reviews^{3,4,5,26,27,125} sought to answer questions about the efficacy of QI in perioperative care. Although these systematic reviews were not designed to specifically evaluate reporting quality, the authors extracted some information from the included articles about data which was missing or

incomplete. One of these reviews focused specifically on perioperative care,³ and five focused on healthcare more generally but included articles relating to the care of surgical patients.^{4,5,26,27,125} Two further systematic reviews, published in 2013²⁸ and 2014,²⁹ examined the use of surgical safety checklists, and these are included here because a checklist is commonly reported as a type of QI intervention.¹²⁶

All of these eight systematic reviews except for one, which focused specifically on the technical application of control charts,¹²⁵ commented on the heterogeneity of the contextual features of QI research. The authors explained that differences between setting,^{3,4,26,27,28} leadership^{28,29} and charisma⁵ (which are arguably all aspects of context)³⁹ can complicate attempts to establish firm recommendations about the efficacy of a specific intervention, but they did not comment on the completeness with which this contextual data was described. Other reporting items which were incomplete were reported. For example, harms and unintended consequences of QI was found to be poorly reported by Nicolay et al³ and Treadwell et al.²⁸ The reporting of cost was found to be poor by Nicolay et al,³ who found that cost reported in only two of 34 included studies, and by Alexander et al,²⁶ who found that cost was reported in only six of 185 included studies. Not all authors reported this finding, and a deeper discussion on cost was provided by Thor et al²⁷ and Treadwell et al.²⁸ Nicolay also found outcome data reporting to be generally poor.³

When the 11 systematic reviews described above are evaluated using the externally validated¹²⁷ 11-item 'Assessment of Multiple Systematic Reviews' (AMSTAR) tool,¹²⁸ it is possible to gain an overall picture of the strengths and weaknesses in systematic reviews of QI research. Using AMSTAR, one point can be allocated for each completely reported item and these 11 systematic reviews scored an average (mean score) of 7.07/11. Doing this reveals that two of the lowest scoring AMSTAR items were the provision of a description of the publication types of the included articles (provided by three reviews)^{4,26,29} and a list of all included studies (provided by five reviews.)^{5,7,28,29,125} These items indicate whether a systematic review has included *all* relevant literature. The reviewers all evaluated the methodological quality of their included papers but

they mirrored concerns expressed by Grimshaw et al in 2003,¹²⁹ that evaluating the methodological quality of QI can be challenging.

Poor reporting in original papers and in systematic review is evidently a problem across many reporting domains such as the QI method, study design, adverse events and economic evaluation. This provides a strong indication that QI reporting in perioperative care is likely to be problematic, and is thus a focus of my thesis.

2.2 What reporting tools exist to improve reporting in QI?

Reporting guidelines are tools which authors can use to check whether they have reported on every item of their study. The first reporting guideline was published in 1989, for the reporting of controlled trials.¹³⁰ In 1994, concerns were still being raised about inadequate reporting in medical research¹³¹ and the Standards of Reporting Trials Group (SORT) recommendations were also published.¹³² This led to the development of the CONSORT statement in 1996,¹³³ offering reporting guidelines for authors of RCTs.

A large number of reporting guidelines have now been published. In a recent systematic review of reporting guidelines endorsed by journals, 101 reporting guidelines were identified,¹³⁴ and as at 10 November 2016, 345 reporting guidelines were listed on the Enhancing the Quality and Transparency of Health Research (EQUATOR) network website (<http://www.equator-network.org>), which is a comprehensive searchable database of reporting guidelines. Recent guidelines include those relating to: Health Estimates Reporting¹³⁵ (published in June 2016), clinical guidelines¹³⁶ and health interventions using mobile phones¹³⁷ (both published in March 2016).

Specific guidelines for intervention reporting emerged after concerns about the adequacy of reporting of non-pharmacological interventions had been raised by researchers including Glasziou and Hoffman. They analysed: a consecutive sample of 133 randomised trials describing surgery, therapy, or education;¹³⁸ a sample of 200 randomised trials describing physiotherapy;¹³⁹ and a sample of 60 systematic reviews describing stroke interventions.¹⁴⁰ These studies concluded that poor reporting of intervention characteristics is common.

Many different reporting guidelines are relevant for authors of QI studies. For example, reporting guidelines have been developed for different types of study design including: the RCT;¹⁴¹ observational studies;¹⁴² case reports;¹⁴³ systematic reviews;¹⁴⁴ qualitative studies¹⁴⁵ (and the synthesis of qualitative studies);¹⁴⁶ diagnostic studies;¹⁴⁷ and protocols.¹⁴⁸ Since QI research can be conducted using many of these design types,³⁰ they may all be of use to a QI author.

Guidance has also been produced to aid reporting of quality improvement reports, specifically: The Standards for Quality Improvement Reporting Excellence (SQUIRE) 2.0.¹⁴⁹ SQUIRE 2.0 includes a reminder for authors to include a description of their intervention rationale (a reason why it was thought to work), the intervention itself, and the contextual elements thought to influence the intervention. SQUIRE 2.0 encourages authors to write “in sufficient detail that others could reproduce it” (p.2).¹⁴⁹ SQUIRE 2.0 may therefore be used alongside other guidelines which have been designed specifically to guide reporting of an intervention.

I identified seven publications (Table 1) which are relevant to reporting each separate component of an intervention which can be described in any evaluative study.

Table 1 Intervention reporting tools

Reporting tool	Reference
TIDieR	Hoffmann TC; Glasziou PP; Barbour V, et al (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. <i>BMJ</i> . 2014;1687(March):1–13. ¹⁵⁰
ITAX	Schulz R; Czaja S; McKay J; et al (2010). Intervention Taxonomy (ITAX): Describing Essential Features of Interventions. <i>Am J Heal Behav</i> . 2010;34(6):811–21. ¹⁵¹
EPOC checklist	McAuley, L; Ramsay C (2002). Cochrane Effective Practice and Organisation of Care (EPOC) Review Group, Data Collection Checklist. Ontario: Institute of Population Health, University of Ottawa. ¹⁵²
CONSORT extension	Boutron I; Moher D; Altman, D et al (2008). Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration. <i>Ann Intern Med</i> . 2008;148(4):295. ^{153,154}
Journal Article Reporting Standards (JARS)	Cooper H (2008). Reporting standards for research in psychology: why do we need them? What might they be? <i>Am Psychol</i> . 2008;63(9):839–51. ¹⁵⁵
The Western Journal of Nursing Research intervention checklist	Conn V (2012). Unpacking the black box: Countering the problem of inadequate intervention descriptions in research reports. <i>West Jn Nurs Res</i> . 2012;34(4) 427-433. ¹²³
STARI	Pinnock et al (2015). Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. <i>Implement Sci</i> . 2015;10(1). ¹⁵⁶

Two references are notably different from the others in this list. The EPOC data collection checklist and the ITAX are not reporting guidelines. The EPOC checklist was originally intended to guide systematic reviewers on the extraction of data from primary studies,¹⁵² and ITAX is a taxonomy,¹⁵¹ which is a method of managing information on a particular topic. But they helpfully categorise the elements of an intervention, and authors might use them as a resource to understand what fully complete intervention reporting should look like.

These guidelines have a number of features in common. They all require a description of: the intervention setting, who delivered the intervention, the mode of delivery of the intervention (for example, face to face) and the dose or timing of the intervention. Six of the seven checklists request a description of the rationale for the intervention,^{123,150,151,152,155,156} and six request a description of the materials used to deliver the intervention and the fidelity of the intervention.^{123,150,151,154,155,156} However, only four request details of any modifications carried out during the course of the study,^{123,150,151,156} or how the intervention was titrated for individual use.^{123,150,151,154}

Two of these guidelines^{150,156} directly state that they have used Moher et al's guidance for developers of health research reporting guidelines,¹⁵⁷ which are endorsed by the EQUATOR network. The remaining guidelines were written before Moher et al's guidelines were published. Three authors (of the CONSORT extension,^{153,154} ITAX¹⁵¹ and JARS¹⁵⁵) reported that a consensus process was used, similar to that described by Moher et al.¹⁵⁷ The process by which the EPOC and Western Journal of Nursing Research intervention checklists were developed was less clear.

All seven intervention guidelines have been cited by other PubMed indexed research. As at 1 November 2016, the TIDieR checklist¹⁵⁰ and the CONSORT extension¹⁵⁴ have each been cited over 100 times. The other intervention reporting guidelines have been cited between one (for the Western Journal of Nursing Research intervention checklist and the EPOC checklist) and 36 (for the JARS guideline) times. The use of citation rate has been criticised because it may not be an accurate reflection of impact or clinical relevancy, but it is a helpful indicator of the popularity and spread of this work.¹⁵⁸

Authors of QI research in perioperative care are often reminded that their interventions are likely to be complex. Thus, they may also benefit from using reporting guidance for describing complex interventions.⁸⁷ Reporting guidance for describing the contextual features of the organisation or setting within which the intervention operates have not been developed, but the Model for Understanding Success in Quality (MUSIQ) tool, which is described as a 'conceptual model' for contextual features in QI, could be used as an aid for reporting.³⁹

In addition to guidelines to improve intervention reporting, the authors of QI research in perioperative care are provided with guidance to improve intervention design in both QI¹⁵⁹ and surgery.^{160,161} The use of these tools during the conception and early design of a study could also facilitate more complete reporting when it is time to write the study up later on. Finally, the authors of QI research in perioperative care may benefit from referring to the QI Minimum Quality Criteria Set (QI-MQCS).¹⁶² This tool was not designed as a reporting guideline either, but its 16 separate domains, which are recommended for critical review of QI research, could be used as an indicator of what good reporting might look like. For example, some of the QI-MQCS items include: the intervention rationale, intervention description, implementation, study design, fidelity, and organisational characteristics. The QI-MQCS could be used alongside 194 other tools put forward to evaluate the quality of non-randomised studies,¹⁶³ and at least three tools put forward to evaluate the quality of randomised studies.¹⁶⁴

The effect of individual reporting guidelines on reporting completeness has been evaluated in specific disciplines such as rehabilitation medicine,¹¹⁶ and in healthcare literature more broadly. Howell et al¹⁶⁵ evaluated the impact of the SQUIRE guidelines (SQUIRE version 1.0)¹⁴⁹ seven years after they were first released. They examined 56 QI articles published before the SQUIRE guidelines and 97 QI articles published after the SQUIRE guidelines and found no significant difference between the two groups in completeness of reporting. In 2006, 10 years after the publication of the CONSORT guidelines, a similar study by Plint et al¹⁶⁶ identified eight articles which had evaluated reporting before and after the publication of the CONSORT statement in 1996.¹³³ Using

journal endorsement of CONSORT, which is typically a statement encouraging authors to adhere to the guideline,¹³⁴ to evaluate CONSORT adoption, they found that use of CONSORT tended to influence RCT reporting for the better, with some reporting items (such as method of sequence generation) affected more than others (such as blinding).¹⁶⁶ Although CONSORT seemed to have a positive impact on reporting, researchers including Altman and Chan remained concerned about the adequacy of methodological reporting in randomised trials,¹⁶⁷ but a recent Cochrane review has demonstrated that journal endorsement of the CONSORT checklist does improve reporting.¹⁶⁸ In contrast to this, a review which excluded CONSORT (because so much previous work has been conducted on its impact) could not identify a relationship between journal endorsement of seven reporting guidelines (BMJ economic checklist, CONSORT for harms, PRISMA, QUOROM, STARD, STRICTA and STROBE) and the completeness of reporting.¹³⁴

2.3 Why is reporting problematic?

The literature reported in section 2.1 suggests that QI reporting is generally poor across a range of items including description of the intervention, QI method, and internal and external validity. A plethora of reporting guidance, tools to aid intervention design, and tools to aid critical appraisal already exists, some of which have been endorsed by journals (section 2.2). This leaves a puzzling question: if researchers have access to guidance and tools for the improvement of reporting, why then is reporting so hard, to the extent that items remain frequently missing from published reports and intervention descriptions are incomplete?

This section will explore the literature on healthcare reporting more generally, first on challenges associated with using reporting guidelines, and second on the challenges associated with the writing and publication process itself.

2.3.1 Challenges of reporting guidelines

The literature describing the impact of reporting guidelines (section 2.2) draws attention to many problems. First, research can take years to have an impact on practice and for action to be taken.¹⁶⁹ A good example of this can be taken from the recent work of Altman and Simera.¹³⁰ They explain that concerns about the

communication of findings from medical research were raised as early as 1934, and a need for high reporting standards was emphasised throughout every following decade. Yet, it took from 1934 until 1989 for the first reporting guideline to be published. To use a more up-to-date example, it could be suggested the second, more recent iteration of SQUIRE (version 2.0),²³ published in 2015, has not yet had time to influence reporting practices. The problem of time lag could explain the vexing conundrum that endorsement of CONSORT seems to have affected reporting practice, yet endorsement of other reporting guidelines has had less influence. Perhaps CONSORT, as the first guideline to receive wide acceptance, has had more time to influence practice and to stand alone during its early years without facing competition from hundreds of other reporting guidelines.¹¹⁶

Second, reporting guidelines and checklists can be problematic because many interrelated factors influence their use. In a recent editorial in the *International Journal of Nursing Studies*, Buss and Agdal¹⁷⁰ suggest that peer reviewers should judge research articles using practical knowledge, rather than the literal application of reporting guidelines. They argue that literal use of reporting guidelines could cause articles to be inappropriately rejected which could be damaging because potentially helpful research could remain unpublished, invisible and unable to contribute to the evidence base. Thus, the act of reporting scientific research is not a linear process which relies on the mechanical application of rules. Instead, it is complicated, non-linear, and can involve teamwork, intuition and practical experience.¹⁷⁰

Finally, the sheer number of reporting guidelines could be cumbersome. At least 345 reporting guidelines (listed by EQUATOR) and 194 quality evaluation tools (listed by Deeks et al)¹⁶³ exist. This means that authors may feel overwhelmed and unsure which tools to select¹¹⁶ and they may not have adequate training or guidance on how to use them.¹⁷¹ For example, Wells et al¹⁷² used in-depth case studies to establish that authors should accompany an intervention description with sufficient detail about context, but they do not advise which reporting tools should be selected to do this. Researchers have addressed this in the field of public health, by proposing a sequential approach for deciding how to select

reporting tools,¹⁷³ though such a framework has not yet been developed for the authors of QI research in perioperative care.

2.3.2 Challenges of the writing and publication process

The process of submitting an article for publication has been examined by Smyth et al¹⁷¹ who interviewed 59 clinical trialists, 48 of whom were chief investigators. This qualitative study was conducted as part of a larger project called ORBIT^{171,174,175} which aimed to explore the extent of the problem of reporting bias in clinical trials. Many trialists (41%) did not have their article accepted at submission to the first journal, and they resubmitted it elsewhere. Smyth et al¹⁷¹ identified various factors (the journal's impact factor, the direction of their research findings and whether they felt the journal would publish negative results, and word count restrictions), which influenced where the article would be published and how long the article would be. Smyth et al¹⁷¹ also suggest that difficulties with guidance from editors and peer reviewers on how to redraft an article are common and this can create a mixture of reporting problems such as missing information.

Reporting difficulties in QI specifically have been examined qualitatively by Davies and colleagues,²³ who authored the SQUIRE guidelines. Using interviews and focus groups with 29 users of the SQUIRE guidelines, they identified that while the SQUIRE guidelines were useful, some uncertainties, like feeling unsure how to report mechanisms of change or intervention modification, persisted.²³ Qualitative interview studies to investigate how QI research is conducted and reported, beyond the use of the SQUIRE guidelines, and what could be done to improve it, have not yet been completed.

The question of what to report and how to report it is accompanied by more fundamental challenges about wider incentives and motivations. For example, using systematic review, Scherer et al¹⁷⁶ identified that just over half of abstracts presented at conferences were later published. The 27 included articles used a mixture of open and close-ended questionnaires to ask the abstract authors why their research was not published. Lack of time was the most prevalent reason, and ranked by the participants to be the most important. Scherer et al¹⁷⁶ report that less frequent reasons given for non-publication

include: “trouble with co-authors”; “publication not an aim”; “publication not permitted by sponsor”; and “negative results.” While Smyth et al¹⁷¹ explain why publications might include incomplete (partially reported) items or omitted items, Scherer et al¹⁷⁶ expose the reasons why some research may never be published at all, creating poor reporting by invisibility.

The literature includes some suggestions of how the reporting problems of missing information, or non-publication, might be addressed. Gattrell et al¹⁷⁷ compared the reporting quality of published RCTs which declared (or did not declare) the support of professional medical writers. They found that support from medical writers improved the completeness of six out of 12 CONSORT reporting items.¹⁷⁷ Manuscript preparation time was not evaluated as part of this work, but lack of time¹⁷⁶ could be a factor which leads to enlisting the help of medical writing support. Conn,¹²³ author of the Western Journal of Nursing Research intervention checklist, explained that challenges to reporting might be overcome by involving a team of active co-authors; by using citations rather than detailed descriptions of well-established methods, and by using the active rather than passive voice, thus saving space for a more complete intervention description.

Chalmers and Glasziou recommend that awareness of reporting guidelines should be raised, and they emphasise the use of reporting guidelines in their model of the “stages of waste in the production and reporting of research.”¹⁷⁸ They further recommend that authors and journal editors should be trained in guideline use and that peer review should be supplemented by expert review from methodologists and the consumers of the research.¹⁷⁸ In the field of QI, organisations dedicated to healthcare improvement have also responded to the challenges associated with ensuring reporting is complete. They have developed practical tools such as online writing tips (www.ihl.org), webcasts (a video broadcast over the internet)¹⁷⁹ and writing conferences (www.squire-statement.org.)

2.4 Why is reporting QI likely to be particularly problematic?

This section will explain that reporting in the field of QI is likely to be a particular problem due to the manner in which it has evolved (and continues to evolve) as a scientific field.

2.4.1 The evolution of QI as a scientific field – type of reporting

Various authors have expressed the notion that QI is an ‘emerging’ or new field of science^{23,24,25} and this has involved some disagreement about *how* authors should present their work scientifically. The urge to present QI as a science can influence many things, including the choice of scientific language and terminology, the scientific reporting of the study design (such as randomised or observational), and the QI method (such as SPC, which is used in QI work to identify variance which is ‘special cause’ (data which vary in an unpredictable manner) and ‘common cause’ (variation which is inherent) in a system.^{45,46}

The scientific language which has been used in recent surgical QI research includes: cause and effect,³ prediction to estimate risk,¹⁸⁰ and effect size,³ which are helpful to understand the relationship between an intervention and an effect, but writing about QI and understanding it as a scientific field¹⁸¹ may require a different approach. Davidoff helpfully points out that improvement interventions are social treatments, not pills.¹⁸² The dose, frequency, ingredients and titration of a pill can be replicated exactly, but QI interventions can involve many different technical components, applied in different combinations,^{26,183} which can be modified as more iterative cycles of measurement occur.³⁰ The extent of this complexity can be highly variable between studies.¹⁸⁴ Also, while the setting for a pharmaceutical intervention (the human body) could involve a multitude of variables, the setting for a QI intervention (a social unit such as an operating theatre, ward or whole organisation) is perhaps even more complex, variable and difficult to predict. Thus, QI interventions and their settings are often more difficult to describe than a pill. Taxonomies have been designed to help authors select which ‘type’ of intervention their work should be categorised as,^{120,185,186} but these can be problematic, as Colquhoun et al¹⁸⁷ explain: one author may use the behaviour change wheel¹⁸⁸ to refer to an intervention as ‘environmental restructuring’, and another may use the EPOC framework¹⁸⁹ to describe the

same thing as a 'structural intervention'. Thus, there is considerable heterogeneity in the interventions themselves, but also in the tools and terminology available to researchers to classify and describe them. Colquhoun et al are responding to this problem as part of an international working group that aims to provide a single consensus on the classification of interventions.¹⁸⁵

QI research is not the only field in which interventions are complex and hard to describe: interventions in surgery¹⁶¹ and rehabilitation^{140,190} can also be described as complex, and difficult to describe. For example, the surgical procedure alone can be performed with "an almost infinite set of subtle variations" (p1).¹⁶¹ The contextual and social features which Davidoff¹⁸² points out to be important in QI can be an important reporting feature in non-QI studies.¹⁷² Surgical research includes many examples of social and behavioural processes, such as the competency and skill of the person delivering the intervention and the notion that the surgeon may become part of the intervention itself.⁹⁵ However, Davidoff's analogy, which distinguishes QI as an especially difficult field in which to describe an intervention, is helpful because the QI literature has a particular focus on the contextual and social features of QI work.

Contextual features including leadership, a supportive culture, motivation to change, and QI team norms are the most powerful determinants of success in QI in healthcare settings.^{102,191} Social actions such as "persuading people that improvement is possible, encouraging and maintaining participation, and learning from each other" (p.198)⁵⁴ have also been proposed as key determinants of success in QI.¹⁹² These reporting features, which require particular attention in QI, are vast in number. Context, for example, is proposed by some authors to be disaggregated into 25 separate characteristics,³⁹ and these can be linked to eight categories provided by Estabrook et al:¹⁹³ culture; leadership; evaluation; social capital; informal interactions; formal interactions; structural and electronic resources; and organisational slack. These contextual features may be poorly prioritised, leaving authors unsure about which should and should not be reported.¹⁹⁴ To use the words of Catchpole, "untangling the knots in people and systems" (p.1)¹⁹⁵ is not easy. Yet authors of QI reports must try to do this, and articulate it in a language which is easy for others to

understand if they would like their published QI work to be helpful to others in settings which are distant to their own.

QI literature has also raised the “curious methodological wrinkle” (p.395)²⁷ that a QI method (such as SPC) can be regarded as part of an intervention.^{27,196} The overlapping acts of measurement, method and intervention are characteristic of QI, but may make it difficult for authors to understand which features of their study should be described in which sections of their report. The type of methods can vary enormously between studies.²⁶ Nicolay et al³ described at least eight different types of QI method used in perioperative care. There is considerable heterogeneity in how each of these methods is described. For example, audit and feedback have been found to include up to 17 different modifiable elements.¹⁹⁶

Researchers acknowledge that as the scientific field of QI evolves, a large number of terms are being used to describe QI,¹⁸⁵ and these can frequently change over time.^{24,55} Thus, it is perhaps not surprising that a single accepted definition of QI is still lacking.¹⁹¹ To add to this, as the scientific field evolves the volume of publications in QI is growing,⁵⁹ and the work is likely to be scattered across a large number of journals and databases.¹⁹⁷ The inconsistent use of key terms and a wide spread of subject categories for indexing QI work could make literature searching difficult. If it is hard to identify the QI literature in the first place,^{197,198} systematic reviewers may not be able to obtain *all* relevant literature. Also, if the literature is hard to find and includes inconsistent terminology, QI authors who wish to use previously published reports to aid the write up of their own¹²³ may find the task of writing hard.

2.4.2 The evolution of QI as a scientific field – depth of reporting

There is little consensus on how much detail should be provided for descriptions of context, implementation and behaviour¹⁹⁹ that are typical of QI work. Some Cochrane reviews of QI have named these variables as ‘effect modifiers’²⁰⁰ (the presence of a third variable which can modify an intervention’s effect),²⁰¹ and choosing how tightly to describe these ‘modifiers’ could be affected by the type of study design.¹⁶¹ For example, researchers implementing a QI study of randomised design may want to apply tight specifications on the types of

population and interventions studied to limit differences between cases and controls. Yet, where social interventions, like QI interventions, are evaluated researchers may need to accept that variability exists that can be accounted for by research designs (rather than controlled out).²⁰² Deciding how the study design can affect intervention description is complicated, not least because QI research can involve a wide variety of study design types.³⁰

2.4.3 Understanding how to report QI in perioperative care

Reporting QI could be particularly problematic if the training needs of those who are actually doing the reporting are not met. A need for surgical staff to better understand research methodology,²⁰³ quality improvement and quality measurement²⁰⁴ has been raised and the ability of curricula to deliver this information could contribute towards successes and failures in QI reporting practices.

Three systematic reviews have investigated i) the effect of a specific QI training curriculum on perioperative staff's QI knowledge and ability to implement QI activities,^{205,206} and ii) the methodological quality of the curriculum content.²⁰⁷ These reviews included a total of 39 articles (of which 22 were controlled trials) and 59 separate QI curricula. Interestingly, the curricula varied in quality, and QI-specific educational objectives were not always well applied.²⁰⁷ Nevertheless, these reviews conclude that the perioperative learner's confidence in QI application improved.^{205,206} This offers relevant insight because although these curricula are not immune to problems²⁰⁷ and are not widely applied outside of the US,⁷³ proper training and continuing education in research methods (including QI) may in turn promote the reporting and publication of sound QI project and research outputs. Specific training on how to use reporting checklists¹⁷⁸ and literature searching²⁰⁸ may also be beneficial.

2.5 What are the consequences of poor reporting?

This section will draw on the literature to explain why getting reporting right is so important. Three main consequences of poor reporting will be highlighted: threats to the conduct of systematic review, threats to the replication of successful studies, and research waste. If systematic review conduct is inadequate; if successful studies cannot be replicated; and if research is

wasted, then the ability of policy makers, clinicians, healthcare managers and researchers to use research to benefit patients will be threatened.

2.5.1 Threats to the conduct of good systematic review

The problem of clinical (diversity in intervention type and setting) and methodological (diversity in the design and conduct of a study) heterogeneity is well recognised in systematic reviews of perioperative QI research.^{3,4,5,26,27,28,29} This degree of variance could make the task of reporting complicated, and if this is the case, then the task of systematic review also becomes more complicated.

The extent of heterogeneity in reporting across QI publications can create difficulties for the systematic reviewer at each stage in the review process. First, if terminology varies considerably between studies it may be hard to capture *all* relevant literature,^{24,197,198} and if a reviewer is not able to be fully exhaustive, less reliable estimates of the effectiveness of QI interventions may be produced.¹¹⁷ Second, inconsistencies in reporting can affect the process by which meta-analysis is applied. Meta-analysis synthesises evidence by combining or ‘pooling’ data and the selection of studies with a low degree of heterogeneity between them is encouraged so that data from similar studies can provide a helpful summary of results.^{117,118,119} Meta-analysis can be conducted using a number of different statistical models to examine diversity within and between studies.¹¹⁷ These models, including the fixed effect, random effect, or meta-regression models, have been applied in QI research,^{126,209} but incomplete reporting of parameters at population, intervention and outcome level can make it difficult to accurately assess studies for similarity.^{116,120} Sometimes, heterogeneity is so problematic in QI that it is not possible to apply statistical modelling for heterogeneity at all.¹⁰² If useful comparisons cannot be made between QI studies, consumers of QI work may struggle to understand whether a study should be replicated or not.

The solution of finding a suitable way to combine data, or to present heterogeneous data when meta-analysis is not possible, is developing as the scientific field of QI evolves. While meta-analysis usually uses parameters at population, intervention and outcome level, contextual features can be also regarded as important ‘influencers’ of QI interventions.^{63,210} For example, an

intervention like educational material may not work effectively without the context of a supportive group setting.⁶³ Thus, while context tends not to be used during meta-analysis, it could influence the magnitude of observed clinical heterogeneity between QI studies, and the decision of whether (and how) to combine the data.

Finally, the problem of heterogeneity in the language and terminology used in QI could create difficulties during the last stage of a systematic review, which involves dissemination of the work. When a systematic reviewer writes for publication, wide variation in reporting practices could make it difficult for them to use previously published reports to aid the write-up of their own, potentially adding to the cycle of poor reporting.¹²³

The issue of reporting bias in QI could also create difficulties for systematic reviewers. Reporting bias includes both publication bias, which tends to cause non-publication of articles with negative findings,²¹¹ and selective reporting bias, which tends to cause non-reporting of harms and outcomes within articles that are published.¹⁷⁴ Both types of reporting bias are known to be problematic in QI,^{124,126} but this thesis seeks to examine the completeness of reporting of what *is* eventually published, thus the possible reasons underlying selective reporting bias, rather than publication bias will be considered. A recent study called ORBIT,^{174,175} explains why this type of bias occurs in systematic reviews of clinical trials. Kirkham et al¹⁷⁵ evaluated a cohort of 283 Cochrane systematic reviews, and found that 55% of them did not include data for the primary outcome from all included studies.¹⁷⁵ The omission of data occurred not only due to data extraction errors, but also because the primary outcome did not occur in patients during the trial, and thus the reviewers opted not to report it even though it was documented in the original article.¹⁷⁵ Saini et al,¹⁷⁴ evaluated a cohort of 931 studies (within 92 Cochrane reviews) and found that 86% of them did not provide full data on harms because the data was incompletely reported in the original included studies.

Understanding why reporting bias occurs in systematic reviews of clinical trials is a useful starting point from which understanding can be developed about why some reporting items are frequently missing from published reports and

systematic reviews of QI. It is important to overcome the problem of reporting bias because when selected outcomes are omitted, systematic review may be more susceptible to generating inaccurate conclusions about the impact of interventions.^{211,174,212} Systematic reviewers are also warned against failing to select studies with negative results and succumbing to selection bias themselves.^{213,214}

2.5.2 Threats to successful replication

Editors and authors share a joint responsibility to ensure that published articles are written explicitly,¹¹⁰ which means the writing should be 'fully and clearly expressed; leaving nothing implied'.¹⁷⁹ This enables the reader to extract data effectively during systematic review and to replicate a successful intervention in her/his own setting.^{116,215} Attempts to replicate successful QI studies have reported mixed success, but to improve the ability to replicate, complete reporting must be the starting point. This form of reporting might be termed 'writing for replication'.

Replication is to copy or to create a 'replica', and in research, the desire to replicate successful clinical interventions began in the field of clinical trials. In research, replication can take place for two reasons. An intervention can be replicated for immediate application to clinical practice in settings other than that of the original study,²¹⁶ or it can be replicated for further repeated testing. This repeated testing can be conducted by different investigators in the same setting, which commonly occurs in basic science,²¹⁷ or in different clinical settings to explore which intervention components should be changed.²¹⁸ Thus, the desire to replicate is also associated with a drive to identify which components are more likely to succeed or fail and how they interacted to produce an outcome. These components have become known as 'active ingredients'.²¹⁹

The identification of active ingredients can be difficult and requires good descriptions from the original study authors. The Enhanced Recovery After Surgery (ERAS) literature provides a useful example of how such ingredients might be identified. ERAS is not routinely referred to as a QI intervention in the literature, but like QI interventions, ERAS pathways contain multiple elements (such as early mobilisation, intra-operative fluid balance and carbohydrate

loading) to achieve an overall effect (reducing surgical stress response to reduce length of stay) which improves an existing clinical process (perioperative care).^{220,221,222,223,224} In a study by Smart et al,²²⁵ multivariate and univariate analysis was applied to data from a consecutive series of 400 patients who underwent colorectal resection, and ERAS to identify which elements of the intervention were predictive of failure (delayed discharge).

These elements included: continued intravenous fluid infusion, lack of a functioning epidural, lack of early mobilisation, vomiting requiring nasogastric tube insertion and re-insertion of a urinary catheter.²²⁵ It is helpful to watch out for those patients who fit into one of these five categories so that remedial action can be taken early, but the ability to understand exactly what each item means and why they may occur adds to the usefulness of this data. In an earlier paper, by two of the same authors (Kennedy and Francis), data from a cohort of 196 colorectal ERAS patients were examined.²²⁶ Taking early mobilisation as an example, the authors carefully defined what they meant by 'early mobilisation' and found that paralytic ileus and analgesic failure were both associated specifically with deviation from the early mobilisation intervention.²²⁶ Thus, full intervention reporting and writing for replication could be important for creating an exact copy, as well as for the purposes of predicting failure and understanding how failure is mediated (by paralytic ileus and analgesic failure).

2.5.3 Previous examples of replication in QI

Some efforts have been made to replicate successful QI programmes although these publications are rare. As Ioannidis suggests in another context, this could be because so much emphasis is placed on innovation through new interventions, rather than replication of old ones.²²⁷ In QI, one notable replication attempt has been the Matching Michigan programme.²²⁸ This English programme attempted to replicate a successful QI project undertaken in the US (the Keystone project), which aimed to reduce bloodstream infections from central venous catheters (CVC-BSIs).²²⁹ The Keystone project reduced CVC-BSIs from a mean of 7.7 to 1.4 CVC-BSIs per 1,000 patient days with sustained improvement at three years²²⁹ and in the UK, this outcome was matched but the improvements could not be confidently attributed to the programme.²²⁸

Examining this replication attempt raises two key points. First, as Ioannidis suggests, for replication to be successful, an a priori understanding of the essential features which need to be replicated is needed.¹¹² The reporting items described in the original Keystone project could have been important to enhance decision making about whether to replicate and how to replicate. However, it is not possible to understand from the original Keystone article²²⁹ whether the authors intended to enable their readers to create an exact replica, or what the key mediators of success were, as this came in a later ex-post theory.²⁵ Understanding which ingredients should (or should not) have been faithfully replicated, and how much detail should be provided, is not yet fully realised in the literature.

Second, a 'replica' indicates an exact copy should be made, but literature exploring what happened during Matching Michigan suggests that where cultures and contexts are so different, the creation of an exact replica might not be possible or desirable.¹⁹² For example, in the UK, staff acted in an insular way, with little interaction between groups, so a collaborative community (that was important to the success of Keystone)²⁵ was hard to establish.¹⁹² This raises the question of whether QI authors should be writing to encourage exact replication, or not. Providing complete accounts of how the intervention operated in practice could allow either exact replication, or informed decision making about which elements to modify in another setting. Both may be legitimate responses to the challenges of spreading improvement, and both require the publication of a high-quality account of the original.

Whether a study is replicated, or 're-created' by modifying it to suit the local circumstances, the ability to reproduce successful outcomes could be influenced by the fidelity with which the intervention was originally delivered.²³⁰ Literature suggests that exact replication could be enhanced by documentation that the original intervention (including details such as dose, number and type of intervention components) was delivered with a high degree of fidelity over time, exactly as it was planned.²³⁰ Yet, QI research is not always implemented with a high degree of fidelity, and one of the reasons is that social, contextual and behavioural features tend to be an integral part of delivery. For example, Rycroft-Malone et al²³¹ used PDSA cycles to evaluate the implementation of

recommendations about preoperative fasting, but the fidelity of the roles of the PDSA facilitators was variable because the roles were linked to everyday activities. This level of variation may seem at odds with what high fidelity is trying to achieve: a higher degree of confidence that the study findings can be explained by an intervention which was consistently delivered.²³² However, Rycroft-Malone et al²³¹ draw on the work of Hawe et al²³³ to confront this dilemma by suggesting that standardisation and fidelity may not always be the best approach when the function and process of the intervention delivery can be examined instead. More work could be done to identify which peculiarities of QI reporting require the most attention so that any consequences of poor reporting (such as an inability to replicate) can be avoided.

2.5.4 Research waste

Writing explicitly and completely to reduce waste has been pitched as a laudable task for many years. For example, Altman and Simera¹³⁰ used a quote from Waife in 1959, to illustrate that a drive to reduce waste formed part of the rationale for the development of reporting checklists:

*“Certainly, it would be a sad waste of effort to allow reams of data to lie yellowing in a dusty file, while in other laboratories workers are unnecessarily repeating the study.”*²³⁴

Research waste is a possible consequence of poor reporting, and reducing this waste is an important ethical principle informing this thesis. During 2014 a group of researchers published a series of five articles about research waste in the medical journal, The Lancet.^{113,208,235, 236,237} The Lancet series provided the ‘human’ reasons why research should be well reported, such that the benefits derived from research should be enjoyed by everyone, but it also explained the possible financial cost of wasted research. Chan et al report that only half of the €6 billion provided by the European Union (EU) for research between 1998 and 2006 led to published research reports.¹¹³ These concerns are echoed in the surgical literature where a cross-sectional observational study of the clinicaltrials.gov database identified that one in five surgical trials were discontinued early and one in three were unpublished.²³⁸ In addition to waste by

invisibility, waste can be caused by incomplete reports which can then be deemed to be unusable.²³⁶

Much of this waste is expected to affect basic science and drug trials, which receive a heavier proportion of funding than applied health services research,^{178,208} but The Lancet series also raised issues in applied research such as poor intervention description.²³⁶ Inadequate intervention descriptions could mean that policy makers and healthcare managers may not understand how to allocate the resources needed for implementing and sustaining an intervention in a new setting. This is important because if an intervention's implementation can be planned with adequate resources, it is more likely to be fully absorbed by the implementing organisation.²³⁹ Complete reporting is also important to avoid the unnecessary duplication of research projects because research should only be conducted if the research question cannot already be answered.²³⁶ Much research waste is avoidable, and just as avoidable adverse events which occur in surgery¹⁴ should be reduced, so should the potentially avoidable errors in reporting.

The Lancet series raised the proposal that to avoid waste through invisibility or sub-optimal reporting, sustainable infrastructure²³⁶ is required. This ethos has already been embraced in surgery by the Restoring Invisible and Abandoned Trials (RIAT) initiative,²⁴⁰ which encourages the publication of all research outcomes, and the IDEAL collaborative,¹⁶¹ which encourages accurate and transparent intervention development. Events such as the fourth World Conference on Research Integrity,²⁴¹ publications such as the Nuffield bioethics series on research culture,²⁴² and the REWARD alliance (**RE**duce Research **WA**ste and **R**eward **D**iligence - <http://researchwaste.net/>) (formed by the lead authors of The Lancet series), also drive forwards a reduction in waste in healthcare more generally. Chan and colleagues¹¹³ place reporting checklists and their systematic adoption by journals, among their list of key recommendations to reduce research waste.

This literature on research waste makes a strong case for the notion that reporting should be complete, and able to explain how the intervention led to the study outcome and what resources are needed to implement it. This will

allow clinical teams to replicate successful interventions, thus contributing towards the spread and adoption of QI in perioperative care. This is vital in enabling the millions of patients undergoing major surgery each year to gain the best possible benefits from any learning from QI studies.

2.6 Researchable questions, yet to be answered

This introductory chapter has shown that QI is intended to improve perioperative care. Any publications that result from QI activity should be capable of enabling the reader to obtain a clear and accurate picture of what happened to patients and healthcare staff during the study. Yet reporting quality of QI generally (not just in perioperative care) appears sub-optimal and each type of poor reporting is described in Table 2 with an example selected from the literature to illustrate each one. The consequences of poor reporting could have a significant impact on the ability of other researchers (such as systematic reviewers), and clinicians to use the evidence generated from QI work for the benefit of patients.

Despite the fundamental role of reporting, it is not yet known which reporting items are the most problematic in the QI perioperative literature. This will be an important question to answer to enable more specific recommendations to be generated about how to move the problem of poor reporting forwards. Therefore the first research question to be answered is: “What is the current standard of reporting of QI in the perioperative literature”?

This introductory chapter has partially explained why reporting might be poor in research in general, in QI, and in perioperative care specifically. It has outlined a number of possible solutions for overcoming the problem of poor reporting that have been proposed, including reporting guidelines. These guidelines exist to enable authors to helpfully and completely describe each aspect of their research, but the literature indicates that the reporting guidelines should not be perceived as a panacea. There is a paucity of qualitative research exploring why poor reporting exists and why the problem has been persistent over many years. This gap in understanding needs to be rectified so that any suite of interventions developed to improve reporting can be fit for purpose, and tackle

the root of the problem. Thus, the second research question to be answered is: “Why is reporting QI research in the perioperative literature so hard?”

Therefore, this thesis seeks to add new knowledge to the field of perioperative QI by exposing why reporting can be difficult and exploring what needs to be done to improve reporting for a range of stakeholders who are interested in perioperative QI and are actively working in the field. The next chapter will explain how systematic review will be used to answer the first research question (“What is the current standard of reporting of QI in the perioperative literature?”) and how qualitative methods will be used to answer the second research question (“Why is reporting QI research in the perioperative literature so hard?”) Each following chapter will develop a core argument to contribute towards a whole picture of why reporting QI in the perioperative literature is hard.

Table 2 Types of poor reporting

Type of poor reporting	Mediators of poor reporting	Example from literature
1. Poor reporting by incompleteness: giving a partial description of reporting items	Lack of knowledge: Variable level of QI education among authors (not knowing what should or should not be included) Bias: reporting bias Lack of reporting guideline: checklists not endorsed; checklists not applied correctly; checklists not used at all Lack of word count: insufficient space to provide full details	Incomplete descriptions of PDSA cycles are often reported, with the number of cycles but not their length or duration. ⁸
2. Poor reporting by inaccuracy: giving an incorrect description of reporting items	Lack of knowledge: Variable level of QI education among authors (not knowing how to describe a reporting item). Lack of awareness: Unaware of existing literature	QI curricula are varied in their methodological rigour and quality. ²⁰⁷
3. Poor reporting by omission: omitting reporting items entirely	Lack of knowledge: Variable level of QI education among authors (not knowing what should or should not be included) Lack of reporting guideline use: Journals have not endorsed QI checklists; checklists not applied correctly; checklists not used at all Lack of word count: insufficient space to describe full details	A systematic review of QI, reported articles frequently failed to report missing data (75%) and generalisability (67%). ⁷
4. Poor reporting by deviance: reporting items provide a distorted view of what actually happened	Selective reporting bias	Selective outcome reporting is a common problem in QI. ¹²⁴
5. Poor reporting by invisibility: not publishing an article	Bias: Publication bias Funding: Study is shelved - ran out of funding	Lack of time can be a barrier to publishing at all. ¹⁷¹

3 METHODS

3.1 Introduction to the methods chapter

This chapter will set out and justify my research methodology, methods and approach to data analysis. I will explain why specific research design decisions were made and how the research was conducted. I will show that I have brought my own experiences to this research process and I will reflect on how this motivated the design of my research. Finally, I will explain how my research was managed in an appropriate research governance framework.

This thesis aims to answer two questions:

What is the completeness of reporting in the perioperative literature on QI methods and interventions and which elements are most frequently missing?

Why is the reporting of QI in perioperative care so hard?

To achieve the research objectives, complementary strategies were used. First, systematic review, completed in 2014, was used to describe the completeness of reporting of QI in the perioperative literature. Second, an interview study was conducted between September 2015 and March 2016 to investigate why the reporting of QI methods and interventions in the perioperative QI literature can be difficult.

3.2 Personal motives

This section begins with a statement of my personal motives. These are important to document, because they shaped the formulation of research questions and my reasons for wanting to answer them.²⁴³ The research questions were the root from which all other elements of the study were designed.⁹¹ My personal motives for completing this research lie in my experiences as a novice researcher, a clinician and a surgical patient.

The problem of poor reporting has been vexing to me as a researcher, clinician and patient. As a novice researcher I recognise the importance of producing a reproducible written account of my research, but I also recognise that writing

with the clarity required to achieve this can be challenging. As a clinician (I am a practising physiotherapist) I have ongoing concerns that research papers do not always provide an adequate description of how an intervention should be implemented so that I can recreate the same results for my patients. As a surgical patient I have observed a system, and the people working in it, for the entirety of my patient journey, on which several experiences have stimulated my interest in reporting in QI. Perhaps the most profound was experiencing long periods of perioperative fasting. I have a clear memory of feeling nauseous, thirsty and uncomfortable during long delays while I waited for a nurse to contact the surgical team for news on my status – even though the guidance is quite clear about what should happen. Such experiences prompted me to ask: ‘If researchers have identified how to improve fasting times, can they communicate their work thoroughly, explicitly, and accurately so that the clinicians looking after me would be able to use it?’

These experiences helped me to recognise the research problem, but when combined with the literature, they also helped me to identify how to approach the problem through the research design I describe here. For example, they led me to prioritise ‘context’ in my interview schedule. As a clinician, I have been immersed in the physical working world of the NHS for 18 years. The day-to-day discourse I have encountered has included topics such as resource allocation and leadership, which, in addition to other characteristics of organisational settings, have been scrutinised in the literature under the broad term ‘context’.³⁹ Reading the literature helped me to understand how the contextual factors I had seen enacted in a clinical setting (such as resource allocation and leadership) might affect how QI interventions work. For example, Delphi processes³⁹ and systematic reviews¹⁰² have been used to isolate the clinical microsystem (a group of staff working together to provide care for a population),²⁴⁴ and the functioning of the QI team as the contextual factors which most directly influence QI success. Thus, I applied the lessons I have learned from clinical practice and the literature in designing a study which is capable of finding out which contextual features are important to include in a QI report, and also which contextual features might help to explain why reporting can be difficult.

3.3 Systematic review

Systematic review was selected to answer the question: 'What is the completeness of reporting in the perioperative literature on QI methods and interventions, and which elements are most frequently missing?' Specifically, I decided to use the systematic review method to ensure that any conclusions drawn about the completeness of reporting in the perioperative literature were as reliable as possible. Reliability pertains to something which can be trusted, which is consistent in its performance.²⁴⁵ Systematic reviews overcome a range of problems associated with traditional approaches to reviewing academic literature: a failure to state a clear review question; poor description of inclusion criteria; selection bias during paper identification; reporting bias during aggregation of data; and failure to capture all relevant literature.^{117,246} The systematic review method enabled me to assemble a pool of literature which was complete and specific to the topic of QI in perioperative care, and to manage that pool of data in a reliable manner. Consequently, systematic review can be regarded as a convincing and reproducible method to inform healthcare and research delivery,²⁴⁷ including answering questions about the completeness of reporting of QI in perioperative care.

A systematic review team of clinicians and social scientists was formed with the intention of assembling a group who could understand all aspects of QI interventions as described in the perioperative literature. Clinicians could ensure that technical aspects of the intervention were correctly identified, such as the speciality in which it was operating, or the clinical objective of the intervention, such as improving compliance with venous thrombosis prophylaxis. Social scientists could help to hone thinking about the adequacy of the description of the intervention and its context.

The roles of each person in this team were clearly defined from the outset. I am an extended scope physiotherapist specialising in the care of hip and knee surgical patients, and I took responsibility for leading this team. During this systematic review I acted as first reviewer. I designed a small pilot study to test the study selection process, the search strategy and the data extraction sheet. I independently extracted data and allocated scores to all 100 papers using the checklists described in section 3.3.4. I worked closely with the second reviewer,

Dr Nicholas Lees (NJL), who is a Consultant in Anaesthesia and Critical Care at the Royal Brompton and Harefield NHS Foundation Trust. NJL and I completed both single and double data extraction together, also explained in section 3.3.4. My two supervisors, Graham Martin (GPM), Professor of Health Organisation and Policy at the University of Leicester, and Mary Dixon-Woods (MDW), Professor of Health Services Research at the University of Cambridge, helped me to revise versions of the search strategy and data extraction sheets after piloting, as well as to resolve uncertainties about which papers should be included. Each person's contribution at each stage of the review is detailed below.

The PRISMA statement was used to guide the design of each element of this systematic review to improve its value and utility.^{144,248} AMSTAR,¹²⁸ an externally validated tool¹²⁷ designed to assess the methodological quality of systematic reviews, has also been widely used across healthcare.²⁴⁹ In line with recommendations that critical appraisal guidance can be used alongside the PRISMA statement,^{146,246,250} I referred to AMSTAR to pre-empt methodological problems which could arise, and build in strategies to overcome them. For example, I designed a data extraction sheet to include a record of the full reference of each included study, and I kept an audit trail of decision making throughout the review process. To increase transparency and robustness, the systematic review protocol was registered with PROSPERO (CRD42014012845). It was also submitted for external peer review, and the protocol was published.²⁵¹ The first page of this publication is provided in Appendix 11 to form a record of publications arising from this PhD.

This review was conducted by adhering to a series of predefined, reproducible steps to identify, select, and critically appraise relevant research.^{144,247} Each step is described below.

3.3.1 Study selection: Overcoming conceptual and terminological confusion

The accurate identification and selection of relevant research was complicated by a degree of conceptual and terminological confusion over the term “quality improvement intervention” (as discussed in the preceding chapters). Poor identification of terms which have high face validity among clinicians and

academics is a recognised problem in systematic reviews of complex interventions, including QI.^{140,252} Determining what qualified as a ‘quality improvement’ study was not straightforward. Although study selection reporting does not usually include a discussion of why some papers were excluded and not others, authors are encouraged to explain when much arbitration was required to resolve disputes.¹⁴⁴ In order to provide an account of this systematic review which is reproducible, I will give a brief explanation to illustrate why uncertainties arose about which papers to include, and how they were resolved.

Many complex interventions have features in common with QI, but would not necessarily be considered to constitute a QI intervention themselves. For example, Enhanced Recovery After Surgery (ERAS) pathways combine existing knowledge (such as early mobilisation, intra-operative fluid balance and carbohydrate loading) to achieve an overall effect (reducing surgical stress response to accelerate recovery and reduce length of stay). This can then improve the quality of a clinical process (perioperative care).^{220,221,222,223,224} QI also tends to use existing knowledge to combine several interventions to improve clinical processes, although ERAS is not routinely referred to as a QI intervention in the literature. Methods of improvement (such as PDSA cycles or SPC) are sometimes referred to as interventions, yet so too are the quality interventions that such methods seek to implement. For example, the literature may use the term ‘intervention’ interchangeably to describe both the application of PDSA cycles (which I have called a *QI technique*—see below) and a quality improvement intervention (which I have called *quality intervention*—see below), such as a checklist. While initially vexing, discovery of this confusing feature of the literature did help me to realise that not all complex interventions are quality interventions and not all interventions to improve quality of care involve quality improvement.

To improve my confidence in my ability to accurately classify studies as QI or not, I adopted the suggestions of Shepherd and colleagues.²⁵² They propose that applying known taxonomies and contacting experts in the field can help to overcome difficulties in reviewing studies which evaluate complex interventions.

The taxonomy I chose was that of Shojania et al,²⁵³ who describe a classification system ranging from provider reminders to financial incentives, which is built on other well-established taxonomies of behaviour change interventions.^{254,255,256,257,258} Shojania et al's original taxonomy which identified nine QI strategies (listed in Table 3), was later amended by replacing the category organisational change with 'team changes' and adding 2 new categories (case management and electronic patient record), to make up an 11 item tool.¹²⁶ Shojania et al's original taxonomy²⁵³ was accompanied by detailed guidance notes. I opted to use Shojania et al's earlier nine-item taxonomy²⁵³ because it was succinct and easy to follow, but I modified it to create a new 11 item version of my own, which classifies the strategies numbered 1-9 in the original as examples of quality interventions, and those numbered 10 and 11 as QI techniques (Table 3).

I distinguish between *quality interventions* and *QI techniques* by defining *quality interventions* as specific changes to clinical or organisational systems. I defined *QI techniques* as the methods used to support the change, which characteristically involves a pre-defined set of steps. Thus, a reminder system for hand-washing would be classified as a *quality intervention*. Methods such as PDSA cycles, which aim to support the implementation of the reminder system, would be classified as *QI techniques*. Of note, the distinction between a quality intervention and a quality improvement technique is not hard and fast, but is rather more of a heuristic and, to some extent, context-specific. Thus, feedback is listed in the taxonomy as both an *intervention* and a *technique* because feedback can be delivered as part of a quality intervention¹⁹⁶ such as a reminder system, but it can also be delivered as part a QI technique, such as audit and feedback.

Shojania et al^{126,253} instruct their readers that some QI interventions, such as patient reminder systems, do not intervene to change practice alone and would only be counted as a QI strategy if they added another approach to organisational change such as team work or Continuous Quality Improvement (CQI). I developed this further by specifying for this review that the quality intervention should be combined with any QI technique (not just CQI) to count as a QI strategy. Therefore, to qualify for inclusion in this systematic review,

articles had to report both a quality intervention (strategies 1-9 in Table 3) *and* an associated QI technique (strategies 10-11 in Table 3).

To ensure reliable classification of articles using the taxonomy, a series of article selection training exercises were conducted among the four contributors. MDW and I jointly considered selected articles for inclusion, and discrepancies were resolved with GPM. To tighten our ability to use the taxonomy reliably, a second round of this exercise was completed where GPM and I independently considered a selection of seven articles for inclusion and discrepancies were resolved by NJL. Selection criteria were refined to ensure consensus and reliability. I also improved the reliability of article classification by studying how others had used Shojania's taxonomy in previous systematic reviews. Tricco and colleagues,² for example, used the 11-item version of Shojania's taxonomy¹²⁶ in their systematic review on the effectiveness of QI strategies on diabetes care.

Table 3 Quality Improvement Taxonomy†

	QI strategy	Definition	Example Methods	Surgical Examples
Articles reporting any QI intervention (1-9) must include one additional item (10-11).				
1	Provider reminder systems	Any 'clinical encounter-specific' information intended to prompt a clinician to recall information or consider a specific process of care	Decision aids Reminders	MEWS The WHO Surgical Safety Checklist
2	Facilitated relay of clinical data to providers	Transfer of clinical information from patients to the provider (not during a patient visit)	Telephone call Postal contact	Relay of BP measurements to the pre-assessment team
3	Provider education	Dissemination of information	Educational outreach visits Distribution of educational material Clinical guideline information	Component separation training and recurrence rates Cadaveric training and surgeon confidence.
4	Patient education	Dissemination of information	Distribution of educational material Individual or group sessions	Tri-modal pre-habilitation programmes compliance and effect on LOS

	QI strategy	Definition	Example Methods	Surgical Examples
5	Promotion of self-management	Access to a resource that enhances the patient's ability to manage their condition	BP devices Patient diaries	Follow-up phone calls with recommended adjustments to care
6	Patient reminders	Any methods of encouraging patient compliance to self-management	Appointment reminders	SMS exercise reminders before bariatric surgery
7	Organisational change	Any change in organisational structure	Multidisciplinary teams Communication Health records	Changes to staff rota to facilitate early patient mobilisation after elective arthroplasty
8	Financial, regulatory, or legislative incentives	Any financial bonus, reimbursement or provider licensure scheme	Positive or negative incentives for providers or patients.	18-week wait target for elective orthopaedic surgery
9	Feedback	Any feed-back of clinical performance	Distribution of feedback via staff education sessions or e-mails. Can occur as part of SPC or audit and feedback.	Percentage of patients achieving target LOS

	QI strategy	Definition	Example Methods	Surgical Examples
Articles reporting any QI technique (10-11).				
10	Audit and feedback	Any feedback of clinical performance summarising percentages of patients who have achieved a target outcome which has been measured at intervals over time.	PROMs LOS Morbidity and mortality	Percentage of patients achieving target LOS
11	QI methods	Systematic techniques for identifying defects in clinical systems and making improvements, typically involving process measurement and re-measurement	PDSA, Six Sigma, TQM, CQI, SPC, Lean	Improving processes for acetabular cup placement in minimally invasive hip surgery

†Adapted from Shojania et al (2004) Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 1: Volume 1—Series Overview and Methodology). Technical Reviews, Rockville (MD): Agency for Healthcare Research and Quality (US).

As a novice researcher I was initially unsure which papers reported both human factors (HF) and QI interventions, and which papers reported only HF interventions. To avoid misclassification of articles, I contacted some experts in the field who shared their views with me on the key differences. This informal coaching helped me to form a basic level of understanding which I later built on through more structured reading. I generated a short series of queries relating to specific papers where I was unsure if the intervention described a HF or a QI approach. The HF experts responded to these queries and helped me to become consistent in my classification of interventions. Developing a working relationship with this extended circle of experts also enabled me to work with them to formally record what we had learned, and we published an article which explained the similarities and differences between QI and HF.²⁵⁹ This publication can be found in Appendix 11. I also contacted the authors of papers identified by the search strategy on six occasions to ensure I had correctly classified the intervention they described. While this was helpful, recall difficulties or heterogeneity in the experiences of different authors on the same paper may have threatened the reliability of the accounts provided.

3.3.2 *Study selection: Search strategy*

The search strategy (Table 4) sought to capture terms relating to (i) surgery, (ii) quality improvement and (iii) methodology. Deciding which combination of search terms to select to ensure the capture of *all* relevant literature was difficult. There is considerable ambiguity about how terms related to QI should be applied¹²⁶ and they tend to be used inconsistently in abstracts.²⁴ To ensure I selected terms which have been commonly used to describe QI activities I obtained key terms from: abstracts of known QI papers (which were included in a previous systematic review of QI in surgery³); a published QI search strategy conducted by the Health Foundation;⁵³ and a small literature scoping exercise to explore QI terminology. The quantitative search terms were obtained from a published search strategy which is available via the Cochrane Collaboration in Naumann's document 'How to develop a search strategy for a Cochrane review.'²⁶⁰ Another small literature scoping exercise identified published search strategies using perioperative^{261,224} and qualitative^{262,263} search terms, and any terms still felt to be missing were added during discussion with my supervisors.

To reduce unwanted 'noise', free-text search terms needed to be sensitive enough to ensure appropriate papers were identified and specific enough to ensure that inappropriate papers were not. Broad free-text search terms were used, for example 'quality adj2 improvement' captured titles such as 'Can quality circles improve hospital-acquired infection control?',²⁶⁴ which would not otherwise be identified by 'quality improvement.' The syntax 'adj2' is used in MEDLINE to capture words within two words of each other, in any order. However, this also resulted in the capture of articles which described quality of care in general, rather than quality improvement specifically, and so the search strategy suffered from poor specificity. I therefore also used Medical Subject Headings (MeSH) terms—controlled vocabulary used for article indexing that can help to overcome problems with specificity²⁶⁵—such as 'General Surgery/' and 'Operative Time/'. This search was comprehensive because it used a combination of keywords, subject headings (sh), multi-purpose words (mp), publication type (pt) and Medical Subject Headings (MeSH). During search strategy pilot tests, a selection of known surgical QI papers³ was successfully captured using the search strategy in Table 4.

Table 4 Search Strategy

Combine column A and column B using AND, then combine using AND to either column C or column D			
A: QI term family	B: Perioperative term family	C: Qualitative term family	D: Quantitative term family
Deming.ti,ab 6 sigma.mp. (Six adj1 sigma).mp. (Lean adj1 sigma).mp. Measurement for improvement.ti,ab. (quality adj2 improv\$).ti,ab. (quality adj1 management).ti,ab (improv\$ adj2 science).ti,ab. (process adj2 improv\$).ti,ab. (Plan and do and study).ti,ab. PDCA.ti,ab. pdsa.ti,ab “plan do check”.ti,ab. (method adj2 improv\$).ti,ab.	preoperative.mp. Per-operative.mp. surg\$.ti,ab. exp General Surgery/ exp Surgical Procedures/ Operative/ or Operative Time/ Perioperative.ti,ab. Perioperative.ti,ab. Pre-operative.ti,ab. Operative.ti,ab. Cancer\$.mp. “enhanced recovery”.ti,ab eras.ti,ab. “rapid recovery”.ti,ab. (fast.mp. AND track.ti,ab.) Operating theatre.mp.	Qualitative.ti,ab. (Focus Groups.mp. OR exp Focus Groups/) (exp Interview/ OR interview.mp.) (Interviews as Topic.mp. OR exp Interviews as Topic/ ethnograph\$.mp. content analysis.mp. grounded theory.mp. grounded approach.mp. (exp Qualitative Research/ or qualitative.mp.) Phenomenology\$.mp. “discourse analysis”.mp. “constant comparison”.mp.	Randomized.ti,ab “Randomi?ed controlled trial”.pt. “Randomized controlled trial”.pt “controlled clinical trial”.pt. “randomized controlled trials”.sh. “random allocation”.sh. “double blind method”.sh. “single-blind method”.sh. single?blind method.sh clinical trial.pt. exp clinical trial/ (clin\$ adj25 trial\$).ti,ab.

Combine column A and column B using AND, then combine using AND to either column C or column D			
A: QI term family	B: Perioperative term family	C: Qualitative term family	D: Quantitative term family
(health.mp. AND behavioural economics.ti,ab) "operations research".ti,ab. "decision science".ti,ab. Shewhart.ti,ab. Pareto?chart.ti,ab "Statistical process control".mp. "Statistical quality control".mp. Toyota.mp Paretochart\$.mp "Control chart".mp	"Operating room\$".mp Anesthe\$ Anaesthe\$ Trauma.mp	"observational method".mp "theoretical sampl\$".mp. "thematic analys?s".mp. "improvement report\$".mp.	((singl\$ or doubl\$ or trebl\$ or tripl\$) adj25 (blind\$ or mask\$)).ti,ab. placebos.sh. placebo\$.ti,ab. "comparative study".sh. exp evaluation studies/ prospective studies.sh. (control\$ or prospective\$ or volunteer\$ or retrospective\$).ti,ab. "Interrupted time?series".ti,ab Time-series.ti,ab "Repeated measure\$".mp Cohort.mp "Case?control".mp "Case control".mp

Combine column A and column B using AND, then combine using AND to either column C or column D			
A: QI term family	B: Perioperative term family	C: Qualitative term family	D: Quantitative term family
			(Assess\$ adj3 process quality.)mp Evaluat\$ adj3 process quality).mp Compliance.ti,ab "Quality control".mp Process control.ti,ab Control chart\$.ti,ab (Adherence OR adherence.ti,ab) Group adj3compar\$.mp Control.ti,ab Before adj2 after.mp Eval\$.ti,ab Variability OR variation OR variable.mp

Bibliographic databases were selected for their representation of both surgical and QI literature. The Ovid SP version of MEDLINE, and Scopus was searched on 28 May 2014. The Cochrane Central Register of Controlled Trials (CENTRAL) was searched on 28 May 2014 using the terms ‘quality improvement’, ‘quality improvement in surgery’ applying the limit of ‘trials.’ CENTRAL was also used to search for records registered by the Cochrane Effective Practice and Organisation of Care (EPOC) group (which indexes interventional studies focused on improvement in health care delivery). The “related articles” function of PubMed was also searched. The restrictions of publication year (2000-2014), humans NOT animals, and NOT infants were applied.

The search results were supplemented with hand-searching of the reference lists of full-text articles and an existing systematic review by Nicolay et al on the application of quality improvement methodologies to surgery.³ Papers included in the systematic review by Nicolay et al³ were included in my systematic review, unless they were published prior to 1 January 2000, or included paediatric cases.

3.3.3 *Study selection: Eligibility criteria*

This systematic review sought to include:

- All studies published (including those published online ahead of print) between 1 January 2000 and 28 May 2014, so as to capture articles indexed since the publication of the Institute of Medicine’s “To Err is Human: Building a Safer Health System” report.⁵⁷
- Studies reporting a deliberate effort to produce change in active perioperative care and meeting the criteria for QI as specified in a modified version of a taxonomy generated by Shojania et al²⁵³ (Table 3).
- All surgical specialties.
- Adult surgical care.
- Elective and emergency (trauma) surgery.
- Primary and secondary care.
- Qualitative and quantitative literature.

Exclusion criteria included:

- Clinical audits, unless they explicitly reported on the implementation of a QI technique designed to produce and evaluate a change.
- Qualitative papers reporting exclusively on staff or patient experience of QI.
- Papers reporting on screening and diagnostic techniques such as endoscopy.
- Papers reporting on end-of-life care.
- Papers not published in English.
- Abstracts, conference proceedings and editorials.
- Cadaver studies.
- Paediatric studies.

I acted as the first reviewer and after duplicates had been removed, I independently assessed and then excluded titles and abstracts of the first 13,603 papers. I set aside the remaining 1,115 papers because I could not make a decision alone on which ones to select. Two reviewers (NJL and I) independently assessed titles and abstracts of these 1,115 papers to reach agreement on the selection of the final cohort of 100 full text articles. Disagreements about eligibility were resolved by discussion with a third reviewer (GPM or MDW).

3.3.4 Data extraction

A training exercise was completed where NJL and I independently extracted data from nine published surgical *BMJ Quality Improvement Reports*. Although the journal *BMJ Quality Improvement Reports* was not included in the final review (because this journal was not PubMed indexed at the time of writing), it was useful to include it in a training exercise to ensure each author applied the Template for Intervention Description and Replication (TIDieR) checklist consistently.

For the actual review, data from the 100 articles that met the inclusion criteria were extracted onto a standardised Microsoft Excel template based on items from the TIDieR checklist and a checklist containing a small number of

additional features relevant to QI (Table 5). To enhance rigour, double data extraction was independently carried out by two reviewers (me and NJL) for the first 12 included papers using the data extraction template.²⁶⁶ Each reviewer independently scored these papers using the checklists described above. Then a consensus meeting was held to discuss the scores. The level of agreement between us was high. All scoring on 10 of these 12 papers was agreed at the first consensus meeting. Two of these 10 papers were given identical scores. The largest variation in scoring was 4 points and this occurred in two of the 10 papers. At the second meeting, the remaining two papers, for which the TIDieR scores allocated by NJL and me differed by two points, were discussed and a consensus was reached. I then conducted single data extraction for the remaining 88 included papers. NJL reviewed my completed scores and verified the data extraction. Discrepancies that could not be resolved were then discussed with GPM or MDW. This combination of double and single data extraction improved confidence that the frequency of error could be reduced. It also allowed for pragmatism because the average time spent for single extraction is significantly less than double extraction.²⁶⁶

Table 5 Data extraction template items

	Additional features	Quality Intervention (TIDieR parameters)	QI Technique
1.	Author	Brief name	Name of QI technique
2.	Year	Why: rationale for intervention	Baseline measurement
3.	Country	What: materials used to apply the intervention	Data collection schedule
4.	Surgical speciality	What: procedures processes used in the intervention	Data analysis (e.g. driver diagrams)
5.	Setting	Who: provided the intervention, including level of training	Data volume/duration (e.g. length of PDSA cycle)
6.	Use of PROMS	How: mode of delivery: (e.g. face to face, internet)	Description of prediction of change
7.	Use of SQUIRE	Where (location: emergency/elective, primary/secondary)	Missing data (and reasons given)
8.	Adverse events	When and how much: duration, dose, intensity	Description of generalisability
9.	Presence and type of PPI	Modifications to intervention during the study	Named primary outcome
10.		How well (planned): strategies to improve/ maintain intervention compliance	
11.		How well (actual): the extent to which the intervention was delivered as designed	

3.3.4.1 Use of the TIDieR checklist to assess reporting of quality interventions

The TIDieR checklist¹⁵⁰ contains 12 items relating to intervention reporting criteria. TIDieR is recommended by the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network as an extension of the Consolidated Standards of Reporting Trials (CONSORT)¹⁴¹ and SPIRIT¹⁴⁸ statements to improve reporting across all ‘evaluative’ study designs. The QI interventions (items 1-9 in Table 3) identified in this systematic review were scored using a modified version of the TIDieR checklist (Table 5).¹³⁸ Item 9 (tailoring: *personalization or titration of the intervention*) was removed because the interventions studied were not titrated for individual patients. This resulted in a total of 11 TIDieR checklist items, and the maximum score that could be obtained by any article was 11/11. Scoring was guided by the TIDieR group’s explanatory statement,¹⁵⁰ further clarified through email correspondence with TIDieR’s first author, Dr Tammy Hoffmann.

Articles were scored as “Yes” for each item that could be assessed as reported in full. If the description was unclear or if no description was given, the article was scored as “No” for that item. For example, where a paper clearly described the modifications made to an intervention in a manner judged to be fully explicit, it was rated “Yes” under the TIDieR item ‘modification’ (item 9 in column 2 of Table 5). An example of a paper achieving “Yes” under this criterion described the modifications made as follows: “After multiple trials of various insulin protocols, a simplified high-infusion protocol replaced the low-infusion protocol with intermittent boluses” (p.25)²⁶⁷ Many papers reported on multiple interventions, for example a safety bundle.²⁶⁸ In order to accurately replicate a multifaceted programme, all of its components must be fully described. Therefore, each article was scored once against each TIDieR item, regardless of number of interventions, and was only rated ‘Yes’ if all components reached the required standard.

3.3.4.2 Use of the checklist of data extraction items (Table 5) to assess reporting of QI techniques

QI techniques (items 10-11 in Table 3) reported in articles were scored using a checklist of relevant items (Table 5). The checklist items were based on the EPOC review group's data collection checklist,¹⁵² which has previously been used in systematic reviews to consider reporting features specific to QI measurement.¹²⁹ Relevant items from the EPOC checklist were selected for our checklist, including baseline measurement, data collection schedule, data analysis, missing data, and named outcomes. A further item relating to data volume/duration was added in response to a recent publication by Taylor et al⁸ on the reporting of PDSA cycles.

3.3.4.3 Use of the checklist of data extraction items (Table 5) to assess additional reporting features

As well as the TIDieR checklist and the checklist of QI techniques, the data extraction template included items relating to reporting of: patient and public involvement (PPI); adverse events; patient reported outcome measures (PROMS); and use of the Standards for QUality Improvement Reporting Excellence (SQUIRE) statement²⁶⁹ (Table 5). PPI was defined as the incorporation of the knowledge, skills and experience of patients, informal caregivers, and the public into a study²⁷⁰ and it was included because it is encouraged across all types of surgical interventional studies.²⁷¹ An adverse event was defined as any unfavourable or unintended sign, symptom, or event associated with the intervention; reporting of such events is important to enable the full understanding of possible benefits and harms of interventions.^{186,272} The SQUIRE statement²³ is a reporting guideline for QI studies, and is recommended by the EQUATOR Network so mention of it might be considered a feature of high-quality studies.

3.3.4.4 Absent reporting features

In view of best practice recommendations produced by the York Centre for Reviews and Dissemination¹¹⁷ I provide an honest statement of what was *not* reviewed. Methodological flaws and risk of bias were not examined because the

review did not focus on intervention effect. Bias is a term that will be used frequently throughout this thesis and bias can be defined as the result of factors that cause a systematic (non-random) departure from the true results in data analysis.²⁴⁵

3.3.5 *Data analysis*

Data were analysed descriptively using an Excel data extraction sheet. This extraction sheet displayed textual descriptions of interventions. I used the four steps suggested by Popay et al²⁷³ to guide how I organised the findings. In the first step, existing theory on the benefits of complete reporting was examined in a literature scoping exercise (see Chapter 2). Second, the tabular format of the extraction sheet gave an overview of the main themes of intervention description (the TIDieR items), QI technique description, and the additional reporting items. A vote-counting system was used as described by Popay et al²⁷³ to identify common measures across all the included studies, such as the number of incomplete TIDieR items. For example, if the average TIDieR score across all 100 included papers was found to be 6.3/11, that would mean that, on average, 43% of the recommended TIDieR reporting items were incompletely described or omitted. Arguments have been posed against this type of 'counting',²⁷⁴ but I found that it gave a helpful overall description of the data. Third, relationships were examined between the studies by producing a visual representation of which features were shared across the included studies. The robustness of each included study was not evaluated because the review did not seek to draw conclusions about the interventions' effect.

3.4 **Qualitative work**

My systematic review on how well QI is described in the perioperative literature revealed that reporting is sub-optimal (see Chapter 4). The most frequent incomplete items were intervention fidelity and intervention modification. The systematic review did not identify *why* reporting was sub-optimal, and why some items were incomplete more frequently than others, however. I therefore sought to explore this in my second study. This involved making choices about study design. This section will demonstrate why one study design (a Delphi study to

design an adapted version of TIDieR) was rejected, in favour of another (the qualitative work) to explain why reporting is poor.

At first, it seemed feasible that fidelity and intervention modification reporting is most often incomplete because the TIDieR checklist (see section 3.3.4) might not be suited to the reporting of QI interventions, which typically seek to actively respond to local circumstances by modifying the intervention components as the study progresses. This raises challenges of how to assess fidelity. In many fields, fidelity involves adhering closely to the study protocol and pre-specified intervention,^{245,275,276} because delivering the intervention in the same way can increase the scientific confidence that observed differences between groups are due to the intervention.^{230,232} Thus, in a clinical trial the treatment dose (including the length of the intervention time period and the expected minimum and maximum range of the dose) may be pre-specified.²³² This dose (or dose range) is typically adhered to throughout the study because in a trial a protocol is followed. However, QI interventions may require a different approach. Issues in QI work which relate to fidelity and modification will be presented here to fully explain why a qualitative study design was eventually selected.

In a QI project, or a study of QI, it is not usual to adhere to a strict protocol for the delivery of the intervention.²⁶⁹ Instead, the intervention can change as the study progresses through several iterative cycles of measurement. In some QI studies, a high degree of fidelity may not be possible at all, due to many interacting social and behavioural elements. In others, fidelity may be high during the first period of measurement. Once the QI measurements have been fed back to the QI team, the intervention may be modified in response to the data. Then, when measurement resumes, it may be beneficial to retain intervention fidelity again for the new, modified intervention. For example, a reminder system for the use of hand sanitising gel may be implemented during PDSA cycle 1 using one audible prompt on entering the ward. Following a period of measurement, it might be amended (say, by changing the number of audible prompts at different time points). While the dose has changed (and therefore by the standards of traditional clinical trials, fidelity is reduced), the intervention remains faithful to its intended ethos of reducing infection. Fidelity in QI is thus less a matter of adhering precisely to exacting prescriptions

specified in protocols, and more a matter of ensuring that an intervention is fully operationalised as it was planned during each PDSA cycle (and this may mean qualitative and quantitative changes in the intervention from one cycle to another).

Similarly, the TIDieR checklist also encourages documentation of unforeseen modification, and of modifications which occur between protocol submission and publication of the primary paper. This has been referred to in clinical trial guidance as a 'protocol deviation.'^{114,277} Modification reporting is also important in QI, but again, the way it is reported in QI may be dramatically different. QI can involve modifying the intervention several times during the study because QI is intended to be flexible, and change according to the needs of the patient.⁸⁷ These issues led to my asking whether the TIDieR checklist was suitable for use as a reporting tool in QI, and whether it should be modified to suit the distinctive features of QI research.

At one point, in discussion with my supervisors, we considered the possibility of a Delphi process to design an adapted version of TIDieR. A Delphi process is recommended to generate consensus on items during the development of guidance and reporting tools.¹⁵⁷ However, jumping straight to checklist development created a feeling of unease. Although the systematic review²⁵¹ had revealed that reporting was sub-optimal (see Chapter 4 for results), it was not yet understood *why* reporting is so poor, what constituted the defining characteristics of a QI intervention, and what authors (and journals) needed to support better intervention and QI technique reporting. I did not have any evidence to support a hypothesis that adapting the TIDieR checklist would affect the standard of reporting for the better. The risk, then, would be jumping to a cure before fully understanding the nature of the pathology. Therefore a Delphi exercise was rejected as premature, in favour of a method capable of exploring why reporting is poor and what should be done to facilitate improved reporting (and thus laying the foundations for future work intended to improve reporting standards, whether via a checklist or through other means). Improving reporting is a laudable future research aim because incomplete intervention description threatens the reproducibility of surgical QI work, which may then

produce several other unwanted consequences. For these reasons, I chose to undertake a qualitative study.

3.4.1 Qualitative study objectives

The qualitative study aimed to explore the question: “Why is the reporting of QI in perioperative care so hard?”

The specific study objectives were to:

- Describe the reasons why reporting QI interventions and QI techniques in the perioperative QI literature can be difficult.
- Explore what counts as a QI intervention.
- Inform wider debate about how QI reporting would be best supported.

The next section will explain the design of this study, which involved semi-structured interviews and an author-checking exercise to assess concordance between readers’ interpretations of a published QI paper and the authors’ intended meaning.

3.4.2 Advisory panel

An advisory panel was formed to help ensure this study retained relevance to people with an interest in QI reporting activity (stakeholders) (see Table 6). Panel members met with me separately roughly twice a year for a telephone or face-to-face meeting for approximately an hour. Each panel member could withdraw from the advisory panel at any time.

Table 6 Advisory panel members

Name	Role
Jo Howarth	Patient safety and quality lead, Yeovil District Hospital NHS Foundation Trust.
Rachel Johns	Anaesthetist, Yeovil District Hospital NHS Foundation Trust.
Gill Penny	A patient with experience of cardiac surgical care (including intra-operative never event which include: wrong site surgery, wrong implant, or retained foreign object). ²⁷⁸
Louise Davies	An Associate Professor of Surgery (Otolaryngology) at the White River Junction VA Medical Centre, VT, US, and a senior scholar at the Dartmouth Institute for Health Policy and Clinical Practice, US.
Gretel Stonebridge	Working with the University of Leicester change and improvement team, Gretel has expertise in applying 'systems thinking' in industry.

Advisory panel meetings were structured around what was most pressing at the time the meeting was called. For example, one later meeting in the study allowed Gill (the patient member for the panel) and me to work as partners, reading a selection of excerpts from transcripts together and extracting themes. The panel was helpful in encouraging a culture of embedding new ideas and ways of thinking into the study as they arose. Different levels of involvement for advisory panels have been recommended ranging from consultation to equal partnership.²⁷⁹ In this study, panel members offered opinions on the utility and appropriateness of study documents such as information sheets and the interview schedule. They also commented on whether the study was being planned and conducted in a manner which was relevant and acceptable to QI stakeholders. The panel helped to keep me focused on one of the intended benefits of reporting, which is to produce improvement in care for patients. This extra focus on how my study might add value to the existing pool of research on QI reporting was very motivating for me. They worked collaboratively with me to

resolve problems with jargon, so that any recommendations produced by this study could be readily understood by different audiences. They also worked collaboratively to help with recruitment.

The literature provided insight into the optimal methods for involving an advisory panel.²⁷⁹ The governance and management of the advisory panel was supported by:

- Providing the panel with printed copies of a role description, written with them to define their roles and time commitment to the study.
- Building in 10 minutes to the start of each meeting to identify any problems with the way the panel is conducted.
- Keeping a reflective log to document any critical incidents.
- Asking the PhD supervisory team to assist with face-to-face meetings if necessary, to ensure that the advisory panel retained relevance and applicability throughout the study.

3.4.3 Telephone interviews

Interviews can be used to encourage participants to share their experiences and their views about what is important to them.²⁸⁰ Generating data in this way was well matched to my research objective of increasing knowledge about why reporting QI interventions and techniques can be hard and how reporting practice could be better supported. The interviews were conducted by telephone because I aimed to recruit participants from varied geographical locations and many were based abroad. One-to-one interviews facilitated discussion on a subject which required some concentration.

3.4.4 Inclusion criteria

Participants were eligible for inclusion in the telephone interview study if they were:

- Willing and able to give informed consent for participation in the study.
- Male or female, aged 18 years or above.
- A QI author, QI consumer, or QI gatekeeper (see definition in section 3.4.5 below).

3.4.5 *Definition of recruitment groups*

The starting point for this part of the study was the question of *why* reporting is hard in the perioperative QI literature and *how* should it be improved. Groups with an interest in reporting activity (stakeholders) were defined at the 2015 EQUATOR and REWARD conference to include: researchers, research organisations, funders, editors, publishers, communications experts and research users.²⁸¹ These groups have previously been described in the literature (which defines ‘research users’ as clinicians, systematic reviewers, patients and policy makers)²³⁶ and in the UK’s Concordat to support research integrity.¹¹⁰ Accordingly, the stakeholders were included in this study population if they wrote up perioperative QI work for publication (termed QI authors); read perioperative/surgical QI literature in order to change care (QI consumers); or made decisions about publishing QI literature (QI gatekeepers).

The stakeholder groups are further defined as follows:

QI authors: This group included individuals who were involved in the development of QI approaches and projects, and who reported their QI work in the published literature. The main qualifying criteria were, first, that they were listed as an author on a paper reporting QI in perioperative care which has been published in a PubMed-indexed journal. Second, authors were recruited if they had published a paper since the production of the Institute of Medicine “To Err is Human: Building a Safer Health System”⁵⁷ report in 2000. This report is thought to have triggered a much higher volume of original patient safety and systems-based research articles.⁵⁹ Consequently, a study population of QI authors writing papers after 2000 would provide insight from individuals who have contributed to a growing specific field of quality and safety literature.

QI consumers: This group included managerial and clinical healthcare staff who read QI reports and use them to inform changes in the delivery of perioperative care. Managerial staff included Directors and Associate Directors of Patient Safety and Quality or Quality Improvement Chief Operating Officers who have used QI reports to plan their improvement strategies. Clinical staff had a range of clinical roles – nursing, medical, physiotherapy, and so on – and

had used QI reports to improve their service. This group were using QI reports to improve their service in the 24 months prior to interview.

QI gatekeepers: This group included individuals who have conducted gatekeeping activities, particularly in relation to the reporting and publication of QI projects and approaches. These activities included making decisions about how QI should be described in perioperative QI reports (reporting guideline authors) and about which perioperative QI papers should be published (journal editors). In the findings chapters, (chapters 5, 6, 7 and 8), the reporting guideline authors are referred to as ‘influencers’ to differentiate influencers from editors. These people were not required to have a specific interest in surgery, but they were interested in promoting complete reporting in surgery and perioperative care as part of a requirement to safeguard high standards of reporting in healthcare more generally. This group had undertaken this QI gatekeeping activity in the 24 months prior to interview.

Editorial team members were selected only if they had a direct role in managing the scientific content of the journal such as handling manuscripts and providing strategic guidance.^{104,282} Defining this study population was complicated by a wide variation in editorial job titles which I initially struggled to understand. To overcome this knowledge gap, I discussed variability in editorial roles with senior colleagues. Literature such as the Council of Science Editors White Paper on promoting integrity in scientific journal publications also provided clarification on role definitions.²⁸² I realised that editorial teams commonly consist of individuals in a hierarchical structure. Job titles for individuals most directly responsible for scientific content include: senior editors (sometimes known as editors-in-chief, or chief editors), deputy editors (sometimes known as editor of scholarly content, or section editor), and associate editors. Members of the editorial team mostly responsible for dictating the overall tone of editorial policy (sometimes known as editorial board members) or for the administration of the journal (sometimes known as managing editors or editorial assistant) were not part of the defined study population.

Participants were grouped according to their predominant professional role, but some participants had parallel current interests which aligned with a second

group. For example, QI authors also mentioned a need to ‘consume’ the QI writing of others. Some participants had experience from previous roles which they could draw upon during interview. For example, a reporting guideline author also held extensive previous editorial experience, although he had not held a senior, deputy, or associate editor position during the 24 months prior to his participation in this study. While the varied experiences of individuals in each group could threaten between-group comparisons, complete standardisation was not desirable for this study. The participants do not exist in a vacuum, and their varied QI-related activities add depth to the data.

These three groups (authors, consumers and gatekeepers) were chosen to provide deeper insight into how reports are written and what helps or hinders their use. I hoped that three groups would provide a diverse range of perspectives, but a possible limitation is that not every type of reporting stakeholder was included. In particular, patients were not recruited.

Literature has supported patient and public involvement (PPI) in the conception, design, delivery, and write up of scientific research.^{283,284} PPI has been encouraged in the *delivery* of QI work by UK organisations such as NHS Improvement,²⁸⁵ and international organisations such as the Institute for Healthcare Improvement (IHI) in Boston, MA, US.²⁸⁶ The GRIPP (Guidance for Reporting Involvement of Patients and Public) checklist²⁸⁷ is included in the EQUATOR Network’s library of reporting tools to prompt researchers to report their PPI activity. However, patient involvement in QI reporting has not yet been widely observed in the QI literature and reporting PPI activity in surgical research is known to be poor.²⁷¹ Also, researchers may need more time to optimally implement the new SQUIRE guidelines,¹⁶⁵ (published in 2008¹⁴⁹ and 2015²³) and improve the reporting standards of QI itself,²⁸⁸ before attempts are made to add PPI to the reporting requirements. For these reasons, patients were not defined as key stakeholders in surgical QI reporting. This was agreed with the patient member of the study advisory panel, who also added that the interview schedule was likely to involve technical language relating to methodology, which might be explored more easily by participants who were experienced in QI reporting.

3.4.6 Participant identification, eligibility assessment and recruitment

A sampling frame is a list from which the sample can be selected.²⁸⁹ The sampling frame for this study included individuals who were active in the fields of higher education, publishing, research funding or healthcare. These individuals were working in organisations such as The Health Foundation, IHI, universities, Collaborations for Leadership in Applied Health Research and Care (CLAHRCs) and hospitals. As described above, the sample was stratified according to author, consumer or gatekeeper, and these groups were defined in advance.²⁹⁰

Purposive non-probability sampling methods were used.²⁴⁵ The use of pre-defined study population characteristics to identify suitability for selection reduced the risk of selection bias (section 3.4.5). Participants were deliberately selected with the expectation that they would contribute to answering the study question: “Why is reporting so hard?” They were chosen purposively and their characteristics (such as job role and publication track record) ensured their participation in in-depth interviews could produce relevant subject matter. Snowball sampling was included in the study protocol as a complementary non-probability sampling method. It involves asking recruited participants for names of others who might be interested in taking part.²⁹¹ As the study progressed, however, it became clear that a large sampling frame could be obtained without using ‘snowballing.’ This reduced any potential further burden on the participants enrolled into the study because I did not need to ask them for any additional information.

To satisfy a requirement to recruit a sufficient number of QI consumers, and to represent consumers from different healthcare settings, four NHS trusts were accessed as recruitment sites. Four was deemed an adequate number because these trusts supplemented a wide spread of recruitment across non-NHS organisations and they provided adequate breadth of hospital types. NHS employees were only recruited at these four sites. The four organisations were purposefully sampled to capture a spread of views using the following criteria: trust size (measured by number of beds);²⁹² type of trust (Foundation or non-Foundation); number of manuscripts published in *BMJ Quality Improvement*

Reports with authors affiliated to the trust in question; and the type of QI activities undertaken. All four trusts were actively involved in implementing QI projects and were signed up to NHS England's 'Sign up to Safety' campaign, which commits them to turn their improvement aspirations into a concrete safety improvement plan.

I identified, approached, screened (checking that the participants' demographics matched the inclusion criteria) and recruited potential participants using a range of techniques. No incentives were used.

A poster was designed to advertise the study. It was e-mailed directly to individuals who were known to the study team through word of mouth, personal contacts, or their QI publications. This e-mail included a standard letter of invitation which introduced me and the nature of the study, and invited them to contact me (Appendix 2). The letter, which included a participant information sheet (Appendix 3), was approved by the University of Leicester (UoL) as research sponsor, and the UoL's Research Ethics Committee (REC). This letter of invitation was sent to:

- The QI authors whose papers I had included in the systematic review²⁸⁸ (Appendix 9). These authors were approached using the correspondence details given on the publication. QI authors were also recruited through personal contact with the Leicestershire Improvement, Innovation and Patient Safety unit (LIIPS).
- QI consumers, such as quality improvement leads and directors of patient safety, who were recruited directly using personal contact details or through contact details on publications which are available in the public domain.
- QI gatekeepers, who were approached directly from: *BMJ Quality and Safety* (<http://qualitysafety.bmj.com/>), the *International Journal of Quality in Healthcare* (<http://intqhc.oxfordjournals.org/>), the TIDieR research group (Centre for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine, Bond University, Queensland, Australia, 4229), the SQUIRE group (<http://www.squire-statement.org/>), and The Health Foundation (UK) (<http://www.health.org.uk>). Professionals from journals or universities were approached using their work e-mail addresses.

Individuals not known to the study team were also recruited through the method described above. A poster was placed on the SQUIRE guidelines home page <http://tinyurl.com/zav5d9v> . A telephone number and e-mail address was provided so that individuals who had not been approached directly, but are active in QI research, could contact this study team. Participants were able to register their interest using an online registration form at the study website: www.le.ac.uk/qi-project-emma-jones

Recruitment began in two phases. UoL sponsorship, UoL indemnity insurance and UoL REC approval were received by 18 September 2015. From this point recruitment of non-NHS participants commenced. When Research and Development (R&D) approval was gained for each site, recruitment of NHS staff could begin at that site. R&D approvals were gained between 19 October and 6 November 2015.

3.4.7 Number of participants

I estimated that a sample size of 30-40 would be sufficient for this study. This number was estimated as a 'best guess' of how many participants would be needed before no new themes arose from the data. A sample size of 30-40 has previously been shown to be enough to reach theoretical saturation in similar studies where telephone interview data about QI was analysed.^{88,293} In order to ensure the recruitment (and retention) of at least this number of people, the sampling frame (section 3.4.6) for this study included twice the number of people I thought would be required. Approaching a larger number of people ensured that even after some people had declined, and some participants changed their mind and withdrew later, a sufficient pool of participants remained to participate in the study.

Theoretical saturation is described by Charmaz²⁹⁰ as the point at which theoretical insights no longer emerge by analysing new data. Theoretic insights do not mean coded items, but patterns in the data. Patterns are generated by looking for relationships across and within sampling groups. They are then used to build an understanding of the study questions.^{290,294} During data analysis I recorded the point at which the same patterns consistently emerged to monitor

when concepts reached theoretical saturation. 'Keeping a close eye' on saturation in this way helped to inform me when data collection could stop.

Meetings with supervisors were a safeguard against assuming too early that saturation had been reached because an open dialogue about the data was encouraged. For example, the feeling that some journal editors struggled to understand the nature of QI was identified across the first seven transcripts. Rather than claiming saturation, further analysis was needed to explore what the 'nature of QI' means, and how editors were perceived to enact a lack of understanding.

It was hoped that roughly equal numbers of participants would be recruited in each of the three groups. Participants were not selected from specific surgical specialities but it was expected that at least one clinician from general surgery (breast, colorectal, endocrine, upper gastro-intestinal, and transplant) and one from orthopaedics would be interviewed, because these are the two most common types of consultant surgeon in the UK^{295,296} and the second and third most common respectively in the US.²⁹⁷

3.4.8 Participant interview schedule

I interviewed each participant once and each interview lasted between 30 minutes and one hour. Interviews began on 30 September 2015 and were completed on 18 March 2016. Each interview consisted of two parts, an interview and a QI report checking exercise.

The interview schedule (Appendix 4) design was informed by reflecting on the results of the systematic review, and through discussion with the advisory panel (see section 3.4.2 above). The interview schedule was 'tested' for ease of use by practising interviews with colleagues in the Social science APPLIED to Healthcare Improvement REsearch (SAPPHIRE) group at the University of Leicester. These mock interviews allowed me to ensure that my questions were easily understandable and that interviews could be kept to an appropriate length.

The interview involved asking the participants *why* QI intervention and QI technique reporting might be hard. An opening question about general

experiences of QI gave each participant time to speak without interruption. This eased them into the interview and gave useful further demographic information. The interview schedule was designed to elicit participants' experiences of writing or reading QI reports, and why they thought reporting might be difficult.

The interview also involved asking participants to read excerpts from a specific QI report by McCulloch et al.²⁶⁸ This report was selected because, first, it met the inclusion criteria for our systematic review of QI in perioperative care. Second, this paper scored 7/11 using the TIDieR checklist in the systematic review, which was the most common (modal) score and therefore represents the level of reporting which participants are most likely to encounter. Third, it used PDSA cycles, which our systematic review identified to be one of the most commonly reported QI techniques in the perioperative literature. This report is, therefore, in many ways a typical example of the existing QI literature in surgery.

Some of the text was highlighted to draw participants' attention to the sections most directly related to the intervention; the paper was also anonymised to avoid the participants being influenced by recognisable names or organisations. The first page of this report is included (Appendix 5) to illustrate how it was highlighted and anonymised. Each participant was given this paper two weeks before their interview. If participants had not read the paper ahead of the interview, time was allowed during the interview. Participants were asked to describe whether they felt they could, in their own settings, re-create the intervention, thus encouraging more talk about replication. In addition to describing a specific intervention, the participants were also encouraged to elaborate on what they thought about the reporting within this article more generally, with questions such as: what were the most helpful things about how QI was reported which might enable you to replicate what they did; what were the main problems with how QI was reported; and what could be done to make reporting better? (Appendix 4). This enabled the intervention description task to act as a prompt for further discussion about intervention reporting.

At the end of the interview participants were asked if they had anything they wished to add which had not yet been said. This helped to ensure that important

themes were not missed. Participants were also asked if they wished to receive copies of any published manuscripts produced as a result of this study.

3.4.9 *Author-checking exercise*

Participants in the telephone interview study were asked specifically for consent for anonymised sections of their transcripts to be shared with the authors of the McCulloch et al QI report described above. I explained to each participant that this would provide insight into the degree of mismatch between how they understood the intervention described in the paper and how the authors of the paper felt it had actually been designed and delivered. Exposing the extent to which the authors' and participants' views aligned was intended to highlight whether participants correctly interpreted the paper and if not, which parts were misunderstood. It would be impossible to conclusively establish this without having the authors of the report clarify its intended meaning. I term this process 'author-checking'. Checking whether documented events or items match what is actually perceived in practice has been used in previous research as a validation method for designing checklists for safety reporting in surgery.²⁹⁸

The six authors of this paper are termed the 'Lean authors' because their paper focused on using the QI technique Lean in emergency care. Three of these authors were approached directly by e-mail and they each received an information sheet specifically designed for the author-checking exercise (Appendix 6). These three authors included the first, second, and last authors. The *American Medical Association Manual of Style*,²⁹⁹ which is cited by the Council of Science Editors White Paper on promoting integrity in scientific journal publications,²⁸² suggests that a first author on a manuscript has usually contributed the most to the work. Subsequently, authors usually appear in descending order according to their level of contribution, but a decision can be made to place the most senior author last.²⁹⁹ Therefore, while the views of half the authorship team is not a complete view of the whole team, these three individuals are likely to have contributed most to the project and this is sufficient to obtain a sense of divergent and shared views for this piece of work.

The author-checking exercise was conducted through a series of one-to-one telephone interviews. The interviews each lasted around 45 minutes and

followed an interview schedule (Appendix 7). Two grids (Table 12 and Table 13) were issued to the Lean authors two weeks prior to the interview, and they were used as a topic guide to stimulate discussion about what was difficult to report, and why.

The first grid contained a list of anonymised excerpts selected from the transcripts of participants. These excerpts related specifically to the McCulloch et al²⁶⁸ article. Excerpts were selected across all three stakeholder groups (authors, consumers and gatekeepers). Between two and five excerpts were placed in each of seven categories: word count/journal requirements; the character of QI work; description of materials used; replication; context and second publications; QI method; and theory underlying the intervention (Table 12). These categories had emerged as consistent themes across all participants during coding. I selected the excerpts, which were then reviewed by both of my supervisors (MDW and GPM) to ensure that a wide spread of participant statements was included.

The second grid included 10 items suggested by the earlier interview participants as possible methods by which reporting might be improved (Table 13). A visual analogue scale, drawn as a horizontal line of 100mm, asked the participants in the author-checking interviews to score each item (from strongly agree to strongly disagree). Because these grids contain findings from the interviews, they are presented in Chapter 8, section 8.1.

3.4.10 Data analysis

Interviews were transcribed verbatim by professional transcription services, which were bound by confidentiality agreements. NVivo (a software package designed to aid qualitative data analysis) was used to manage the data.

NVivo has been criticised for imposing an external structure, thereby making analysis technical and more superficial.²⁹⁰ I sought to avoid this pitfall by discussing my analysis on a regular basis with my supervisors and with my advisory panel. This enabled a free-flowing dialogue of ideas which were not strangled by an external structure. I enjoyed verbalising my thoughts because sharing them with others helped me to give careful thought to my interpretation

of the data. Despite its pitfalls, I found NVivo to be extremely useful, because it provided an easy method by which I could navigate a large number of words, but also record of the evolution of the project. NVivo could tell a story of how codes became themes and the process I used to generate meaning from the data.

The methodological literature provides a range of perspectives on whether qualitative research should be conducted as a deductive (use data to confirm or negate an idea) or inductive (generate ideas from the data) process, or as a combination of the two.²⁸⁹ In addition to this traditional definition of deductive and inductive strategies, some have debated whether or not there should be a hard and fast distinction²⁹⁴ between them. Glaser and Strauss's grounded theory and its associated constant comparative method³⁰⁰ emphasises the importance of theory generation from the data and many researchers have felt that qualitative work should intuitively be 'inductive.'²⁸⁹ Opposing arguments have placed value on deductive qualitative studies which test a hypothesis based on pre-existing theory, such as Rashidan et al³⁰¹ who used the behaviour change theory during their qualitative research to explain GP prescribing. This study supports the view that inductive and deductive strategies can be used in the same project,^{231,302,303} and I therefore used both strategies. The deductive research strategy used in this thesis was concerned with using existing literature (which proposes the widely held perception that reporting is poor) to consider arguments as to why this problem exists. This literature informed the development of the research questions and interview schedule. Inductive reasoning was used in parallel to the deductive strategy. The interview transcripts were read and concepts were identified, then re-grouped and re-defined. This allowed entirely new themes to be generated. The themes were used to answer the question "Why is reporting so hard?", and contributed to the development of theoretically significant ideas. This study also supports the view that deductive and inductive strategies are not always mutually exclusive.²⁹⁴ I noticed this during a phase of line-by-line coding when I generated new codes inductively, but at the same time orientated myself towards deductive analysis, recognising a wish to confirm (or not) a theory about the emerging code.

The constant comparative²⁹⁰ method helped me to make comparisons as the analysis progressed.²⁸⁹ I tried to think about the participant's perspective as I read the transcript. This helped me to notice how data at different points in the same transcript could be compared to get a sense of how meaning surfaced from the conversation. I was also able to compare data on the same topic between individuals. I used the constant comparative method to focus my analysis of the story of poor reporting, to explain *what* is difficult to report in QI, *why* it is hard, and *how* it might be improved. The stages of this analysis are described chronologically below.

The first step of analysis involved becoming familiar with the transcripts. I read them and I noted down interesting points.²⁸⁹ Following this, data collection and coding were conducted concurrently, allowing the emergence of new codes to inform the interview schedule. I continually discussed coding and themes with my supervisors; this allowed new insights to be included as the data analysis developed. The second step of coding involved open coding. I analysed the data line-by-line, noting what the essence of each line was in a short phrase. Then I found it more intuitive to use segment-by-segment coding, providing meaning for larger chunks of data. This phase of coding was fast and I generated concise codes made up of short words or phrases, such as 'peer review' or 'reporting guidance.' I then used broad 'bucket codes' to sort my data into categories. Bucket codes are described by Marshall and Rossman³⁰⁴ and I found them to be a useful method of translating data into a tentative list of categories called 'buckets.' I later sub-divided these larger bucket codes, such as 'replication', into smaller 'open' codes, such as: 'replication is desirable', or 'replication is not needed'. Reflecting on this data led me to create strong analytical directions and complete a phase of focused coding. I used the principles of axial coding (relating codes to each other to create new categories) and selective coding (discarding redundant codes and selecting relevant codes). Axial and selective coding was described by Strauss and Corbin,³⁰⁵ but rather than follow axial and selective coding exactly as it was originally proposed, I adopted Charmaz's²⁹⁰ approach of 'sensitising concepts.'

The term 'sensitising concept' originates from the work of Blumer (1969)³⁰⁶ and Charmaz has used it to describe a set of assumptions (or general ideas) from

the qualitative data which, when combined, 'sensitise' the researcher to ask particular questions about a problem.²⁹⁰ As I explored my codes, sensitising concepts were helpful, because they are not defined by a set of clear attributes and rather "suggest directions along which to look" (p.160).³⁰⁷ For example, I allowed my assumptions about the data to produce initial codes such as 'the patient in the middle,' which later became part of a theme about values. Blumer³⁰⁶ emphasises that "meaning is derived from or arises out of social interaction" (p.2) and social interaction is important in generating sensitising concepts. Codes such as 'the patient in the middle' highlighted my interest in the value of patient involvement for QI and the value which would be gained for patients by improved reporting. Line-by-line and segment-by-segment coding helped me to see the data through the participants' eyes, but sensitising concepts allowed me to follow leads that I recognised within the data.

The use of sensitizing concepts suited the requirements of this study because open coding could still be sorted in a logical frame, while at the same time, decisions could be made about how each category should be defined as new links between codes were identified.²⁹⁰ The process of sorting the data allowed me to constantly compare data from the same individual, different individuals, and the categories of author, consumer and gatekeeper. I refined and developed my sensitizing concepts and I was able to eventually build up a level higher than coding, which involved the abstraction of information to create themes and, eventually, theory.³⁰⁸

In the last stage of analysis, one of my supervisors (GPM) reviewed a selection of excerpts from three transcripts (one QI author, one QI consumer and one QI stakeholder) to verify the themes I had identified. GPM read the transcripts and reviewed them against the themes which are described in Chapters 5, 6, 7 and 8. The transcripts reviewed by GPM were selected at random using an internet based random number generator (<http://andrew.hedges.name/experiments/random/original.html>) so that selection bias (the risk that I would deliberately select a 'good' transcript to improve the outcome of the verification exercise) would be reduced. The verification of qualitative data by a third party has been discussed at length in the literature.^{309,310,311} I recognise that the same data can be interpreted differently

by different people and that these differences are often celebrated as part of the richness of qualitative enquiry.²⁴³ My own view is that I could not reasonably present a set of themes without adhering to the process of regular 'checking' and discussion with my supervisors which has been so central to the evolution of this project. This working style has involved differences of opinion, but it has always sought out consensus, and verification is a means by which I believe I can demonstrate a set of qualitative results which can be judged by others.

The steps of analysis recorded above were broadly sequential and easy to follow. The question of when I should conduct the literature review for my interview study was not as easy to unravel. Authors hold differing opinions and while every researcher brings a pre-existing background knowledge,³⁰⁵ some suggest that reading the literature too early might be best avoided so that subsequent data analysis is not affected by preconceived ideas.³⁰⁰ Conversely, Charmaz is clear that timing restrictions on conducting literature review are not necessary. This was fundamental in my study because the systematic review was conducted first, and *a priori* arguments on the problem of replication were constructed early.

According to the Felder Solomon Index of Learning Styles questionnaire,³¹² I am an active and visual learner. Using NVivo is a very visual method of working with data, using colour codes and highlighting. During early drafting of my findings chapters (5, 6, 7 and 8) I translated NVivo references (quotes) from the data into a Word document which had been organised into headings and sub-headings, reducing the data into manageable chunks. Early in the analysis I also used a 'grid' structure to display comparisons within and between individuals about what they felt was hard to report, why they felt it was hard to report it, and how reporting could be made better. This helped me to make sense of the data and eased my ability to translate it into themes as looking down columns and across rows to make sense of data is very familiar to me from my clinical background in hip and knee surgery. I commonly use matrices (for example, the management recommendations for patients with metal on metal hip replacement)³¹³ to synthesise complex information on an everyday basis. This improved my confidence that I had thoroughly examined each participant's story, that regularities and relationships across the data had been

exposed, and that my emergent themes were complete. To present the qualitative findings chapters (5, 6, 7 and 8), I used quantitative data to count how many participants expressed particular views. This helped me to be conscious of examining detail, to avoid giving undue attention to views that were not commonly held and to give an honest and transparent account.³¹⁴

3.5 Qualitative work: Study operating procedures

This section will explain how I aimed to act responsibly and sensitively in response to the ethical requirements of this study. Ethical requirements, including informed consent, confidentiality, withdrawal and safety monitoring are woven into the fabric of the standards I set myself as a researcher and a clinician.

3.5.1 Informed consent

A participant information sheet was attached to every e-mail sent to potential participants inviting them to participate in the study. This information sheet explained: the purpose of the study; any risks involved; the data collection method; how the information would be used; confidentiality; the time involved in the study; and participants' freedom to withdraw at any given time without prejudice and with no obligation to give the reason for withdrawal (Appendix 8). The information sheet also explained that sections of the anonymised transcripts may be shared with the study team, and with the authors of the QI paper used in the task during the interview. The participant was allowed as much time as they wished to consider the information and ask questions. Each participant was given my contact details if they wished to ask questions. Participants were advised that if they would like to take part they could either e-mail me or register their interest on the online registration form at the project website www.le.ac.uk/qi-project-emma-jones.

I then sent each interested participant a second 'interview arrangement' e-mail. This e-mail asked whether the participant had had an adequate chance to ask any questions. If they wished to take part, the participant was asked to complete the consent form (Appendix 8) which was written into the body of the e-mail. The participants returned their consent form by replying to the e-mail from their own personal or work e-mail address. Consent was only obtained by me.

Completed consent forms then printed out and kept in a folder dedicated to consent as source data. This folder was stored by the UoL's Department of Health Sciences in a locked drawer. Participants were provided with a copy of the anonymised QI report by McCulloch et al.²⁶⁸ They were asked to consider highlighted sections of this report prior to their interview but if they did not have time to do so, extra time was allowed on the day. This second e-mail contained a link to an online scheduler which participants used to book their interview appointments (<https://emmajonesleicester.youcanbook.me/>.) YouCanBook.Me is a secure password-protected online booking system available to the public.

3.5.2 Confidentiality and data handling

This study complied with the Data Protection Act³¹⁵ and the Social science APPLIED to Healthcare Improvement REsearch (SAPPHIRE) group's Standard Operating Procedures for handling confidential information.³¹⁶ All interviews were conducted by telephone in a private room using an encrypted voice recorder. The telephone call was recorded while the telephone was on loud speaker. Participants were reminded that the tape recorder could be turned off at any time if they wanted to speak privately and that in any written outputs they would be referred to by number not name. The QI report²⁶⁸ used during the interviews was anonymised by removing the authors' names and affiliated institutions, though participants could have identified this relatively easily if they had tried.

The digital recorder was kept in a locked office drawer at the UoL's Department of Health Sciences and recordings were deleted when the study was completed. All interview transcripts were stored on a password-protected university computer. The interview transcripts were anonymised using study-specific participant numbers and only I knew the identities of individual participants. Where quotations were used no identifiable information was included.

Access to the strengthening quality improvement intervention reporting in surgery (SIQINS) study computer files containing transcripts, and documents such as screening logs was strictly limited to me via username and password. A screening log kept details of individuals who had been screened for inclusion.

This log included their study number and a reason for exclusion if they were not recruited. An enrolment log listed the study-specific numbers of individuals who were enrolled on the study, as well as their contact details (telephone number) and the date of interview. The participant's name and other identifying details such as institution was only included on the e-mail verification of consent. The e-mail verification of consent was printed and stored in a binder dedicated to consent forms at the University of Leicester's Department of Health Sciences, in addition to being stored electronically in a separate password-protected SIQINS study file.

3.5.3 Discontinuation/withdrawal of participants

Each participant had the right to withdraw from the study at any time. Four withdrew before an interview had been arranged because they did not have sufficient time to complete the interview. Participants were told that they could withdraw at any time during the interview. If a participant withdrew during the interview, consent would be sought to use the interview data up to the point of withdrawal. If the participant did not give permission for their data to be used, all data relating to that individual that had not yet been anonymised (and could still be linked to the participant) would be destroyed.

3.5.4 Safety monitoring

No medical harm was anticipated as a result of participating in this study. The study was not an interventional study, no patients were involved, and patient care was not altered. It was unlikely that any untoward, unfavourable and unintended sign, symptom, or disease would be produced in a participant. Nonetheless, I operated on the principle that an interview would be terminated as necessary if a participant expressed discomfort. Regular supervision with my PhD supervisors helped me to ensure that interviews were conducted to the best standard I could deliver, with sensitivity to reducing participant burden. Risks to my own safety were minimal because interviews were conducted by telephone only.

3.5.5 Quality control, quality assurance procedures and approvals

This study was overseen by me and my supervisors (MDW and GPM). We ensured that management systems were in place and used efficiently to maintain high standards of research integrity and behaviour. We met initially more frequently, resolving queries face to face and by e-mail as the study opened. Later, we met at least monthly. The study was conducted in accordance with the current approved protocol and the principles of good research governance which included ensuring that the research followed sound ethical principles such as careful data management and respect for confidentiality.

The Department of Health (DH) research governance framework applies to all research under the remit of the Secretary of State for Health. This framework states that all research conducted with NHS staff requires a sponsor. The sponsor for this study was the University of Leicester, which took overall responsibility for the conduct of this study. The NHS Health Research Authority (HRA) decision tool (<http://hra-decisiontools.org.uk/ethics/EngresultN1.html>) was used to identify that SIQINS did not require NHS REC approval because the study recruited NHS staff as participants, not patients. The DH research governance framework also stipulates that the principles of Good Clinical Practice (GCP) apply to all studies, not just clinical trials. Thus, GCP is the gold standard for non-CTIMP (Clinical Trial of an Investigational Medicinal Product) studies and full GCP certification was completed before recruitment began.

Under the Medicines for Human Use Regulations Statutory Instrument (SI) 2004/1031,³¹⁷ which was transposed into UK law from the EU clinical trials directive in 2004,¹¹⁴ it is against the law to start recruiting until a clinical trial has been approved by a research ethics committee. SIQINS is not a clinical trial and does not fall under the jurisdiction of this law, but the principles of the Medicines for Human Use SI highlight that any research which is subject to the DH research governance framework should undergo ethical review before recruitment can commence. Therefore University of Leicester Research Ethics Committee (REC) approval was sought. The protocol, participant information sheet, and all proposed advertising material was submitted to the University of Leicester REC for written approval. All study documents were approved prior to any participant recruitment.

The process of obtaining the required approvals to begin recruitment was a meandering path which engendered feelings of delight, frustration, boredom and achievement. Reflection upon this process has been useful because qualifying how I conduct myself now has helped me to focus on how I should progress towards a future leadership role.

3.6 Qualitative work: Reflective practice

A reflective diary was kept throughout this study. This helped me to come up with solutions for the various problems I faced along the way. Although this chapter has been presented chronologically, my experience of conducting the research was not so neat. I managed the busy and sometimes messy emergence of overlapping research stages by constant reading and vigilant management of my calendar during recruitment.

I found the telephone interview method to be convenient because they were conducted in my place of work. This meant that I could immediately share my reflections on how an interview went with colleagues in the SAPPHIRE team while it was fresh in my mind. Although a lack of non-verbal cues may cause the participant to feel detached from the interviewer,³¹⁸ I did not find it difficult to build a good rapport over the phone. This could have been because my interviewees were all QI stakeholders who were likely to have a strong ability to present themselves as coherent speakers to others. Also, my participants may have already felt familiar with the existing network of QI researchers which includes me and my supervisors. When I had established a good rapport, I felt that I had built a “climate of trust” (p.143)²⁸⁹ which enabled speech to flow more readily. The telephone did not represent a major barrier to this climate of trust because a conversation about reporting was unlikely to reveal personally sensitive topics.

There were some challenges to conducting telephone interviews, which I have termed ‘technical’, ‘organisational’, ‘operational’ and ‘social’ challenges. ‘Technical’ challenges included accidentally cutting the phone off and battling with poor telephone reception. ‘Organisational’ challenges included learning to navigate the UoL system to gain access to international dialling (which is not installed at the desk of each student). ‘Operational’ challenges related to the

day-to-day operation of the project. For example, I adjusted my working hours to accommodate telephone calls to varying overseas locations, including Australia, which required a very late evening in the office for me and an early start for the interviewee. 'Social' challenges included the realisation that visual cues are important, particularly at the end of a conversation. A little more effort was needed on the telephone to co-ordinate the successful closure of a conversation without seeming too abrupt. I used Christmann's suggestions³¹⁹ to inform my approach to telephone interviewing and kept a log of my experiences which reminded me not to make the same mistake twice. I hope that it will serve as a useful tool for my future career path which is likely to involve further qualitative research.

I also found it helpful to deliberately take a short pause before dialling the telephone number for each interview to remind myself to listen carefully. Although I did not ask my participants directly for their views on what it was like to take part in my study, listening to the cues they gave helped me to adapt the conduct of the research as I went along. For example, the reactions of my participants helped me to realise that I did not always manage to make myself clear. To avoid delays created by needing to repeat or re-phrase complex questions, I adapted my interview technique and I learned to sum up what I wanted to ask in a single sentence. Interviewing successfully felt like a complex skill to learn and a constant combination of reading, reflecting and asking the opinion of my colleagues was vital to complete this study. I was motivated to commit to this continual cycle of reflecting, learning and changing my practice because promoting accurate reporting of what was done and why it did (or did not) work is important to me. I wanted this study to provide insight into how successful QI interventions could be reported with greatest impact for patient care.

In this chapter I have explained how the systematic review and interview study were conducted and I have justified my research design choices. I have provided an honest account of what motivated me to complete this PhD and how I navigated a steep learning curve. In the next chapters I will present the results of my empirical work. The first findings chapter, Chapter 4, will present

the results of the systematic review. The following four findings chapters, Chapters 5-8, will present the results of the qualitative study.

4 SYSTEMATIC REVIEW FINDINGS

As I explained in Chapters 1 and 2, reporting quality in QI and surgery is consistently poor, and has been for some time. Some features of poor reporting, which have been identified in separate evaluations of the perioperative and QI literature, may also be seen in a specific review of the perioperative QI literature, but this review has not been conducted yet. Accurate identification of which reporting items are most frequently incomplete in the perioperative QI literature could enable specific recommendations to be generated about how to move the problem of poor reporting forward. Therefore, the objectives of this systematic review were to answer the questions: 'What is the completeness of reporting in the perioperative literature on QI methods and interventions?', and 'Which elements are most frequently missing?' Systematic reviews are conducted by adhering to a series of predefined, reproducible steps to identify, select, and evaluate relevant research,^{144,247} and findings generated from each of these steps will be reported in turn.

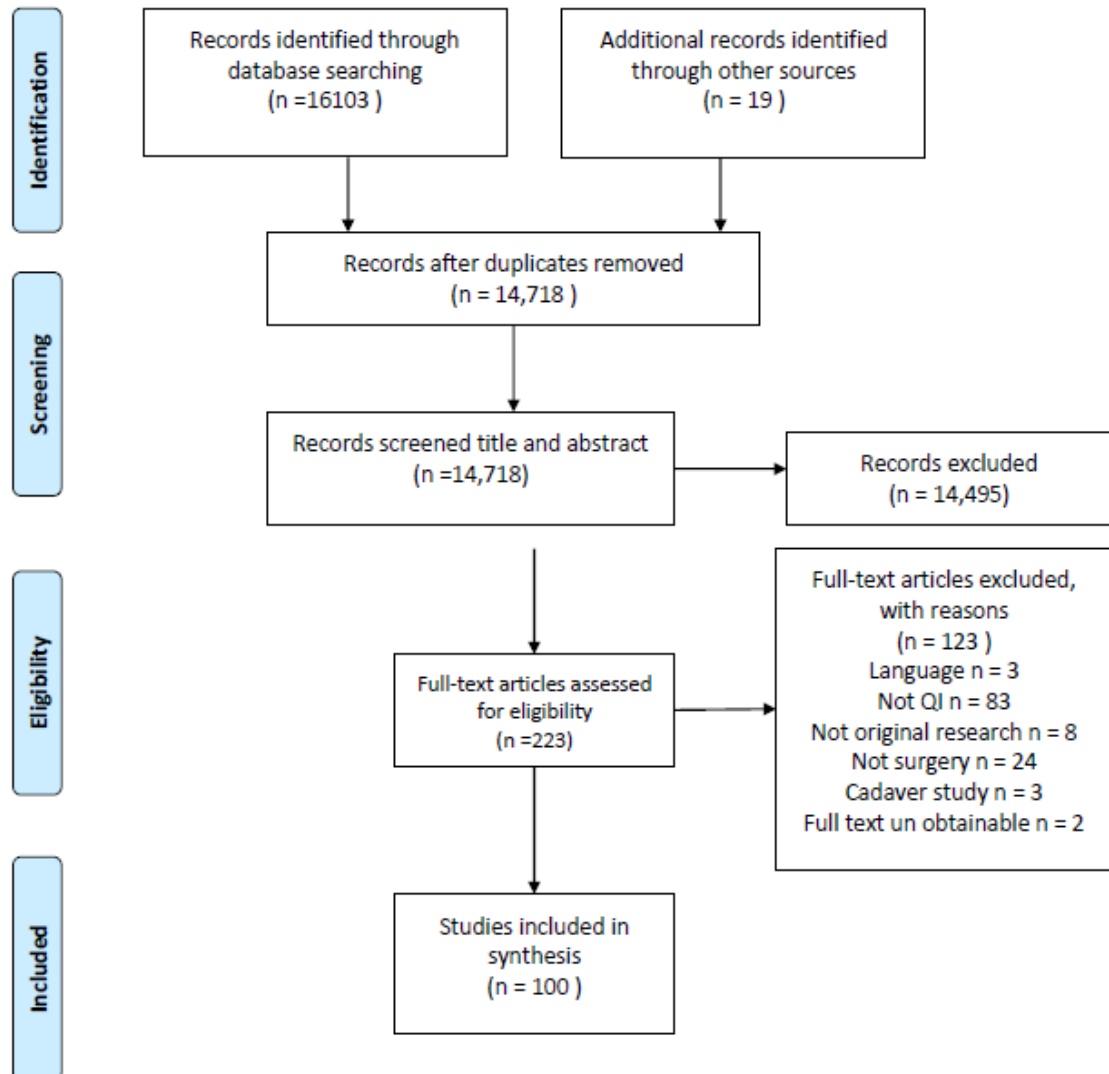
4.1 Article identification

The PRISMA diagram (Figure 1) reports the phases of article identification and selection. The search strategy identified 16,103 abstracts from database searches and 19 from other sources. Duplicates were removed, leaving 14,718 articles. From this, a further 13,603 articles were excluded during initial screening leaving 1,115 potentially relevant articles. NJL and I independently assessed titles and abstracts of these 1115 articles (in discussion with GPM and MDW) where discrepancies arose against the inclusion criteria, resulting in the full-text being obtained for 223 articles. Of these, 123 articles were excluded because they were: not written in English (3), not QI (83), reviews or conference abstracts (8), not surgery (24), unobtainable (2) or cadaver studies (3). On completion of this process, 100 articles were deemed eligible for inclusion.

When the process of article identification is reported, discussion of why some papers were excluded and not others is not usually presented. Instead, authors are encouraged to elaborate only when much arbitration was required to resolve disputes.¹⁴⁴ This systematic review did require much arbitration to design and deliver a suitable article selection process and resolving disputes about which studies to include was often difficult. I provided an account of why such disputes

arose and how they were resolved in Chapter 3, section 3.3.1. Making this explicit is an important step towards reducing variability in QI definition.

Figure 1 PRISMA Flow Diagram



4.2 Study characteristics

Of 100 articles, 40 focused on two or more surgical specialties. The remaining 60 named a specialty including: cardiothoracic (21), colorectal/general (19), musculoskeletal (4), vascular (4), urology (3), gynecology (3), hepatobiliary (1), upper gastrointestinal (1), transplant (1), ophthalmology (2) and otorhinolaryngology (1). Settings included emergency (6), emergency and elective (13) and elective surgery (81).

Study designs were varied (Table 7). Many articles (65) did not explicitly identify their study design, but on inspection were found to be before-and-after studies (a design using data collected at defined time points before and after the introduction of an intervention, also known as the pre-test/post-test design).³²⁰ Nine studies were labelled as cohort^{321,322,323,324,325,326,327,328,329} yet did not appear to feature true observational study designs, and one study was mislabelled as a case-control.³³⁰ The US was the most frequently reported country for study setting (67/100) (Table 7).

Table 7 Study design and country of publication frequency for 100 perioperative QI papers

Study design	Number of papers
Uncontrolled before and after (including Statistical Process Control studies)	85
Non-randomised controlled interventional study	4
Cluster randomised trial	4
Randomised controlled trial	2
Case series	2
Time series or segmented time series	2
Cohort	1
Country of publication	Number of papers
US	67
Netherlands	7
UK	5
Germany	4
Australia	3
Norway	3
Finland	1
Taiwan	1

Table 7 (continued): Study design and country of publication frequency for the 100 perioperative QI papers

Country of publication	Number of papers
Iran	1
Tanzania	1
Turkey	1
Columbia	1
China	1
India	1
Switzerland	1
France	1
Italy	1

The most commonly reported motive for undertaking QI was that of reducing infection (30), followed by improving intra-operative clinical processes (such as reducing never events) (18) and reducing post-operative complications (such as bleeding and prolonged intubation) (15). The least frequently cited aims were: improving the post-operative discharge process (3), improving self-management (3) and reducing the post-operative incidence of venous thromboembolism (VTE) (1) (Table 8).

Table 8 100 perioperative QI publications categorised according to targeted clinical issue

Intended clinical outcome of the QI intervention	Number of papers
Reduce post-operative complications – VTE	1
Improve post-operative process - organisational (e.g. discharge process, complication reporting)	3
Improve self-management/patient and family satisfaction	3
Reduce post-operative error (e.g. medication errors, unnecessary tests)	4
Reduce post-operative complications - pain and sedation	5
Improve pre-operative process (e.g. fasting, admissions documentation)	9
Improve intra-operative process - organisational (e.g. start time, waiting time between cases)	12
Reduce post-operative complications - general (e.g. hyponatremia, lengthy intubation)	15
Improve intra-operative process - clinical (e.g. operative technique, never event)	18
Reduce post-operative complications - infection	30

An analysis of the number of QI articles published per year revealed a steady increase in the yield of publication volume over time, rising from three articles published in 2000, to 17 articles in 2013, with 68 articles published between 2008 and 2013. The journal that each article was published in was analysed using Web of Science (WOS), a web-based platform which provides access to the Science Citation Index. Seventy-three articles were published in journals which were indexed by WOS, and these journals were indexed under 24 category types. The most common categories were surgery (31 included articles) followed by cardiac and cardiovascular systems (10). Other types of category included: health policy and services; urology and nephrology; public, environmental and occupational health; and medical informatics.

4.3 Completeness of reporting: Quality interventions and QI techniques

In this section, I report the appraisal of the completeness of reporting of the TIDieR checklist items and QI techniques (Table 5). A full list of 100 included papers can be found in Appendix 9.

4.3.1 Completeness of reporting: Quality interventions (TIDieR).

All articles used a combination of quality interventions (Chapter 3, section 3.3.1), such as introducing a care pathway, providing staff education, changing the timing of ward rounds and issuing reminders. No specific combination of interventions was used more often than any other. The most commonly reported intervention (classified according to the modified Shojania et al²⁵³ QI taxonomy) was education (59), including any form of teaching and learning, such as workshops. Nine studies provided access to web links for additional material such as web-based educational modules. Checklists were reported as quality interventions in 14 articles; protocols were reported as quality interventions in 43. More than half (51) of the studies included feedback as part of the quality intervention.

The distribution of TIDieR scores for the reporting of quality interventions across 100 papers approximately followed a normal bell-shaped curve, with a slight skew towards higher ratings (Figure 2). The most common (modal) score was 7/11. The average (arithmetic mean) score was 6.31/11. The median score was 6. The TIDieR items that were most usually fully reported were: *why* (complete

in 98 papers), *brief name* of intervention (complete in 94), *where* (complete in 77), *what (procedures)* (complete in 69), and *who* (complete in 52) (Figure 3). The nine articles from the journal *BMJ Quality Improvement Reports*, which were used in a training exercise to improve the consistency of article selection, were also scored separately by me and the second reviewer NJL. These nine articles achieved a slightly higher mean TIDieR score of 8.2.

How well the researchers actually adhered to the intervention protocol and reported intervention fidelity (item 11: how well *actual* – Table 5) was the most frequently incomplete TIDieR item (Figure 3), absent in 74 articles. An example of good reporting of intervention fidelity is provided in Thomassen et al:³³¹ “Our checklist was used in 61% of all anaesthesias during the testing period” (p.1183).

Modifications to interventions were also generally poorly reported (incomplete in 73 articles). Other items that were not fully reported in more than half of the included papers were: *what* (materials – any physical or informational materials used in the intervention and details on how they can be accessed) (incomplete in 62 articles), *when and how much* (incomplete in 60 articles), and *how well (planned)* (incomplete in 53 articles).

Only one article was judged fully complete against the 11 TIDieR items.³³² Extracts of text from this paper (Table 9) offer examples of each completely reported TIDieR item.

Figure 2 Range of TIDieR scores 0-11

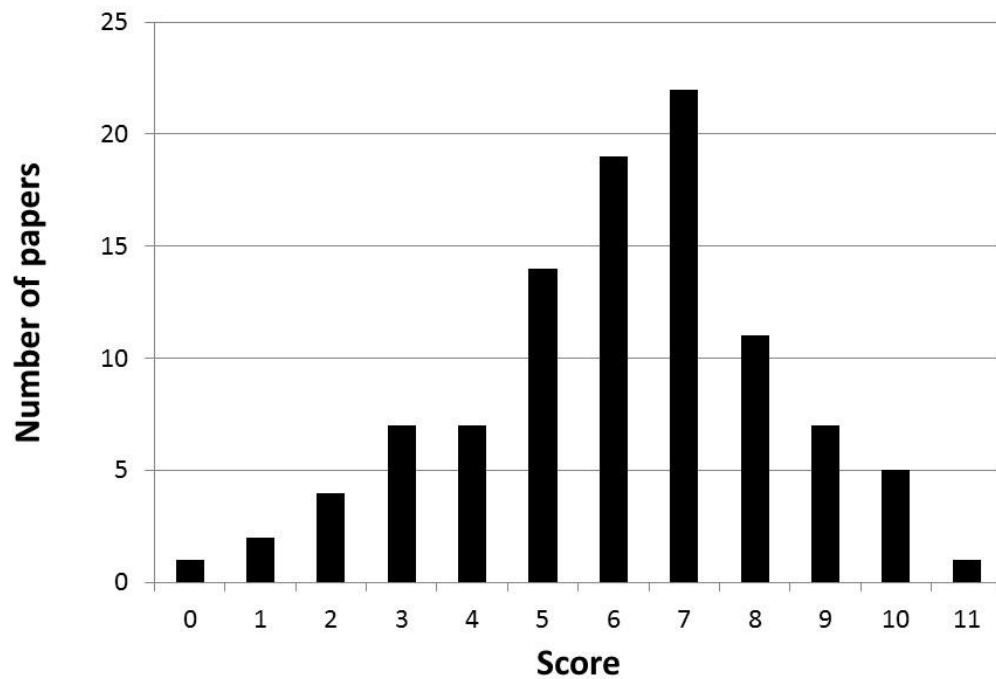


Figure 3 TIDieR scores for perioperative Quality Improvement papers

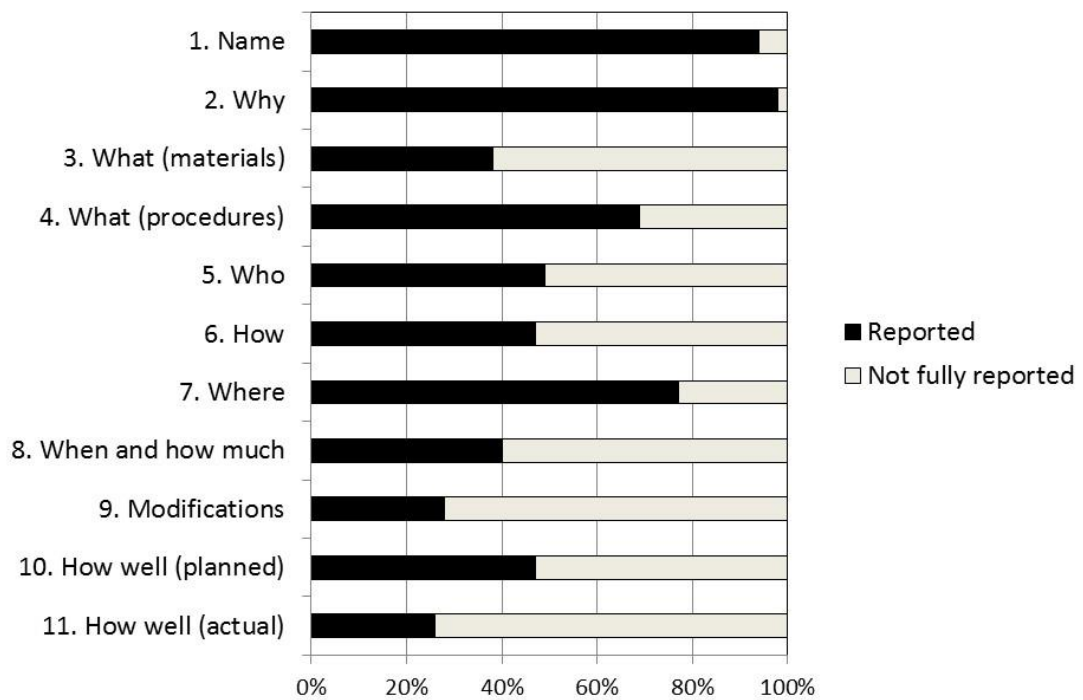


Table 9 Example of fully complete TIDieR items

TIDieR items			
1	2	3	4
Brief Name	WHY: Rationale	WHAT: Materials	WHAT: Procedures
“Multifaceted systems intervention” (p. 2014).	To determine whether a multifaceted systems intervention would eliminate catheter-related bloodstream infections.	Web-based training module (http://www.hopkins-medicine.org.heic/), a standardised checklist to be used during central venous catheter insertion, a daily goals form, a central catheter insertion cart with four drawers and partitions to organise the contents, which can be rolled to the patient's room.	Five interventions, including “empowering nurses to stop the procedure if guidelines were not followed” (p. 2016). An example of reporting item 4 (procedure) for this intervention is: “we discussed with both residents and nurses that the nurse should page the SICU attending physician if the resident, after the nurse identifies a violation, fails to correct the violation” (p. 2016).

TIDieR items			
5	6	7	8
WHO: Staff Group, training & expertise	HOW: Mode of delivery	WHERE: Setting + infrastructure	WHEN & HOW MUCH: Dose/Duration
<p>“An interdisciplinary team including the SICU codirectors, ICU physicians, nurses, and infection control practitioners” (p. 2015). Additional information about how the QI team worked together included: “The SICU leadership met with both groups of providers and emphasized our focus on patient safety and teamwork” (p. 2018).</p>	<p>“All physicians or physician extenders who insert central catheters were required to complete a Web-based training module and successfully complete a ten-question test before they were allowed to insert a central venous catheter” (p. 2016).</p>	<p>The surgical ICU is a 16-bed unit for adult patients undergoing general, orthopaedic, transplant, trauma, and vascular surgery. Extra features of the setting relevant to the intervention included “Hospital Epidemiology and Infection Control (HEIC) required leaders from hospital administration to . . . provide the SICU with the additional resources required” (p. 2018).</p>	<p>From 1 January, 1998, through 31 December, 2002, ALL those who insert central catheters were required to complete a Web-based training module (URL provided) and successfully complete a 10-question test. The checklist was implemented in two phases, which included a 2-week observation period.</p>

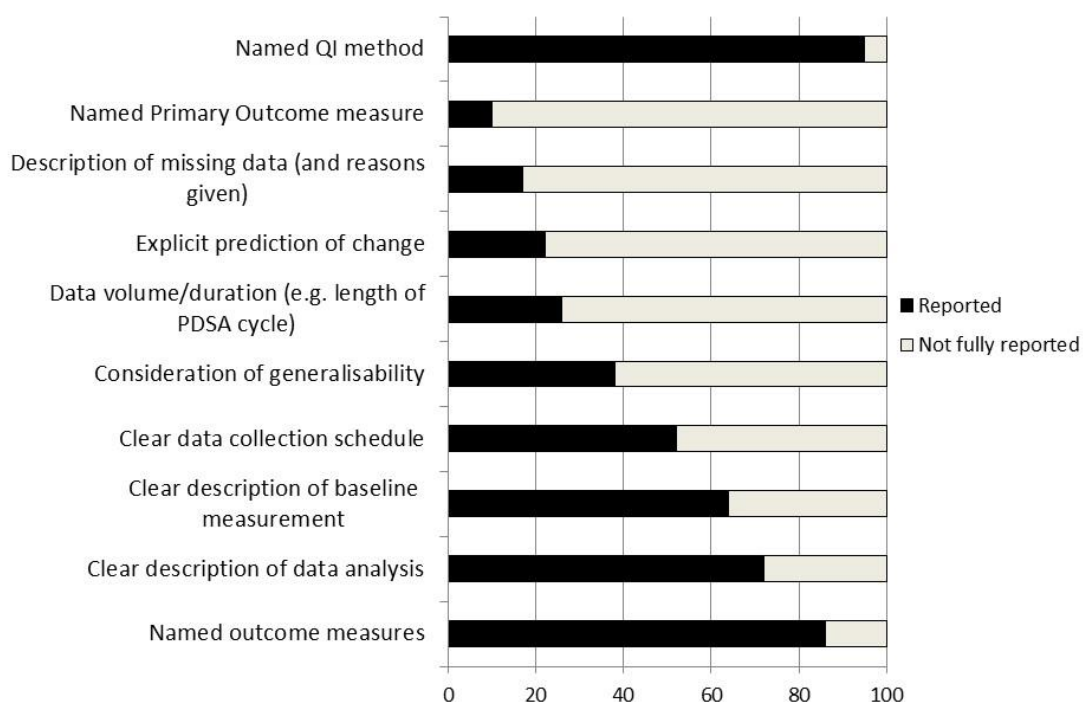
TIDieR items			
9	10	11	
Modifications	HOW WELL: (Planned)	HOW WELL: (Actual)	
<p>The management of central venous catheters once they are inserted did not change during the study period, with the exception of the change in daily patient visits in the study SICU to ask whether catheters could be removed.</p> <p>Following a pilot test, the checklist was modified based on feedback received.</p>	<p>Residents were informed that a checklist was being used. This strategy allowed nurses to feel more comfortable intervening if they observed a violation, because they felt that an expectation had been set and as a result, they were less likely to have an uncomfortable encounter with a physician and they could stop the procedure if they saw a violation of evidence-based practice.</p>	<p>An example of actual compliance with one of the five interventions: “During the first month, nursing completed the checklist for 38 procedures: eight (24%) for new central venous access, 30 (79%) for catheter exchanges over a wire, and three (8%) were emergent. A nursing intervention was required in 32% (12/38) of central venous catheter insertions” (p. 2017).</p>	
<p>SICU, surgical ICU; QI, quality improvement. * Adapted from Berenholtz SM, et al.³³²</p>			

4.3.2 Completeness of reporting: QI technique.

The most frequently reported QI techniques (refer to Chapter 3, section 3.3.1 for the distinction between *QI techniques* and *quality interventions*) reported in the articles were audit and feedback (42), PDSA (28), Six Sigma (16), Continuous Quality Improvement (CQI) (16), and Statistical Process Control (SPC) (10).

Assessed against the QI technique criteria, based on the Cochrane Effective Practice and Organisation of Care (EPoC) review group's data collection checklist (Table 5), the most frequently *complete* items were naming the QI technique (fully reported in 95 articles) and outcome measures (86). The most common *incomplete* items were the description of missing data (not complete in 83 articles) and the provision of a primary outcome measure (missing in 90 articles) (Figure 4). This was followed by incomplete reporting of an explicit prediction of change (missing in 78 articles) and data volume (e.g. length and number of PDSA cycles) (74). Just over a third (38) of articles discussed whether or not the results might be transferrable to another setting (Figure 4).

Figure 4 QI technique scores for 100 perioperative QI papers



4.3.3 Additional study features

Use of the SQUIRE Guidelines was reported in only one article.³³⁰ Only two articles – both in the area of orthopaedic surgery – reported patient and public involvement. In the first of these, Robarts et al³³³ interviewed patients to identify research priorities and service needs and involved patients in a process mapping exercise. In the second, Rycroft-Malone et al²³¹ collaborated with a patient co-researcher throughout the conduct of the study. Six studies used patient-reported outcomes in the form of non-validated visual analogue pain or patient satisfaction scales.^{333,334,335,336,337,338} One study reported adverse effects of the QI intervention: in this case, nurses and junior doctors reported anxiety about implementing the interventions.³³²

4.4 Discussion: Systematic review

This is the first systematic review assessing how well QI methods and quality interventions are described across diverse settings (emergency and elective, and primary and secondary care) in surgery. The methods were designed to allow the review to extend beyond a presentation of raw outcome data, and also to consider: What rationales are provided for the application of specific QI strategies? How is QI defined? To what extent were patients involved? This systematic review has been subjected to peer review and was published in June 2016²⁸⁸ (Appendix 11).

This systematic review has demonstrated that the reporting quality of the perioperative QI literature is sub-optimal when measured against the TIDieR checklist and a second QI method checklist. The reviews strengths include the use of standardised data extraction tools and checklists, and achieving consensus with a third reviewer to limit subjectivity in data analysis. Forming a systematic review team which included both social scientists and clinicians also ensured that messages could be developed which addressed gaps in current understanding. In addition to these strengths, it is possible that there were some limitations to this work.

4.4.1 Limitations of the systematic review

Limitations of the systematic review include problems with: the accurate identification of the literature, the scoring of the results and the use of the TIDieR checklist. Before each of these points is addressed, an opening statement to defend the selection of the systematic review method itself will be presented.

4.4.1.1 Justifying the choice of systematic review method

My initial review of the literature (Chapter 2) revealed that inadequacies in reporting exist in the QI literature, such as poor reporting of missing data, generalisability, the number and timing of PDSA cycles, allocation concealment, unintended consequences of QI, cost and outcome data reporting. Yet, I could not assume that what had been found in other sections of healthcare literature, would be the same in perioperative QI research. Therefore, the question: “What

is the current standard of reporting of QI in the perioperative literature?" still required an answer.

It could be argued that a systematic review 'over-engineered' the task of answering this question, and that selecting a few key journals that frequently publish perioperative QI studies could have been sufficient. I feared that this option would be inadequate because deliberately selecting journals that are known to publish QI work could have caused the selection of only the higher scoring papers. The systematic review method involved exhaustively searching all available literature, and removes the problem that some articles will be selected at the expense of others which could create misleading results.

4.4.1.2 Finding the literature

The ability to ensure the accurate and complete identification of all available literature is one of the most pressing concerns of all systematic reviewers, but the article retrieval and selection process I chose does contain some flaws. First, literature searching can be performed in many different ways. Booth et al³³⁹ argue that topic-based searches using typical databases, which is the technique I chose, may not be the best approach for reviewers who wish to explore complex interventions. Studies of complex interventions often require the publication of more than one article, and Booth et al³³⁹ proposed a methodology to examine these articles in small 'study clusters' (p.1).³³⁹ I did not deliberately use the cluster technique and it is possible that I did not identify all articles within clusters, but my systematic review did identify *some* small clusters using a traditional database searching approach. For example, Kreckler et al¹⁹ provide extra detail about one of the seven interventions originally presented in McCulloch et al.²⁶⁸ Booth et al³³⁹ also recommend using the cluster technique to identify web-based resources and book chapters, but these evidence sources were not of interest to me as I sought to identify the quality of reporting of QI research published in healthcare journals.

Second, the selection of words and phrases for the search strategy was difficult. The possible scope of QI literature is wide because the use of Medical Subject Headings (MeSH) terms and keywords has been inconsistent²⁴ and there is a general lack of consensus on how QI terminology should be applied.¹⁸⁶ For

example, the authors themselves used myriad terms, and some articles (which met the criteria of QI) did not explicitly use the term 'quality improvement' in their titles or abstracts.

To design a search strategy specific enough to ensure that inappropriate studies were correctly rejected, and sensitive enough to ensure the correct capture of studies relating to QI in perioperative care, I adopted a number of strategies. I conducted rigorous pre-reading of the literature, which allowed me to identify and modify a QI search strategy which had previously been employed by the Health Foundation. I used MeSH terms (which operate as part of the National Library of Medicine's controlled vocabulary thesaurus) because keyword searching only finds references where an exact word or phrase, specified by me, appears in the title or abstract. MeSH terms therefore provide some mitigation against the possibility that I may not have thought of all relevant synonyms. I ensured the keywords that I did select were varied, and could accommodate the myriad of terms I had seen in my pre-reading. The term "QI" itself was not used because it reduced the specificity of the search by retrieving a large number of articles relating to other topics such as 'qi' in traditional Chinese medicine.

Another strategy to ensure the comprehensive retrieval of all available literature was the use of multiple databases. This is helpful because while both MEDLINE and PubMed are both linked to the National Library of Medicine journal citation database, journals found in PubMed may not be selected for MEDLINE (and vice versa). Also, there can be a delay of some months between records being indexed in MEDLINE and in the Cochrane Central Register of Controlled Trials (CENTRAL), since CENTRAL is only updated quarterly. Articles from the journal *BMJ Quality Improvement Reports* were not included in the main review, though they were used during a training exercise. This decision was justified because this journal was not PubMed indexed at the time of writing. Also, this journal is different from all of the other included journals because it gives authors a template to guide their QI writing.

Third, it was hard to answer the vexing question of what types of interventions should be described as QI interventions (see Chapter 3, section 3.3.1) in order

to understand which articles should be selected and which should not. There are many articles on checklists, protocols, pathways and electronic handover that could have been included as QI research in surgery. For example, 33 articles were identified by Treadwell which evaluated four different surgical safety checklists, but not all of these were included because I created a firm set of criteria to explain what would count as a 'QI article'. To do this I modified a taxonomy by Shojania et al,²⁵³ to distinguish between what I term *quality interventions* and *QI techniques* (section 3.3.1). The distinction between quality interventions and QI techniques is by no means consensual in the literature and Shojania et al²⁵³ had not specified that QI interventions should always be combined with QI techniques to be counted as a QI strategy. However, I was confident that the original taxonomy of nine interventions had been developed based on appropriate and comprehensive methodology²⁵³ and applied successfully in previous literature.² I was satisfied that my modified version of this taxonomy could be used consistently by both me and my second reviewer to retrieve relevant literature because I conducted pilot tests.

The pilot tests I conducted allowed me to adjust the search strategy to ensure that a known selection of perioperative QI articles was captured. These articles were obtained by searching the reference lists of previously published systematic reviews. Taking the example of Nicolay et al,³ of their 34 included papers, I included 23 because two articles used a paediatric population, and nine were published before my cut-off date of 2000. During this process I identified some articles (for example, Burkitt et al)³⁴⁰ which were not included in a previous systematic review by Levy et al.⁷ Their own inclusion and exclusion criteria may equip them with good reasons for excluding them, which demonstrates there may be many different yet justifiable methods of selecting QI literature. During pilot tests, the search strategy was adapted, but some articles may still have been missed, especially those published most recently. For example, it is possible that my search may not have captured all studies stimulated by the recent Improving Trauma Care Act in the US³⁴¹ or the UK's emergency and urgent care review.¹¹⁵

Finally, it is possible that I may have been prone to over selecting articles with positive results, or perhaps, in this case, articles which contained better

descriptions of interventions. The possible problem of selection bias was mitigated by using more than one reviewer. The exclusion of three articles written in language other than English and of unpublished reports may have also introduced some bias³⁴² but this would have greater importance if the review had intended to estimate the size of the interventions' effect rather than describing their content. It does mean that I can only comment with confidence on reporting standards in the English-language perioperative literature.

It is a difficult balance for the systematic reviewer to obtain enough papers to ensure that nothing is missed while also reducing 'noise' to ensure the project is manageable. More than 16,000 papers were identified, indicating that my search strategy had low sensitivity and specificity and this had to be resolved by detailed review. It is likely that problems in search strategy design were related to lack of consensus on how QI terminology should be applied.¹⁸⁶ The fact that only 100 papers published over a 15 year time period met my criteria might seem surprising, but a stringent set of inclusion criteria was necessary to prevent the inclusion of a vast number of articles about ERAS pathways and checklists which were not QI as we had defined it. The volume of literature obtained seems reasonable given that Nicolay et al's review of a similar field identified 34 surgical QI articles in 2011.

I propose that despite the limitations, the strategies I employed to aid selection of the literature enabled a good compromise between robustness and pragmatism to decrease the likelihood that articles would be missed.

4.4.1.3 Imprecise scoring

The scoring of the QI methods and interventions may have been imperfect. I reported that 90 of the included articles failed to report a primary outcome measure and 78 of articles failed to report an explicit description of a prediction of change. Yet it may not have been appropriate to report them in the first place because papers with a pre-post design do not usually intend to test a causal relationship between the quality intervention and the outcome of interest. Thus, rather than deeming these items to be missing, the absence of certain reporting features could be used to shape a view of what QI research is. Overly negative scoring of the QI intervention may also have occurred. For example, authors

may not report on modifications if the intervention was never intended to be modified, but it would seem unusual for a QI study not to involve modification of some kind because interventions are often changed in response to iterative cycles of measurement.²⁰ Likewise, very positive reporting could have occurred in relation to this same feature because authors may have fully documented the modification of one element of the intervention, when other features were also changed. This would mean the item was scored as ‘fully reported’ when, in fact, other modifications went unreported. I did not contact authors to identify missing aspects of intervention reporting, and the papers were scored as seen. Finally, the task of allocating a reporting item to be ‘fully reported’ or ‘not fully reported’ could be regarded as very subjective.³⁴³ However, this categorisation was reliable when subjected to pilot testing between two reviewers (see Chapter 3, section 3.3.4).

Deficiencies in the reporting of QI may have been underestimated because 83 potentially relevant papers were not included. These 83 papers were all excluded after careful consideration of ‘what counts as QI’ and this involved the use of our taxonomy, wide reading, and collaboration with experts. Some of these 83 articles self-identified as involving QI, but they were excluded because they were most aligned with a human factors (HF) approach. The excluded paper by Catchpole et al³⁴⁴ is a good example of this. Conversely, Smith et al³⁴⁵ was provisionally excluded, but was subsequently included after an e-mail discussion with an expert in the field to clarify the nature of the study:

“This paper does not appear to me to be human factors related. First, it is fundamentally about improving efficiency, not human work or performance. In HF we focus on individuals to do their work better, leading to better performance. This is because HF is fundamentally seeking to make system changes based on human centered thinking. Second, the techniques that were used were not ones traditionally taught and used in HF, such as a task analysis. Third, no-one claiming human factors expertise (or related expertise such as cognitive science) was involved in the project” (e-mail from a Human Factors expert).

4.4.1.4 Use of the TIDieR checklist

This systematic review covers the period 2000-2014, and the TIDieR guideline was published in 2014. It is therefore arguably unsurprising that compliance is limited, because the authors of the articles included in this systematic review could not be expected to use TIDieR as a specific checklist. However, this systematic review was not intended to evaluate compliance with or ‘adherence to’ the TIDieR items per se. Instead, TIDieR was selected because after review of the literature (Chapter 2, sections 2.1 and 2.2), I deemed TIDieR to be the most appropriate tool to establish the current state of reporting. While the authors did not know the specific TIDieR headings, it is reasonable to expect that their attempts to describe their interventions might fit broadly around them, because intervention description has been for many years “a major research activity” (p.2).¹⁵⁰

4.4.2 Explanation of the review findings

The three most frequently incomplete aspects of reporting were: the materials needed to implement the intervention; any modifications applied to the intervention; and how well the intervention was actually delivered according to how it was planned. For the QI method, missing data and a named primary outcome were the most frequently incompletely reported items. These data suggest that authors find it difficult to provide a full account of their QI intervention and technique. Full descriptions are important to determine whether an intervention can be used to deliver a positive effect to benefit patients in a new setting, and how resources should be allocated to achieve this.^{138,150} The poor quality reporting of QI studies is therefore likely to lead to frustration for interested readers who may wish to replicate a QI intervention in a new setting. Incomplete reporting can also be implicated in the problem of research waste because studies that are not fully reported can necessitate additional or futile research that would not be required if the full findings were known. The drive to improve reporting and to reduce waste has gathered some momentum in QI and in perioperative care with the publication of reporting guidelines and taxonomies, and initiatives to reduce waste (see Chapter 2, sections 2.2 and 2.5.4), and this review supports the need for such initiatives.

Complete reporting is necessary to ascertain whether an intervention *can* be replicated, but it has another equally important function, to inform decisions about whether an intervention *should* be replicated.⁸³ When the results of QI studies are compelling and interesting, interventions must be reported in a way that allows recognition of all of their strengths and weaknesses. This involves ensuring any notable caveats are reported, and it was particularly disappointing that I found only one study reporting adverse effects resulting from applying the QI intervention. A QI article should also fully report all the information required to judge *how much* the intervention might contribute towards changing practice for the better across many settings, and this requires an understanding of how the study was designed.

Nine articles^{321,322,323,324,325,326,327,328,329} did not resemble true cohort studies, because they were not observational (they deliberately introduced interventions aimed at change).³⁴⁶ Many other articles (65) did not explicitly identify the design used and this required me to work closely with an epidemiologist to categorise each study. This was challenging, because I learned how each study type should be defined by reading published literature, which itself suffers from frequent incorrect categorisation of study.³⁴⁷ I also faced the problem that conventional descriptors, derived from epidemiological study designs, may not be optimally suited to use of some QI techniques. For example, I allocated ten SPC studies to an ‘uncontrolled before and after’ study design group, but SPC requires the selection of a number of different cases at multiple time points for analysis,³⁴⁸ which could also be described as a time series. However, the SPC work included in this systematic review did not analyse secular trends, which time series require.^{349,350} Instead, the studies collected data before and after an intervention, with the ‘before’ acting as the historical control. Therefore, I categorised the SPC studies as ‘before and after’ discontinuous series without a contemporaneous control.

The appropriateness of imposing a study design label upon 65 studies which did not explicitly describe one could be debated. I decided that an explicit description of a study’s design is helpful so that the nature of the claims the study makes about generalisability and internal validity can be better understood.¹¹⁷ A clear depiction of aspects of study design—such as whether

an intervention it had a historical or population control, and whether it implemented an intervention deliberately or whether it was already occurring—could also translate into better reporting of intervention features such as the ‘when and how much’ TIDieR item, which describes the dosage and timing of the intervention delivery.⁸³ Consistent application of study design terminology could therefore be helpful to QI reporting in the future.

A before-and-after design with a historical control was used in 85/100 included studies. Recent systematic reviews evaluating the effect of QI have focused on studies with a population based control such as an experimental design^{2,351,352} and this has been encouraged by the EPOC review group.¹⁵² If judged by the standards of traditional epidemiology, the before-and-after (sometimes known as pre-post) has weaknesses in controlling for bias and in making causal inferences. Yet, pre-post studies serve a legitimate purpose and they can allow QI researchers to explore interventions in single settings. Noticing how QI studies are presented in the literature allowed me to reflect upon a key difference between QI research and traditional research – that QI work often sets out to secure change in a specific environment, in contrast to traditional research where the primary purpose is to test a hypothesis and generate new knowledge.³⁰

In the course of the systematic review, disagreements arose among the reviewing team members about which articles to select due to disagreements related to broad definitions of QI¹²⁶ and QI terminology²⁴ (Chapter 3, section 3.3.1). Explaining how we formed a consensus on the literature selection process could allow this systematic review method to be reproduced by other researchers who wish to reduce variability in the design and conduct of systematic review in the field of QI. In addition to facilitating replication, the account provided in section 3.3.1 also draws attention to the issue that heterogeneity in how QI reported is problematic for systematic reviewers. Reducing heterogeneity could ease tasks undertaken by systematic reviewers at each stage in the review process. First, heterogeneity can make it hard to capture *all* relevant literature,^{24,197,198} and if a reviewer is not able to be fully exhaustive, less reliable estimates of the effectiveness of QI interventions may be produced.¹¹⁷ Second, heterogeneity in reporting could make it difficult for

reviewers to understand how similar QI interventions are, in order to pool data and compare interventions to generate a helpful summary of results.^{117,119}

Acknowledging the challenges we faced helped me to design an interview study in which I could explore problems with the participants, with some understanding of their point of view. For example, the systematic review was at times difficult because I had never received any formal training in QI as a physiotherapist working full time in the NHS. This led me to consider that members of surgical teams may also experience a lack of training on how a QI study should be constructed, which could translate into reporting difficulties. This systematic review, and the interview study which developed from it, may therefore serve to improve recognition of reporting difficulties, and this is a helpful stimulus for debate about how to move the field forwards.

5 DEFINING QI AND FINDING A SCIENTIFIC FRAME FOR REPORTING

5.1 Introduction to the findings chapters for the qualitative work

The objective of the qualitative study was to explore the question: “Why is the reporting of QI in perioperative care so hard?” The interviews involved asking the participants to describe the reasons why reporting QI interventions and QI techniques in the perioperative QI literature can be difficult and to consider what counts as a QI intervention. The qualitative study also aimed to inform wider debate about how QI reporting would be best supported.

The findings of the qualitative work will be presented in four chapters. These chapters will summarise the themes which emerged from the interviews with participants about what they see as the reasons for difficulties in reporting QI. The themes I derived are as follows:

1. **Defining QI and finding a scientific frame for reporting (Chapter 5):**
This theme outlines participants’ views on inconsistent and complicated use of terminology and their conflicting views on whether or not QI should be regarded as a science. Problematic terminology and a lack of clarity about how to report QI in a scientific manner indicates that QI is a young field which has not yet become fully established. This section outlines why these issues are a barrier to complete reporting.
2. **Influences on reporting (Chapter 6):** This theme is about the influence of journals and healthcare organisations on enabling or inhibiting reporting in QI. Some features of organisations can cause tensions for authors who want to report their work in a manner which is sympathetic to the emerging field of QI, yet also sympathetic to the requirements of conventional medical publishing.
3. **The problem of how to report context (Chapter 7):** This theme is about why the contextual features of QI research (such as leadership, buy-in, culture, teamwork, resources and environment) can be difficult to report. The participants identified pragmatic solutions to improve reporting of context which could be applied to current practice, and aspirational solutions which would need further development.

4. **QI authors and their audience (Chapter 8):** This chapter will present the data from the author checking exercise. This exercise involved asking the participants how they would replicate an intervention described in a perioperative QI article. The article's authors then explained whether their own recollection of the intervention matched the participants' understanding.

To aid the explanation of why reporting is hard, a series of sub-themes were generated. Assigning these sub-themes was not always easy because some (such as word count) could often be related to more than one theme. Thus, data were categorised in the sections that seemed most appropriate for ease of reading: word count, for example, was placed under the theme of context. Throughout the findings chapters the readers' attention is drawn to the links between themes to show how they connect together.

The presentation of the findings begins with an account of the sample size (including recruitment rate and reasons for withdrawal) and the participants' characteristics.

5.2 Sample size

During the participant identification process 73 potential participants were assessed as eligible and they were all invited to take part.

Forty-six agreed to participate in the study and were sent a link to book an appointment. Four did not book an appointment: one person explained that time pressure prevented participation, and three withdrew by providing no further contact. The remaining 42 participants all participated in a telephone interview. The Gantt chart (Appendix 10) reports all of the phases of the study including screening and enrolment.

Nine individuals declined, giving the following reasons: not enough surgical experience (6); not enough QI experience (2); and not enough time (1). The remaining 18 individuals did not respond to the invitation e-mail.

5.3 Participant characteristics

The 42 participants were not equally divided between the three recruitment groups, but a reasonable allocation to each group was attained. There were 12 'QI consumers', 15 'QI authors', and 15 'QI gatekeepers.' The gatekeeper category contained two sub-groups: journal editors (11) and 'influencers' (authors of reporting guidelines) (4). The participants worked in varied geographical locations including England (24), the United States (US) (11), Germany (1), Canada (3), Australia (2) and The Netherlands (1). In the US participants were spread across six states, and in England across seven counties. The participants' professional backgrounds included clinicians (28), academics (13) and a healthcare manager (1).

Clinicians included: surgeons (13), anaesthetists and intensivists (6), internal medicine doctors (4), nurses (2), a radiologist (1), a physiotherapist (1), and a cardiologist (1). Surgeons specialised in upper and lower gastrointestinal (GI) (7), vascular (1), ear nose and throat (1), orthopaedics (1), cardiothoracic (2), and transplant (1). The most common sub-speciality was upper GI, representing four of the 13 surgeons. The intensivists (critical care doctors) in this study started their training as anaesthetists so the two specialties are represented under the umbrella term 'anaesthetists.'

Eleven participants held editorial roles currently, or within the past 24 months. These roles included: editors in chief (2), associate editors (7), deputy editors (1) and assistant editors (1). They worked for six different journals which publish QI work: BMJ Quality and Safety, Implementation Science, International Journal for Quality in Healthcare, Journal of Hospital Infection, British Journal of Anaesthesia, and Journal of Perioperative Care and Operating Room Management.

Table 10 shows the numbers of QI consumers, QI authors and QI gatekeepers grouped by professional category.

Table 10 Participants by professional category and recruitment group

Professional group	Consumer	Author	Gatekeeper	Sum of each row
Surgeon	5	5	3	13
Anaesthetist	3	2	1	6
Internal medicine	1	1	2	4
Nurse	2			2
Radiologist		1		1
Physiotherapist		1		1
Cardiologist		1		1
Academic		4	9	13
Healthcare manager	1			1
Sum of each column	12	15	15	

Throughout the findings chapters for the qualitative work, the participants are labelled according to their allocation to each recruitment group (author, consumer, editor or influencer) and to their professional role (surgeon, nurse, academic etc). Thus, a surgeon who is also an editor will be referred to in the text as Surgeon-Editor.

While participants were grouped according to their predominant professional role, there was some overlap between categories. Some participants had parallel historical interests, defined as experience of working in another role more than two years ago. For example, four of the 13 academics had previously worked as clinicians (as nurses, occupational therapists and doctors), but they were not, for purposes of this study, classified as clinicians because they had not been clinically active within the last 24 months. Some participants had current parallel interests. For example, QI authors also mentioned a need to 'consume' the QI writing of others. While the varied experiences of individuals within each group could threaten between-group comparisons, complete standardisation was not desirable for this study. The participants do not exist in a vacuum, and their varied QI-related activities added depth to the data.

5.4 Defining QI and finding a scientific frame for reporting

Participants characterised QI as a youthful scientific field which does not yet have a firmly established definition and finding the right frame for the scientific reporting of QI has been difficult. Participants explained why these difficulties translate into making reporting hard. This chapter will be presented in three parts.

First, the problem that the language and vocabulary of QI is not fully developed yet will be presented:

“There’s a lot of inconsistent terminology.” (Academic-Author 01)

Researchers depend upon a common language and terminology to develop and relay written scientific ideas to others.³⁵³ Language is made up of words and terms and it provides a means through which others can understand a subject. Terms are technical words which may deviate from the meaning they have in everyday language. Participants raised the problem that QI authors can use many different terms for the same concept, and, even when terminology is consistent, the same word can be interpreted in different ways by different people, including the use of the term ‘quality improvement’ itself.

Second, problems with vocabulary may be the symptom of a much more fundamental issue, that there is poor agreement among QI stakeholders on what QI is and whether it should be regarded as ‘scientific’ or not. This can explain why it can be hard to define what makes QI research, research.

Third, the absence of a consensus on how QI is understood translates into many difficulties in describing QI interventions in a scientific manner in the literature. It has been difficult to find the right scientific frame for QI reporting which can adequately describe complexity, rationale, materials, fidelity, outcomes and active ingredients.

5.5 The vocabulary of QI

Participants found it difficult to explain what QI actually is. They struggled to define what characterises QI as a separate field of science, with little evidence of full consensus across participants. They described a set of terminology that can be regarded as specific to QI, but may be problematic because different

people can understand the same terms (such as Lean) in different ways. Likewise, different terms (such as PDSA and Plan-Do-Check-Act (PDCA)) can refer to the same activity. Participants clearly explained why problems in the vocabulary of QI should be overcome.

5.5.1 What is quality improvement?

Perhaps the first and most important phrase is that of quality improvement itself. All participants were asked: 'What does the phrase "quality improvement" mean to you?' (Interview schedule: Appendix 4) and they responded by indicating the diversity of meanings attributed to it:

"QI means different things to different people" (Academic-Author 54).

When describing quality improvement, participants used words such as 'fast' (*Surgeon-Consumer 07, Academic-Editor 56, Surgeon-Author 66*); 'cyclical' (*Surgeon-Author 26, Surgeon-Consumer 29, Physician-Author 30*) and 'systematic' (*Academic-Author 09, Nurse-Consumer 10, Physician-Author 30, Anaesthetist-Consumer 45, Academic-Editor 51, Surgeon-Author 66*).

Each participant explained their understanding of the phrase 'quality improvement' by selecting different combinations of words and this can be problematic.

"Well articles are words and words have several meanings. Words are gathered in a phrase and also that phrase has several meanings. If you give the same phrase to ten different people and ask them what they have understood from that phrase, you get ten different stories." (Physician-Editor 65)

Participants across all three recruitment groups drew parallels between QI and other types of evaluation. Six consumers, five authors and six gatekeepers suggested that QI was 'similar to' audit (10), change management (1), human factors (1) and implementation science (5).

"People who classify themselves as human factors experts, some of them agree that what they are doing is extremely similar to what is

described as quality improvement, but others say it is fundamentally different.” (Surgeon-Author 66)

Of these 17 individuals, four pointed to the overlap between these fields, but they also provided a clear explanation as to what sets them apart. For example, two editors (*Academic-Editor 13, Academic-Editor 56*) explained the difference between implementation science and QI, but suggested that understanding how terms like ‘implementation’ should be used is difficult for students and more established academic staff.

“When we see QI studies, often the biggest issue is that it isn’t at all clear what is being implemented . . . it may be a locally home grown solution that seems like a good idea to the people who are doing it . . . and that is a key distinguishing feature between implementation research and quality improvement science – . . . people are learning this in a very ad-hoc way . . . it’s not a stable clearly codified set of information for people to absorb and that’s unfortunate.” (Academic-Editor 56)

Four QI editors (*Academic-Editor 56, Academic-Editor 13, Physician-Editor 53, Academic-Editor 49*) suggested that the problem of heterogeneous terminology and a lack of consensus about how QI should be defined, carries with it the risk that QI work will become spread among a variety of journal types because authors adopt a “*scattershot approach*” (*Academic-Editor 56*) to submitting articles for publication.

The views of the participants in this study demonstrated no clear consensus about whether the use of a QI method should be used to define QI. Eight participants, with a mixture of roles across author, consumer and gatekeeper groups, defined QI as having a “*certain terminology, terms and methods*” (*Academic-Editor 13*). Participants named this group of methodologies (including Lean, Six Sigma, PDSA, Audit and Feedback, SPC, Statistical Quality Control, TQM, and Continuous Quality Management (CQM)) as “*improvement structures or strategies*” (*Physiotherapist-Author 05*) which are used specifically in QI:

“Commonly you will use a quality improvement method, whether that’s Lean, PDSA, six Sigma, total quality management, as a method to help you understand the problem that you’re trying to improve.”
(Physiotherapist-Author 05)

These methods (or QI techniques as I termed them in Chapter 3), are strongly associated with the use of methods which originated in manufacturing industry^{3,354} and they each employ a pre-defined set of steps which aims to generate measurable improvements in outcome.⁵³

Conversely, three participants (all QI gatekeepers) reported that the meaning of the phrase ‘quality improvement’ should not be associated with QI methods. This view is also reflected in the literature, where some authors have acknowledged that QI can include a range of initiatives which do *not* depend on established techniques adapted from industry.³⁰

“What I am not meaning [by my definition of QI] is a narrow range of quality improvement techniques which have been introduced to healthcare from the manufacturing industry.” (Academic-Influencer 21)

Participants (11) directly stipulated that QI can be defined by whether it is capable of developing new knowledge and most of them reported that QI could not generate completely new knowledge:

“Once you know that something can work somewhere and then you’re trying to get it to work consistently in a routine practice, in clinical research they call that effectiveness as opposed to efficacy, then you’re sort of starting to come into quality improvement territory.” (Physician-Editor 53)

Some (4) explained that it can be easier and more common for QI to build upon existing knowledge, but did not want to commit to defining QI as being unable to generate new knowledge:

“I think evaluation of quality improvement can be seen as research but I think quality improvement in itself is not...I was going to say it is not

about generating new knowledge, but maybe it is...” (Academic-Editor 62)

All 42 participants highlighted that QI is a complex concept. The wide range of definitions they provided supports the notion that there is no single accepted definition of QI.¹⁸

All participants were asked: ‘When is an intervention a ‘QI’ intervention?’ to establish whether certain types of intervention can be associated with QI programmes. Participants generally felt that the breadth of intervention types in quality improvement is vast. Many of them (20) questioned whether a surgical intervention such as a type of incision or care pathway could be classed as a QI intervention (or quality intervention as I described in Chapter 3). Seven of this group of 20 participants felt that *any* intervention that improved quality was a QI intervention, even a clinical treatment:

ELJ: “A type of surgical approach, for example?”

Participant: “Sure, that is quality improvement.”

ELJ: “Okay. Why is that?”

Participant: “If you have established that the new technique is better and has improved outcomes, by definition you are improving quality so yes I would deem that quality improvement.” (Surgeon-Author 58)

Of the seven participants who reported that a clinical intervention can count as a QI intervention, three clinicians, based in the US, Canada and the UK, drew parallels between Enhanced Recovery After Surgery (ERAS) and QI. As described in Chapter 2, ERAS pathways combine multiple elements (such as early mobilisation and carbohydrate loading), to achieve an overall effect (reducing length of stay) which improves an existing clinical process (perioperative care).^{220,221,222,223,224} Participants said that the structure of ERAS and QI interventions are similar because they use existing knowledge to combine multiple elements into one intervention to achieve an improved outcome. A fourth clinician described the use of machine perfusion to pump

livers as a QI intervention (Surgeon-Consumer 33). The remaining three members of this group were academics and journal editors, and rather than giving a clinical example, they shared the view that:

“It’s unlikely that you intervene to deteriorate a process, so in a way, all interventions [including clinical interventions] are improvement interventions.” (Academic-Editor 49)

Thirteen participants disagreed, stating that the type of incision and ERAS are ‘clinical’ interventions, and that clinical interventions and QI interventions are not the same thing.

“Clinical interventions are those that involve changes in clinical treatment. So introduction of a new drug typically wouldn’t be considered quality improvement. if you’re comparing laparoscopic versus robotic surgery to see if it improves outcomes, that wouldn’t necessarily be considered quality improvement.” (Academic-Editor 13)

Participants who were clinicians also explained that QI interventions are not the same as clinical interventions, reporting that:

“Rather than a clinical intervention, or specific clinical skills, anything that affects the way that a clinical intervention is delivered can potentially be a quality improvement intervention.” (Nurse-Consumer 10)

Examples of what counts as a ‘non-clinical’ QI intervention included: simulated learning (Surgeon-Editor 28, Nurse-Consumer 10), which they described as a bespoke type of patient safety training, and interventions which focus on changing a system:

“QI involves practical, pragmatic intervention, as in taking a system and trying to make the system better than it was yesterday.” (Anaesthetist-Consumer 69)

Some tentative patterns can be seen within this data. Of the 13 participants who said that clinical interventions and QI interventions are not the same thing, seven were not clinicians (editors), but six were clinical staff. Clinical staff felt that quality interventions can include some clinical activities such as simulation.

Conversely, the editors who were not clinicians (7) expressed a clear preference that clinical activities should not be classified as quality interventions at all.

Two editors and one author suggested that the question of whether an intervention is a quality intervention or not, may not be an important question at all:

“I think sometimes that can be a semantic thing, I mean, there’s definitely some grey zones - but I’m not sure what you’re trying to get at with this actually.” (Physician-Editor 53)

5.5.2 The terminology of QI

There are many terms which can be used to describe QI techniques and quality interventions (defined in Chapter 3 section 3.3.1), and this has long been acknowledged and debated in the literature.^{24,129} Thirty participants in this study were concerned that QI terminology is fluid, changing between individuals and over time, and the meaning of specific words can be understood in different ways:

“Then there’s the problem of specific words that people understand differently or don’t understand in some sense at all.” (Academic-Influencer 64)

Participants raised the problem that different people can denote different meanings from the same words, such as Lean, PDSA and complex intervention. They used the word *“loaded”* (Academic-Author 01) to indicate that there may be different types of Lean, PDSA or complexity:

“The term ‘Lean’ is widely misused and used in different ways, by lots of different people, so the word doesn’t necessarily have specific meanings to the reader.” (Anaesthetist-Author 47)

Participants said that it is important to overcome the difficulties associated with terminology because the inconsistent use of terminology can hamper a writer’s ability to convey a clear message to their reader, which exacerbates the problem of poor reporting:

“We need a fairly standardised set of terminology that is used around reporting and discussing and implementing QI. We need to use the same words all the time. So set words that I do recognise include – Plan-Do-Study-Act – but in this place it’s called the Plan-Do-Check-Act. Now, if we want to call it Plan-Do-Check-Act, that’s fine. I don’t mind, but let’s call it one or the other.” (Anaesthetist-Author 47)

Some clinicians expressed a preference for Plan-Do-Study-Act (PDSA), and the terms PDSA and ‘Plan-Do-Study-Act’ were used more frequently (on 96 occasions during speech) than PDCA and ‘Plan-Do-Check-Act’ (used on 17 occasions during speech) during their talk about QI. Participants also described ‘quality improvement’ and ‘improvement science’ as synonyms for each other. They recognised that terminology can change over time. For example, a track and trigger chart has now changed its name to MEWS, which one clinician explained:

“If you are at all familiar with general hospital process in the UK and you’ve got more than five years’ clinical experience, you’ll remember what a track and trigger chart was. It’s now called a MEWS chart.” (Surgeon-Author 66)

Participants explained that the problem that terminology changes over time in a field which is still emerging and quite new, brings with it the conundrum that young or new researchers may not be familiar with terms used by their more experienced colleagues. This can contribute towards difficulties in understanding written reports. Likewise, in QI terminology can change over time as the field becomes influenced by other academic disciplines:

“Many of the concepts that we have badged as new and novel in improvement science, have actually been established - there are already methods for dealing with context, for example.” (Academic-Author 48)

Participants in this study recognised that QI writing requires the use of terms which overlap with other scientific fields including: behaviour/behavioural sciences (17), sociology/social science (12), Toyota/industry/engineering (12), and human factors (1):

“There is so much that goes into this in terms of organisational behaviour and sociology and all the rest of it.” (Academic-Author 09)

Participants did not recall being taught the language of the social sciences and they emphasised that they were more familiar with applying a conventional medical scientific template to their reporting. Two stated that the terminology in the social sciences is:

“Complicated” and “Confusing” (Anaesthetist-Editor 17)

“Baffling,” “social science-esque,” and “Quite boring because they have a combination of words that I don’t understand” (Anaesthetist-Consumer 69).

This data indicates that QI is stuck between different fields and it has not established itself fully as an independent scientific field. Authors struggle to satisfy many competing priorities, for example, adequately reporting issues related to *“organisational behaviour and sociology” (Academic-Author 09)*, while also satisfying the requirements of conventional medical reporting, which some participants were more familiar with. This conflict can complicate reporting first, because QI authors may be forced to conform to the writing style of one field or another.

“There is a tendency amongst the social scientists to over-interpret and use elaborate writing styles which is the opposite of what we try to encourage in scientific medical writing.” (Anaesthetist-Editor 17)

Second, familiarity with a conventional medical scientific template meant that some participants were less likely to seek out and learn from existing QI research because they find it hard to read.

Fear of the terminology of an emerging field which is new or different could then translate into reporting difficulties because it hampers the ability of new QI researchers to learn more about what the science of QI looks like, and how it differs from conventional medical reporting. Editors noticed that terminology used in QI is separate from the language used in traditional scientific medical writing, and some of them recognised that new *“mental models” (Academic-*

Editor 60) might be helpful for authors to learn how to recognise QI as a separate field of science and that certain terms should not be expected at the expense of others. For example, some terms, such as 'p-value', should not be valued more highly than intervention rationale:

"Jeez, we see a lot of stuff where as soon as they get their P value, that's the end of the road." (Surgeon-Editor 16)

While the terminology of overlapping scientific fields was isolated as a specific problem, participants also discussed terminology used in QI more generally. They described terminology used in QI as:

"Jargon" (Anaesthetist-Consumer 37, Anaesthetist-Author 47)

"A written syntax that is quite difficult to understand" (Anaesthetist-Editor 17)

"Academic speak" (Academic-Influencer 21).

Participants also explained why jargon is used, helping to explain why problems with terminology have persisted. For example, one participant suggested that terminology is sometimes imposed by peer reviewers (Surgeon-Author 66). Another reported that complicated terminology can be linked to a wish to elevate one's status:

"There's a tendency in all types of science to make it seem more important and special by using inaccessible language, and, actually, it puts people off, there's a pomposity about it." (Anaesthetist-Author 47)

Participants were frustrated by confusing terminology and definitions of QI. Authors wanted to create clarity about what QI is, because they wanted their QI audience to feel good, rather than confused, alienated or 'talked down to.'

"When I write my papers, I think, 'well, how would I like the people reading them to feel?' and I think of a good experience of reading that I've had." (Anaesthetist-Author 47)

Participants also wanted to resolve the problem of terminology to ease communication between QI specialists and non-QI specialists such as lay

people, economists, and implementation scientists; to ease the precision with which QI articles are understood and how they are distinguished from articles published in related scientific disciplines and to reduce the amount of time it takes to clarify the meaning of words which are poorly understood.

“‘Safety relevant care processes’ – I was like ‘gosh what does that mean?’ Then you have to go and hunt through to see what those things are.” (Nurse-Consumer 38)

This, in turn, may ease the “scattershot approach” (Academic-Editor 56) that QI authors sometimes take when submitting manuscripts for publication, because editors may be more likely to receive articles written with terminology which is appropriate for their journal.

It has long been recognised that providing straightforward engaging writing which includes consistent terminology is essential for adequate dissemination and utility of literature.³⁵⁵ Participants aspired to achieve straightforward and engaging writing in QI and the four authors of reporting guidelines (who have been referred to in this thesis as ‘influencers’) explained that work to tackle problems with inconsistent terminology has begun:

“There’s the problem of specific words that people understand differently or don’t understand at all. An important early task in the process of getting the new [anonymised] guidelines to work better was to make sure that everybody understood what the hell everybody else was talking about.” (Academic-Influencer 64)

5.6 QI reports and QI research

During discussions about the definition of QI, participants suggested that QI reporting can be distinguished into two types: the reporting of QI projects and the reporting of QI research. Participants emphasised that whether QI work is written up as a report or as research, it should be written in a ‘scientific fashion’ which can be understood by everyone.

“I think people who report it [QI], in order for it to be useful, they really need to look at it as a science.” (Surgeon-Editor 16)

Participants stated that QI projects which are reported as 'QI reports' tend to be *"real-world"* (Surgeon-Author 66), a *"science project"* (Surgeon-Editor 16), *"practical and operational"* (Academic-Author 48) and a *"halfway house"* (Academic-Influencer 21). Many participants (10) defined QI as being part of their everyday world, and that QI is a 'rebranding' of something they were doing anyway as part of their commitment to good practice. These participants also raised the issue that the distinction between QI projects (which are reported as QI reports) and QI research can be blurry because there is a bridge between academia (research) and local improvement efforts, rather than a wall which separates the two:

"My job was created to act as a bridge between academic interests and the practical activities linked to improvement science." (Academic-Author 48)

Three participants (Academic-Editor 49, Physician-Author 30 and Academic-Editor 62) suggested that it might not be necessary to distinguish QI reports from QI research, because the purpose and presentation style of QI and QI research can be very similar.

"I don't see a distinction in these two types of communication strategies. If you study quality improvement, you write your report on it, and then you write a scientific paper. It's the same thing. A quality improvement report and a research paper could have exactly the same purpose – to describe the context factors, to assess the scalability of an intervention. I don't think there's a distinction of what a quality-improvement report is and what a quality-improvement research paper is." (Academic-Editor 49)

The majority of participants distinguished QI reports as separate and distinct from research, but they did not reach full consensus on exactly what makes QI research, research. Participants used: the type of journal the QI work was published in, the use of narrative data in QI, how QI data can be evaluated, which includes generalisability (assessment of external validity) and the management of bias (assessment of internal validity),³⁵⁶ and ethical approval, as parameters for deciding whether the work was research or not.

5.6.1 Type of journal

Six participants distinguished QI reports from QI research by referring to the journal *BMJ Quality Improvement Reports*. Reporting in this journal can be done with “*more latitude to be more descriptive*” (Nurse-Consumer 10), it tends to be “*very routine,*” (Physician-Editor 53), and it can be regarded as a “*QI magazine*” (Anaesthetist-Consumer 45), which gives authors a good chance of getting published (Physician-Editor 53, Physician-Author 30). Conversely, participants used examples such as *The Lancet* to describe research reporting in a ‘traditional’ journal, stating that for publication here, the work would need to have “*something novel about it*” (Physician-Editor 53) and would need to be written concisely because:

“Traditional journals may prevent you from explaining things in as much detail as you would like.” (Nurse-Consumer 10)

5.6.2 The use of narrative data in QI

“In the nature of quality improvement, there is a desire to be creative.” (Academic-Editor 56)

The idea that QI projects and QI research can be a naturally creative enterprise involving storytelling was discussed by 17 participants. The use of narrative and qualitative data in QI was discussed more broadly by all 42 participants in this study. However, one author, five consumers and four gatekeepers felt reticent about using a qualitative writing style and suggested that “*woolly*” (Anaesthetist-Consumer 37) qualitative accounts are less valued by the surgical research community. This is explored in more depth in Chapter 7, section 7.5.

In contrast to this view, thirty-two participants reported that narrative data was integral to the ‘storytelling’ of what happened during a QI study and necessary to produce a full scientific account of QI research.

“No single piece of information on its own means anything without the rest of the story or the data that helps make sense of it.” (Academic-Author 01)

Participants felt that telling a complete story of what happened could help them to present complete accounts of QI projects and QI research which are able to explain the results. Nine participants (on 28 occasions during their interviews) suggested that providing narrative accounts could improve the credibility of the QI project, or QI research. They related the credibility of QI to the completeness of the written publication and their ability to decide how much 'scientific confidence' they had in the results. They said that for QI reports and QI research to be complete, narrative data is required to adequately describe rationale, the role of context, and bias.

"The other is the business about credibility, because if the article is hazy or fuzzy or incomplete about implementation, then you can scratch your head and say well, it looks interesting, but I'm not sure I have much confidence in their results, because I don't know exactly what they did and how they did it." (Academic-Influencer 64)

5.6.3 Generalisability and using QI reports in a new setting

Twenty-two participants talked directly about generalisability and the extent to which a published account of a QI intervention could be used to translate the intervention into a new setting. Margolis et al⁸⁴ suggest that it is desirable for published QI work to be generalisable and the participants in this study broadly agreed with them, stating that being able to 'copy' a quality intervention from one setting to another is an important function of reporting QI research.

Nine out of 15 editors suggested that generalisability is an important function of research, but the word 'generalise' may not be appropriate in QI research. These editors suggested that rather than generalising every ingredient of an intervention, it may be better to evaluate how 'transferable', 'feasible' or 'portable' some ingredients are over others. Other editors (3) suggested that it is rare for QI studies to be generalisable because QI studies tend to use a single site. They said that very often the whole study will need to be repeated again in a new setting, before judgements can be made about whether an intervention can be implemented in a new place, beyond the original study setting.

“The nature of how we implement things in practice for QI is that it needs to be redone (re-evaluated) in almost exactly the same way... to check for implementation problems.” (Physician-Editor 53)

Therefore, they classed single-site work as ‘projects’ and described them as being unsuitable for receiving academic credit as research in traditional peer reviewed journals. Academic-Editor 53 felt frustrated by “*narrow*” single-site QI studies whose authors seemed to say ‘just believe us.’

Conversely, nine participants saw value in reporting single-site studies because generating a new idea about how to do something which can be developed, rather than generalised, is just as valuable as being able to create an exact replica. Also, even when QI studies do involve multiple sites, each setting often has “*little quirks*” and “*different practice patterns*” (Physician-Author 25), making generalisability difficult even in the more ideal multi-site study design. A discussion about the difficulties in reporting the features of local settings is provided in Chapter 7 under the theme ‘the problem of how to report context.’ Some participants in this study emphasised the importance of exact replication:

“If other people can't repeat [a successful intervention] then the work is meaningless.” (Anaesthetist-Author 55)

However, the most widely held view of participants in this study was that rather than being able to “*get it off the shelf and pull it in*” (Academic-Author 01), QI interventions usually require some element of re-testing in a new healthcare setting to work out which elements can be kept the same and which elements may need to be changed. It is important to acknowledge this in a published account of QI because:

*“Just saying that it worked is not as useful anymore.”
(Academic-Author 09)*

These participants described another meaning of the term ‘replication’: the creation of another version of the intervention. This indicates that the replicated QI work may not always be an exact repeat of the original. Rather than creating a precise replica, participants felt that it should be possible to:

“Use it [the QI publication] to know what was going on and be able to adapt it [the intervention] for other settings.” (Academic-Influencer 12)

Although generalisability was provided as a parameter which could distinguish research from QI projects, the participants were not sure how the generalisability of each component of a quality intervention should be judged.

5.6.4 Bias in QI

Bias was discussed by 25 participants, including 13 out of 15 editors, nine out of 15 authors and three out of 12 consumers. This participant group suggested that actions to diminish bias were usually seen in research studies, rather than QI projects, and that in research, bias is an important item to report.

Some participants referred to the absence of equipoise as something which could introduce bias in scientific research.

“You have got a hypothesis, you have a theory about how something works and you then set about trying to prove the hypothesis wrong. Now that is a very challenging thing for people who are not used to scientific thinking because what you want to do is to prove your pet idea right, but if you are no longer in equipoise, you are biased and if you are biased you will select results that are positive and ignore those that are negative.” (Physician-Author 55)

Other participants argued that in QI equipoise was neither achievable nor desirable, because QI interventions are deliberately designed to improve outcome, and gaining ‘buy-in’ to the intervention right at the start is very important. One of the surgeons interviewed said:

“QI is more real-world and it is not research. It will inherently have all the biases. It’ll have clinical biases, selection biases, reporting bias, buy-in from staff, it’ll have all the biases one can think of.” (Surgeon-Author 14)

Many participants said that studies with controls can and should be done in QI. Six participants emphasised the benefits of using a study design with a contemporaneous comparator or control because it can: *“make the project stronger” (Surgeon-Consumer 07), “protect generalisability” (Physiotherapist-*

Author 05), and “avoid bias” (Academic-Editor 17). Participants also explained that contemporaneous controls are difficult to implement in both QI (*Academic-Author 09, Surgeon-Author 66*) and in perioperative care (*Academic-Editor 13*) because the intervention may not be delivered in a uniform way.

*“Different surgeons may be doing things slightly better, or slightly worse.”
(Academic-Editor 13)*

Participants also talked about studies with historical comparators including time series (8) and before-and-after (or pre-post studies) studies (10), suggesting that these designs are used more commonly in QI. Study designs without a historical control were regarded as weak.

“Often what happens is you don’t even have any pre intervention data, all you have is just the intervention was implemented and then some time-series data.” (Surgeon-Author 14)

Authors were concerned that some types of QI study were less likely to be accepted for publication than others and this affected how they chose to report their work. QI techniques like statistical process control, and study designs like time series analysis, have been regarded by journal editors and authors alike as *“new-fangled,” “suspiciously dark” (Academic Author 48), and “a little bit out on the far edge, suspicious and squishy” (Academic-Influencer 64).* However, participants recognised that this type of work could be strengthened to give the article more value. Methods of strengthening QI research included using: a historical control (18), a longer period of data collection before the intervention (3), and process measures (10) such as measures of compliance, or healthcare staff experiences which can be used to track an interventions implementation. By strengthening the work, and managing the risk of bias, *“suspicious” (Academic-Author 48)* methods such as time series may then produce *“statistically valid and robust information” (Academic/Influencer 64)* which can be *“very powerful” (Radiologist-Author 24)* and the research would also become more publishable.

“If you did a good piece of interrupted time series work . . . you would probably get a decent quality journal to publish it.” (Surgeon-Author 66)

Another participant suggested that the scientific profile of QI work could be strengthened by adapting the “*traditional evidence based medicine pyramid*,” (Radiologist-Author 24) which tends not to recognise study designs like time series. He considered whether a hierarchy which is similar to the traditional pyramid³⁵⁷ could help elevate the scientific profile of QI by making each level of QI scientific writing, including the use of narrative data, explicit.

This data suggests that participants could not fully explain how bias should be reported in QI, or to what extent bias can (or should) be mitigated in QI research.

5.6.5 Ethical approval

A small number of participants (one editor, one author and one consumer) used ethical approval as a method of distinguishing whether QI was research or not. They used features of QI such as ‘fast’ and ‘cyclical’ (see section 5.5.1) to explain that QI investigators often want to make a difference for “*people in the here and now, not for future generations*” (Academic-Editor 56), and this can provide incentives to commence the QI work quickly. When regulatory and oversight activities slow down the work being done, it may be harder to get things done in a timely way, and this creates less incentive to design the QI work as ‘research.’ This may be a driver for how QI projects are written up and whether they are written up at all.

5.7 Reporting QI interventions

Participants reported that descriptions of interventions are often poor in published accounts of QI.

“I haven’t seen many [QI articles] where people have given detailed instructions.” (Physician-Consumer 72)

They felt that it is important to overcome this and “*showcase*” (Academic-Editor 62) the full intervention. A full intervention description would be more “*compelling*” (Radiologist-Author 24) to read, which would enable others to learn from it and replicate it. However, the participants observed that intervention description might be difficult because QI is a ‘new-fangled’, ‘evolving’, ‘new’,

and ‘emerging’ field, which is gradually coming to be recognised as a separate field of science:

“It is an evolving science and it seems like every year there is something new that is developing where part of the science of quality improvement is emerging or clarifying or becoming clearer.” (Physician-Author 25)

While this scientific field emerges, it might be difficult to isolate exactly how some elements of quality interventions should be reported. Participants (25) said that complexity, rationale, materials, fidelity and modification, outcome measurement and active ingredients should be described in all published accounts of QI, and although these reporting items have been widely encouraged in other types of scientific writing such as clinical trials, they can often difficult to fully articulate in QI. Participants suggested that some aspects of quality intervention description are distinctly different to other scientific fields, such as the precise method by which active ingredients should be identified, and the rapidly mutating nature of quality interventions which are frequently modified.

5.7.1 Complexity

Twelve participants recognised that *“QI and patient safety [interventions] are always more complex” (Academic-Author 54)*, but that ‘complex’ is a “loaded” (Academic-Author 01) word and different types of complexity should be fully explained. Participants reported that complexity can mean:

“The way it’s developed (its rationale and what it actually is), the way it’s delivered, what its individual steps and components are, and the more subtle changes that are made along the way.” (Academic-Influencer 12)

Complex interventions can include many different “people” (Nurse-Consumer 10), “components” (Cardiologist-Author 25) or stages, rather than one “defined event that is being judged” (Surgeon-Consumer 33) and this can be difficult to properly describe, implement and replicate.

5.7.2 Rationale

Twelve participants raised intervention rationale as a feature which is often reported poorly despite its importance. Rationale, aims and underlying theory appear together in the TIDieR checklist as reporting item 2 (Section 3.3.4 Table 5). This item was fully complete in 98% of the perioperative QI papers identified during my systematic review (Chapter 4, section 4.3.1). Davidoff et al³⁵⁸ explain that different types of theory can be used in QI, such as grand, mid-range and programme theory. Participants did not name any particular types of theory but they explained that having a rationale relates to:

“Having a reason for an intervention, being clear as to why every bit of the QI intervention was implemented . . . whether it was literature guided or theory guided.” (Academic-Influencer 12)

Participants said that rationale, aims and theory should be documented right at the start of a QI study because *“the why you did it part is almost as important as what you did” (Surgeon-Editor 16)*. Participants said that QI studies which include a complete, *‘thoughtful’ (Surgeon-Influencer 11) and ‘robust’ (Anaesthetist-Editor 17)* account of the rationale were examples of good reporting because this can help the reader to understand how the intervention intended to address the identified problems (*Physician-Editor 53*). These 12 participants gave two reasons why reporting of intervention rationale is poor. First, authors find it difficult to analyse what the problem is. Second, editors fail to realise the importance of the description of rationale, which they may regard as *“taking up too much space” (Physician-Editor 65)*.

5.7.3 Materials and procedures

Thirteen participants said that the reporting of the materials (which are physically used to deploy an intervention) and procedures (the activities or processes used to implement the materials and deliver the intervention)¹⁵⁰ required to implement an intervention should be reported. However, they recognised that materials and procedures are frequently inadequately described and they expressed frustration about this.

“I want to know the specifics! . . . sometimes it is a little vague – oh we used a validated blah – I’m like, can you show me what that blah was!”
(Surgeon-Author 58)

Participants were frustrated about the poor reporting of materials, first, because they wanted to be able to use the information to replicate the intervention themselves and second, because including examples of the actual materials used, and what procedures were required, is a means of allowing the reader to grasp what how the intervention was delivered more quickly.

“When you give me an example I’m able to focus in on it very fast”
(Anaesthetist-Author 47).

Three participants said that materials should be easy to report because:

“That is just a matter of describing them and providing access to them”
(Academic-Influencer 12).

However, the majority of this group of 13 participants recognised that providing *“granularity of detail”* (Surgeon-Editor 16) can be very difficult, first, because it requires a lot of *“intellectual labour”* (Surgeon-Influencer 11) to create the materials which authors may not wish to release for free use by everyone else. Second, the *“extra effort”* (Academic-Author 54) involved in describing materials may not be fully relieved by the provision of an online repository for supplementary resources.

Participants wanted to be able to read a complete account of an intervention, but they weighted the importance of intervention description differently. One participant explained that intervention description was *“number one”* (Surgeon-Author 14) and the most important thing to focus on to improve reporting in QI. Conversely, another participant explained that intervention description is not weighted as highly as the reporting of data analysis.

“In terms of weighting how I recommend for acceptance or rejection, I’m basing that decision more heavily on whether I think that analysis is robust, because the description can be improved, whereas the analysis is often harder to improve.” (Radiologist-Author 24)

5.7.4 Fidelity and modifications

Fidelity is the extent to which a treatment is given in the way it was planned.²³⁰ Recording fidelity helps the reader to understand whether a significant or insignificant result was due to the intervention, or due to factors added into or omitted from the intervention.²³² In experimental research, low intervention fidelity can introduce performance bias because scientific confidence that the results are due to the independent variable (the intervention) is reduced.³⁵⁹ Few participants (4) used the word fidelity without any prompting, but many of them discussed modifying the planned intervention.

Sixteen participants recognised that QI can be difficult to report because quality interventions are frequently modified through several iterative cycles as the study progresses. Modification and the absence of typical ‘clean’ fidelity was used to distinguish QI from clinical trials.

“That’s one of the differences between improvement work and natural science. In natural science, particularly clinical trials, you have to stick to a protocol. If you deviate from it it’s considered a deficit and it weakens the study. But in improvement, making changes in what you’re doing, is exactly the whole point.” (Academic-Influencer 64)

Participants reported that modifications are hard to report because modifications were associated with QI being a ‘dirty business’ (*Academic-Author 54, Physician-Author 30*), grounded in the messy reality of everyday life. This conflicts with the value of ‘clean’ high fidelity interventions which are highly prized in clinical trials and participants found it hard to resolve this tension. Participants concluded that trying to conform to a traditional scientific perspective which values high fidelity would not be the right thing to do. Instead, the frequent changes implemented during QI work should be seen in a positive light because of the beneficial effect it can have on patient care.

Emphasis on the type of modification seen in QI, differs from other types of scientific endeavour, and may be unique to QI. For example, modifications observed in clinical trials tend to be implemented to protect patient safety or surgeon autonomy:

“If somebody says, ‘well we’re not going to transfuse a patient unless the haemoglobin is less than seven’, you would get a lot of arguments from people.” (Surgeon-Author 26)

Other modifications, such as the deliberate and constant changing of an intervention in response to new data during iterative cycles of measurement, may be more specific to QI.

“QI doesn’t achieve success by having everything decided up front and controlled, it achieves success by cycling, by having iterations of new interventions and new tweaks, new adjustments to try to make things better as they are spread.” (Radiologist-Author 24)

Participants recognised that modifications are *“very rarely reported,”* which is consistent with the finding of my systematic review (Chapter 4, section 4.3.1) which shows modifications were incomplete in 73% of papers. Participants raised the point that modifications can be purposeful, and conducted in response to cycles of data collection, but they can also be unexpected, in response to practical challenges identified during the course of the study:

“As soon as you start doing the project you’re like ‘oh! There’s this barrier, and we realised we needed to approach this differently” (Academic-Author 01).

Unexpected modifications can be problematic for reporting because they may require the author to admit *“that didn’t work for us so we had to change it” (Academic-Author 01)* and because authors *“don’t keep track of it very well through the passage of time” (Surgeon-Influencer 11)*, so information about modification may be lost. In addition, authors may feel that writing up expected and unexpected modifications in a final report may not be publishable.

“There’s a lack of scope for that kind of reframing of problems as your project progressed.” (Academic-Author 01)

Modifications may also be rarely reported because there are often so many small changes that if every detail of what was done was recorded:

“You could probably write a book, but nobody is going to read it. There is only so much you can do.” (Surgeon-Consumer 29).

Participants offered several solutions to the problem of inadequate reporting of intervention modification including: greater encouragement from journal editors and those providing QI training to document and ‘track’ the refinements which are made to an intervention over time, and the degree of impact this has on patient care (*Physician-Author 30, Academic-Influencer 64*).

5.7.5 Outcome measurement

Twenty-one participants discussed the issue of outcome measurement reporting in QI. They said that reporting outcome measurement can be hard because it is difficult to know which ones to select in the first place, because quality can be defined in so many different ways, and because some things can’t be measured (or are very difficult to measure).

Only two participants reported that measurement was usually good in quality improvement reports. The majority view, held by nineteen participants, was that outcome reporting is particularly hard in QI because it is an emerging field with poor consensus on which outcomes should be selected:

“Measurement is one area where the vast majority of papers fall down and that’s partly analysis but it’s partly just having metrics that are valid, sensitive to change and indicative of quality.” (Radiologist-Author 24)

Some participants recognised that outcomes can relate specifically to quality:

“Donabedian described the different components or domains of quality as being about: efficacy, effectiveness, efficiency, optimality, acceptability, legitimacy, equity. So you pick one of those domains to measure and say does it [QI] impact it?” (Surgeon-Author 26)

Participants also acknowledged that quality can be defined in different ways by different people, and a wide array of outcome measures which relate to different opinions of what quality comprises of can be selected. For example, investigating whether or not compliance with DVT prophylaxis was achieved

may be understood by evaluating compliance, as well as measures which relate to quality of life and cost effectiveness (*Surgeon-Author 26*).

“You can't get away with just one outcome, you want much more to be able to say that ‘this is what we wanted to improve, it improved and it improved in the way that you would expect based on these other things that we also collected.’ (Physician-Editor 53)

Participants said that it is important to understand “*how they selected the outcome measures*” (*Academic-Author 09*) and whether they are “*reliable and valid*” and “*robust*” (*Nurse-Consumer 38*). However, even when authors explain why and how their outcome measures were selected, they may still face the problem that they may not have measured all of the features which influenced their QI study because consensus has not yet been reached on how to measure the various aspects of context such as teamwork:^{194,360}

“There’s a whole lot more things than we can measure that go into the daily operation of a QI programme . . . you may not be able to measure all of the other things that were going on.” (Surgeon-Author 26)

Participants raised the unfortunate problem that authors are not just faced with selecting measures which are relevant, valid and meaningful for the QI study in question, but they must also be able to reach a consensus on what the outcome is designed to do.

“I was astonished about outcomes and the way that outcomes are described I saw that [groups of surgical staff] had a totally different understanding about the definitions.” (Physician-Editor 65)

5.7.6 Active ingredients

Active ingredients have been described as the contents of an intervention,²⁵² or more specifically, as the components which can create change.¹⁸⁵ The question here is how the active ingredients could be identified in QI, and how the consumer of QI research can know whether the active ingredient should be replicated exactly. Seventeen participants explained that reporting QI is difficult because identifying “*which are the most important ingredients with the greatest*

weight” (Surgeon-Author 26) is particularly technically challenging. Participants provided two reasons to explain why isolating active ingredients can be hard in QI.

First, contextual features (such as organisational culture) can be regarded as active ingredients because they are especially important in QI, and can have a significant effect in one setting. However, although they might be the most active ingredient in one place, it can be hard to identify them (see also Chapter 7, section 7.3) and hard to work out which contextual features are crucial to describe, because it is not always clear whether the same ingredient will be most important in an alternative setting.

“Part of the active ingredient might have been inadvertently the culture or the attitudes of the people in the organisation which you may or may not have somewhere else.” (Academic-Editor 13)

Other participants did not directly refer to contextual features as active ingredients, but they suggested that they might be *“crucial to describe”* (Academic-Influencer 12) or a *“magic ingredient”* (Surgeon-Influencer 11).

Second, isolating active ingredients can be hard in QI because the scientific framework which should be used to present data on what the most active ingredients are and which mechanisms may facilitate their effect³⁵⁸ is not yet firm. Participants did not know how to isolate which ingredients were the most active ingredients.

“[Isolating active ingredients] is not an easy task and part of the science of improvement will be to try and disentangle these issues . . . we do have some way to go.” (Academic-Editor 13)

Participants suggested different mechanisms by which active ingredients could be isolated by QI authors. These suggestions ranged from *“make an educated guess”* (Surgeon-Influencer 11) to more structured suggestions such as: the use of external evaluation to identify active ingredients (Nurse-Consumer 10, Surgeon-Author 14); *“actively study the process of how it happened”* (Academic-Author 09); and using the previous experience of a clinical lead who may instinctively know what did and did not work in their setting (Surgeon-

Influencer 11). Some participants criticised this view, suggesting that *"two guys in an office scratching their heads"* (*Academic-Editor 13*) is not an optimal way to identify the most active ingredient, and argued that further research is required to enable QI researchers to identify active ingredients with accuracy (*Surgeon-Consumer 07, Surgeon-Author 26, Radiologist-Author 26*). This is aligned with the opinion of some QI scholars who suggest that while individual experience and observations can be useful knowledge, it is limited because it is not scientific.⁸³

Participants elaborated on how active ingredients should be reported by suggesting that authors should not claim the intervention was more sensational than it really was (*Academic-Influencer 21*), and they should manage the expectation of their readers honestly. For example, the reader should be warned that it will take time for some active ingredients, such as changing staff relationships, to 'bed in' (*Academic-Author 60*), and they should be provided with a realistic view of what is needed rather than allowing the reader to believe that they can *"just get it off the shelf and pull it in"* (*Academic-Author 01*).

Participants raised the point that the consumers of QI reports should not use a published account of a QI work as a recipe, and they should take some responsibility for identifying which ingredients will be the most active in their own setting. For example, QI consumers should use the experience of their team to aggregate what others think will be the most important ingredients (*Surgeon-Influencer 11*) and ensure that their QI team is able to:

"Recognise it might not be ideal but also see that there might be some benefit so they will give it a go and adapt to [suit] their local circumstances" (*Academic-Editor 60*).

This influences reporting because authors cannot be expected to provide an account that can stand alone and provide all the answers for a QI team in a new setting.

5.8 Discussion: Defining QI and finding a scientific frame for reporting

The pursuit, evaluation and study of QI can be examined alongside the wider field of scientific research, and the overarching theme of this chapter is that the easily recognisable frame of conventional scientific reporting offers an uneasy fit with QI. Participants unanimously agree that all types of QI work (including projects and research) should be written about in a scientific manner. They said that scientific reporting involves brevity and reporting all details of methodological quality (including internal and external validity), which should enable the reader to understand whether the work is legitimate, credible and replicable. However, tension can exist between QI, which can stray from more traditional notions of research, and existing fields of science. Thus the fit between QI and wider scientific fields is not a perfect one.

QI projects and QI research belong to a scientific field which is still in its infancy. Participants identified the problem that QI authors can use a collection of different words to signify the same (or subtly different) meanings, or that the same word can be used to signify greater differences in meaning. Participants also recognised that different scientific fields have contributed to QI and this impacts the language used during writing for publication, and to some extent, where the work is eventually published. Whether the broad collection of terminology is a consequence of a poor agreement on what constitutes QI, or a precursor to it, is not clear, but poor agreement on terminology and what constitutes QI creates confusion about how QI is understood and how it should be reported. It is important to overcome these problems because the conscientious use of scientific terminology has become systematically embedded as a cultural norm in academic organisations such as journals, to encourage the spread and adoption of useful evidence derived from both practice and research.^{15,198}

These problems of language and vocabulary may be a superficial symptom of a deeper challenge: those conducting QI projects and QI research may not yet agree exactly what kind of science QI is, or in fact whether QI is a science or not. Different researchers may have different understandings. For example, some definitions of research state that its defining characteristic is the generation of new knowledge.¹⁰⁵ Participants in this study used novelty (or lack

of it) to define QI, with some participants suggesting that QI work is not usually completely novel, but others suggesting that new and compelling ideas are generated frequently. Participants also suggested that adopting features characteristic of the broader field of health services research (such as multiple sites to aid generalisability, or a control group to increase validity) is indicative of a research study. However, others disagreed, suggesting that studies conducted at a single site can also offer illuminating insights that can be used elsewhere if features such as context are appropriately described.

Participants actively disagreed about many aspects of QI work such as novelty and whether a QI intervention can achieve the same results in a new setting, and these disagreements stem from a major problem: it is not clear what type of science QI is and what kind of science QI should aspire to become. Participants presented QI as a different type of science. For example, rather than creating an exact replica, QI interventions should be re-tested in new settings as part of a scientific jigsaw to eventually extrapolate some elements which can be universally replicated, and some elements which need to be modified.

Participants found it difficult to satisfy the competing priorities of reporting features which are typical of QI while also satisfying what is required of a conventional research report. For example, QI authors may report parameters associated with traditional research, such as bias, while also attempting to write a pragmatic account of what works in their own clinical environment. When discussing generalisability, participants used terminology relating to conventional medical science, such as contemporaneous controls, but they also raised issues which are more typical of QI, such as a desire to report honestly 'what really happened' to demonstrate to their reader what can and cannot be used elsewhere. Some participants (8) associated the QI method SPC, with time series, which also selects different cases at multiple time points.³⁴⁸ They judged the time series approach according to conventional epistemological standards and they regarded it to be a weak level of evidence, but they also considered how this could be strengthened to produce a publishable account of their QI work.

Participants did not know whether QI reporting should fully conform to the norms of other disciplines such as the social sciences, implementation science, and health services research. Disagreements about whether QI is a science at all, and to what extent QI reports should satisfy competing scientific demands to 'fit' into the landscape of existing sciences, can make reporting hard. Thus, rather than trying to force QI reporting into a style it does not quite fit, QI researchers may look towards positioning QI *between* these existing scientific fields with some adoption of some existing norms, or to carve out a completely new field with a new set of values and norms which are unique to QI as a separate scientific field. These key issues will be discussed in more depth in the final discussion of the thesis in Chapter 9.

This chapter has highlighted that many different people are involved in QI projects and QI research. For example, those running QI projects may be clinical staff who are conducting 'real world' projects. Conversely, those involved in writing QI work for submission to traditional journals may draw on an additional repertoire of skills which are used to present the work in a specific scientific, academic format. The interests and concerns of these groups of people could be very different, and this will be explained further in the next chapter, which explores the working lives of QI stakeholders more directly, to consider how the organisations in which QI work is operationalised and published can influence reporting practice.

6 INFLUENCES ON COMPLETENESS OF QI REPORTING

This chapter will report participants' views about what motivates authors to report their QI research, and which barriers can prevent them from providing fully complete descriptions of QI in publications. The analysis exposes the problem that when writing up their work, QI authors may face competing priorities. These influences on reporting can shape the content of the publication that is eventually produced.

6.1 Motives for good reporting

This section explains what participants' value about QI reporting. They identified three main reasons: academic/career reward, patient benefit and reduction of waste.

6.1.1 Academic/career reward

Twenty-three participants (10 authors, eight gatekeepers and five consumers) said QI authors want to write up and publish their work for academic reward. Academic reward can involve *“some sort of personal gain”* (Surgeon-Consumer 68), or the more altruistic benefit of contributing to a wider body of knowledge.

Clinicians and academics alike valued the kudos that publication can bring, using phrases like:

“Personal credit, ambition, glory – all that kind of garbage.” (Anaesthetist-Author 55)

Participants highlighted many types of personal academic reward which varied depending on whether their primary occupation was in clinical practice or in academia. Academics were motivated to produce publications and have some of them published in 'high impact' journals for *“organisational visibility”* and *“academic career progression”* (Academic-Author 01). Participants defined career progression as obtaining a better *“faculty rank”* or an *“ability to get research grants”* (Academic-Influencer 64). Three UK-based academics viewed publication as a requirement by their employing university. The quantity of published work can affect the allocation of research posts and the distribution of funding and research volume has been used as a measure to evaluate and compare researchers.³⁶¹

“As an academic you are expected to produce four, three to four star papers over four years coming up to the next REF.” (Academic-Editor 62)

Seven participants (two authors, three consumers and two editors) suggested those leading the QI projects may not always be traditional academics, and accordingly are not *“driven by the currency of publishing”* (Academic-Author 09). They provided a different set of reasons why these clinicians were motivated to report their QI work:

“To get things on your CV.” (Surgeon-Consumer 68)

“To do QI projects as part of the royal colleges’ exams.” (Physician-Author 30)

“To get our voices heard [junior doctors].” (Physician-Consumer 72)

Two editors (Academic-Editor 49, Physician-Editor 53) explained that *“rank and file clinicians”* (Physician-Editor 53) who do not have a *“compelling incentive”* to publish and use their spare time to try to write up their QI work, could find the task of QI reporting onerous. These editors said that a negative QI study is often more difficult to report because it is hard to explain why an intervention still didn’t work if it was implemented as intended. It can also be hard to explain that an intervention did not work because it was modified (not implemented as intended). Consequently, clinicians lose incentive to publish at all because they may worry that time spent on writing will be wasted.

ELJ: *“What do authors find most difficult to write about in quality improvement?”*

Participant: *“Stuff that didn’t work! [LAUGHS]....I think publishing null studies is always hard and a lot of people don’t do it.”*
(Surgeon-Author 58)

Nineteen participants in total raised the problem that reporting bias (both publication bias and selective reporting bias) exists in QI, and they wanted to address it. They said that the scientific field of QI is well placed to encourage a writing style which allows readers to learn from *“what works and what doesn’t work”* (Academic-Author 01), and this should be encouraged because:

“So much time and effort are wasted by individuals recreating interventions which have been shown not to work.” (Physician-Author 30)

Two editors (*Academic-Editor 49, Physician-Editor 53*) suggested that if aids which intend to improve reporting are further developed, authors of negative studies should receive specific guidance on how to overcome the particularly demanding and technically challenging task of portraying why a QI intervention did not work. This would then give QI authors of both negative and positive studies an equal chance of writing a QI publication which is fully complete and of good quality. Further findings which relate to reporting bias are presented in Chapter 7, section 7.4 during a discussion about selective non-reporting of the negative contextual features of QI research.

Nine participants (four clinical staff and five academic researchers who were: authors (5), gatekeepers (2), and consumers (2)) used the phrase ‘impact factor’ (4) or high or low-impact journals (5). An impact factor is a bibliometric measure which can be used to indicate a journal’s level of prestige.³⁶² A journal can have a two or five year impact factor, and this is the average number of times articles from a journal published in either a two year or five year window, have been cited in the journal citation reports year.³⁶³ An impact factor of 6, for example, signifies that on average, a paper published two years ago within that journal has been cited six times. Four participants perceived that authors are motivated to publish in high impact journals to enable their work to become highly cited and to enhance *“academic reputation”* (*Surgeon-Consumer 32*). Becoming published in high impact journals was regarded as challenging because they tend to encourage more quantitative data, and QI reporting can sometimes require qualitative descriptions (especially of context) (*Academic-Editor 13, Academic-Author 54, Academic-Author 01*). To enhance the impact of their reporting, clinicians and academics (3) suggested that *“young would-be academics”* (*Physician-Editor 53*) who need to achieve publications could be more motivated to use technology and social media. Participants also said that clinical staff who are actively involved in QI reporting might use other measures of impact, such as:

“How many people have read the paper, how many people try to duplicate the intervention.” (Cardiologist-Author 25)

Participants raised the risk that publishing for personal academic reward might have negative effects on reporting. For example, QI authors can sometimes compete with each other.

“The elephant in the room with any research, [is that] there’s vested interests, people want to own the data, get their name first on the paper.” (Anaesthetist-Consumer 37)

This type of competition can cause problems when QI authors who do not want their peers to realise that their QI project did not go according to plan, might edit some information out.

“We hold our cards pretty close to our chest because there is that whole competition thing going on – we like to put forward perfect things that worked.” (Academic-Author 09)

Publishing for academic reward may also cause some information to be edited out in order to satisfy the requirements of traditional research reporting.

“A lot of papers that I’m assessing seem to be written just to get things published and that’s the wrong motivation. I would like to ask the authors simple questions: what really motivates you to the study? What do you think is exciting about it? Why should others think about replication in their setting?” (Academic-Editor 49)

Three participants associated the lure of personal reward with “cheating” (Physician, Author 30); “gaming of the system” (Surgeon-Author 26); or “having the wool pulled over your eyes” (Anaesthetist-Consumer 37). For example, one participant had seen the deliberate selection of sections of SPC data to demonstrate a large effect (Physician-Author 30). Another participant became suspicious of selective reporting when, at regional QI collaborative, he noticed that all of the QI programmes received at least a three star status, which is the highest quality rating awarded by the QI collaborative:

“Somewhere between 5 and 10% of all [anonymised] programmes are positive deviants,” thus, “when 17/17 have three stars, that would suggest to me that there is gaming of the system” (Surgeon-Author 26).

Thirteen participants wanted their reporting to be complete and coherent in order to add to a body of knowledge which could be capable of impacting patient care. Participants suggested that this library of knowledge should be developed by the scientific, medical and healthcare community, to be used by frontline clinical staff, academics and policy makers. These participants said that improving QI reporting should be recognised as *“a call to arms to move the field forward” (Surgeon-Editor 28).*

ELJ: “Why should we bother [to improve reporting]? Why is it important?”

Participant: “Because QI is an emerging field and it’s becoming increasingly obvious that it needs to become, will become, an integral part of the work of healthcare professionals. The sooner we get reporting right in terms of rigour and in terms of replicability, the faster the science will develop.” (Nurse-Consumer 10)

6.1.2 Patient benefit

Thirty-four participants valued well-reported published QI research because it can be applied to change practice and provide patients with better care.

Participant: “It would be nice if you see success in your intervention, that other people will be able to replicate it.”

ELJ: Why is that important?

Participant: “Because we should be able to learn from others, steal ideas and interventions that have worked shamelessly to benefit patients” (Nurse-Consumer 10).

These participants were motivated to improve reporting because: they had been patients themselves (2); they had made a mistake and did not want it to happen again (2); they could see how QI could benefit their own patients (24); and they

had seen published QI work create a positive change in culture in their healthcare organisation (6).

Two participants said that improving reporting mattered to them because they had been patients themselves, and they also wanted to benefit from improved QI reporting. In a third example, one participant told me while the tape was turned off that being a patient inspired him to become involved in QI research.

“One day we will all be patients; in fact, I have been, so we do it for our patients, for our families and for ourselves (Anaesthetist-Author 55).”

Two participants (one consultant and one junior physician) described specific clinical incidents which personally affected them to explain why good reporting is important. They aspired to better, more complete reporting to enhance clinical learning so that the same mistakes would not happen again. One incident involved the participant attending the coroner’s court after antibiotics were not prescribed for a patient. The other incident related to incorrect patient information being handed over during a shift change. In both instances, the clinician was motivated to encourage an *“open learning culture”* (Physician-Consumer 72). They undertook their own QI projects to address the problems they faced and wanted celebrate their success by sharing it with others through reporting.

“We’ve got some really spectacular results . . . We haven’t had a death from sepsis in our trust for about nine months” (Anaesthetist-Consumer 45).

Twenty-four participants valued good reporting of QI in surgery because they wanted to be able to use QI work to benefit patients. Three quotes are used illustrate this point, one from each QI stakeholder group (consumer, author and influencer). Each quote was given as a quick reaction to my question about why improving reporting in the perioperative QI literature is important:

“Because everybody’s interested in improving patient outcomes and making patients’ experiences better” (Surgeon-Consumer 07);

“Because it makes a difference to patient care, number one” (Surgeon-Author 14); and

“Because it matters to patients” (Academic-Influencer 64).

Participants said that to enable QI publications to be used for patient benefit, they must be written in a manner which has relevance for practitioners. For example, the QI research should describe *“similar problems to the problem that you’ve got” (Physiotherapist-Author 05)*. Participants also said that for QI publications to be easily used in practice, they should be presented in a manageable format (see section 6.2.1).

“How can anyone look at 305 references at the end of a paper – it’s nonsensical” (Anaesthetist-Consumer 37).

Finally, participants said that to enable QI publications to be used for patient benefit, the published work should be of high quality, which means it should be written in *“a very consistent way, demonstrably so” (Anaesthetist-Author 47)*.

“Unless the reporting is good, the full potential of QI won’t be realised from the patient’s point of view, and that’s the thing that matters” (Academic-Influencer 64).

Six participants explained that the publication of their QI work influenced the culture of their healthcare organisation, when their QI team received official recognition for their QI publication, and when other clinical teams in their organisation realised that they too could *“do similar things” (Surgeon-Consumer 32)*. Participants suggested that QI publications stimulated a growth in positivity and confidence by demonstrating that their organisation is learning and can shift from one which *“only celebrates success and blames failure”* to a *“continuous learning culture” (Physician-Author 30)*. Participants suggested that the positive culture that publishing QI work creates can fuel future improvement work for the benefit of patients.

“We published it and then people said, ‘Oh, we’re good at that’, and ‘good job’, and it sort of fuels the culture to want to improve” (Surgeon-Author 26).

6.1.3 Reduction of waste

Seventeen participants felt motivated to improve reporting to reduce research waste and to reduce wasted healthcare resources.

ELJ: “Why should we improve reporting?”

Participant: “Because without it, there is so much replication, not replication but reinvention going on everywhere, people start off on roads that if they were familiar with the literature, they wouldn’t start from there . . . we need to prevent research waste, everybody’s wasted efforts” (Academic-Influencer 21).

When a QI report is incomplete, or when it does not reach its intended audience at all (poor reporting by invisibility or poor dissemination), staff in other settings may repeat the same mistakes. Five participants used the phrase *“reinventing the wheel”* (Surgeon-Author 58, Academic-Editor 62, Physician-Author 30, Manager-Consumer 70, Academic-Influencer 21) and they felt frustrated that practitioners often come across the same barriers which were incompletely reported by another person.

Participants recognised that reducing waste can be a key rationale for completing QI work because QI can involve *“reducing costs or saving time to deliver the same standard of care”* (Academic-Author 48).

Three participants wanted published QI reports to explain *how* resources could be used appropriately during the QI work to *“maximise value”* (Radiologist-Author 24) or provide a *“value-added programme”* in their own hospital (Surgeon-Editor 16). The value-added by QI to an organisation is a recognised reporting item which has often been lacking,²⁶ yet value-added activity is a key focus of Lean methodology which is often used in perioperative QI research.³ Participants recognised that a complete QI report should include an evaluation of resources required and consumers of QI work should be able to use this information to reduce wasted time, money and ‘staff capital’ (effort spent on staff engagement activities) during any attempt to replicate or scale up a QI study.

“Resources are part of the context, and resources can and should be quantified.” (Academic-Editor 49)

6.2 Influences on reporting associated with organisations

This section will report participants' views of why some features of organisations such as hospitals can make reporting hard. An organisation can be defined as an entity which has 'social collectivity, organisational and individual goals, coordinated activity, organisational structure, and is embedded within an environment of other organisations'³⁶⁴ (p.11). An organisation's features can be defined under three broad headings: organisational structures (such as skill mix of staff or corporate roles); organisational processes (such as incentive systems);^{365,366} and organisational context. Organisational context has been defined as including eight dimensions: leadership, culture, evaluation, social capital, formal and informal interactions, structural/electronic resources, and organisational slack.¹⁹³ A clear distinction between organisational structures, processes and context has not been easy to locate in the literature and items between these categories can overlap.³⁶⁶ Thus, rather than forcing my data into the categories of structure, process or context, I have presented data in this section using broad headings to explore participants' views on organisational 'features.'

The participants reported that features of hospitals which influence reporting included organisational resources (time and money) and organisational culture. These organisational features can create barriers to reporting and these barriers sometimes required the participants to make choices between competing priorities.

6.2.1 Influences on reporting associated with hospitals

Fourteen participants reported the problem that they needed to report QI work in order to try to influence the organisation they were working in, but simultaneously, the QI work that is needed can be hard to complete because their clinical work load is so heavy.

“If we want money for things to happen, we have got to show other people, the evidence for it has got to be measureable” (Surgeon-Consumer 29).

These participants used words and phrases like *“feeling battered” (Anaesthetist-Consumer 45)*, *“on a hamster wheel,”* and *“wading through treacle” (Anaesthetist-Consumer 37)* to describe the challenge of doing QI work, and writing it up, while simultaneously looking after patients.

“We’ve got in our little group of enthusiastic individuals, but actually getting anything done is quite difficult....it’s like wading through treacle on a number of fronts” (Anaesthetist-Consumer 37).

The influence they wanted to make on their organisation (making a case for their QI work to be supported), and the influence the organisation was having on them (a heavy work load) represents a choice that participants needed to make.

“There’s only so much energy people have, and you have to focus it - they just don’t have enough to divide their time and energy between all these different things that we’re asking them to do, as well as take care of patients by the way” (Surgeon-Editor 16).

Participants had to decide how much time they could dedicate to conducting QI, and then on top of that, writing up a good QI report when caring for patients was their key priority. Although their motives for writing up QI was to benefit their patients (section 6.1.2), the task of conducting and reporting QI often had to come second.

“I am very good at modifying it, editing it ... but don’t ask me to start doing it [the QI report]. That would take me a year to get there I am so busy. I just cannot do it” (Physician-Author 30).

Some participants said that they were *‘self-taught’* and had very little support while they tried to navigate the field of QI which they found hard to understand (*Anaesthetist-Consumers 69 and 45*). Juggling two very demanding tasks (a clinical workload and QI writing) was harder than juggling two tasks where one

is less demanding than another. If QI authors “*already have some kind of knowledge [about QI]*” they might be less inclined to think:

“Oh God, here is more work to do on top of my already busy clinical schedule” (Surgeon-Editor 28).

Thus, when the task of reading literature was perceived to be onerous because it was time-consuming and hard to understand, participants would either opt to spend their time in clinical practice, rather than trying to improve their QI reporting, or spend their own time in the evenings working to improve their understanding of QI.

Participants proposed two ways of reducing the pressure of a perceived lack of time. First, three participants (*Physiotherapist-Author 05, Anaesthetist-Consumer 37* and *Surgeon-Editor 16*) recommended that hospitals should allow QI teams to have protected time away from their clinical work to focus on QI project tasks so they are not “*running around doing different things*” (*Surgeon-Editor 16*). Thirteen participants (six authors, five gatekeepers and two consumers) suggested that this protected time could include access to a structured programme of QI education, such as: mentorship schemes, online web-based programmes of learning, and formal QI education which starts at medical school and continues throughout clinical training. Participants said that this training is needed first, to enable authors first to understand what QI is. To be able to make sense of QI, write it down and portray it to a reader a particular set of skills needs to be learned. Participants said that education should encourage authors to use patterns of thinking which are flexible rather than protocol driven (*Academic-Editor 60*); experiential ‘on-the-job learning’ (*Physician-Consumer 30*); and existing literature to learn how to develop a writing style appropriate to QI (*Radiologist-Author 24*). Participants recommended that education about QI should be “*co-led by clinicians and researchers*” (*Academic-Editor 28*) and that having educational support from the start could enhance how “*publishable*” the QI work becomes (*Surgeon-Consumer 07*).

“Most of the time we are not trained in actually doing QI. ... To improve reporting, medical students, nursing students, therapy students, everybody should learn about improvement” (Physician-Author 30).

Second, 14 participants (seven gatekeepers, four consumers and three authors) suggested that the process of writing up QI work would be enhanced by collaboration between academics and clinicians. They expressed a wish for healthcare and academic staff to work together more frequently. This collaboration could occur by developing “joint roles” (Nurse-Consumer 10), for clinical-academics, or by the creation of a multi-disciplinary writing team (MDT). This MDT could include: social scientists, implementation scientists, systems engineers, clinical staff and patients. Active involvement of clinical staff in this MDT was regarded as “critical” (Physician-Author 30) to produce QI publications which can be “understood by their peers” (Surgeon-Consumer 32). Participants valued the contribution of experts from different disciplines, because individuals who are most appropriately trained to complete a task would be available to carry it out.

“We should not train the quality improvement people to be implementation researchers, because they don’t want to be implementation researchers” (Academic-Editor 56).

Nineteen participants suggested that practical support for clinical staff leading QI work is often informal and scant which exacerbates the struggle to understand, conduct and write up QI work. Six participants reported that the MDT’s or clinical-academic roles which could support QI reporting could be operationalised within local QI units working like local clinical trials units within individual hospitals, or regional QI units operating to support a several hospitals within a geographical area. Participants suggested that other designated QI staff could include: QI nurses (for help with data collection); statisticians; methodological experts; communication experts (for help with the dissemination of QI work); and librarian support (for help with locating the QI literature).

“If you’re thinking of sort of blue sky, I can imagine that you know, in the very same way as we have clinical trials [units] we should have quality improvement units” (Academic-Editor 13).

A lack of financial support was another barrier to producing complete reporting in QI. Some participants perceived that cheaper studies are at risk of informal measurement and a lack of co-ordination which makes ‘roll out’ to other departments and reporting more difficult (*Anaesthetist-Consumer 37, Surgeon-Author 14*). Three clinicians from England’s NHS recognised that financial reward (or avoiding financial loss) can affect whether a QI study is started in the first place. Participants noticed that “*people higher up*” (*Surgeon-Consumer 68*), such as chief executives, medical directors and middle managers, prefer to run QI studies which can avoid fines (such as those generated when ED targets are breached),³⁶⁷ and payments (for the care of patients who are readmitted within 30 days).³⁶⁸ Thus, ideas for QI studies which are relevant to clinicians may never be progress to become studies at all and this could skew the type of QI literature which is available. Even at the conception of a study, QI authors may need to make choices between what they deem to be a relevant and useful QI study, and what their organisation will support.

Six clinicians (four anaesthetists and two surgeons), who worked in England raised the problem that conflict between ‘quality improvers’ (those doing the QI work) and NHS hospital management or the Department of Health (which provides strategic leadership for the NHS in the UK) can often force a QI team to make choices about what can and cannot happen during a QI study. This can affect the type of study that is eventually conducted and reported. For example, participants suggested that QI studies can sometimes feel like they had been “*imposed*” or “*pushed down*” by a “*faceless bureaucrat*” with a “*hidden agenda*” (*Anaesthetist Consumer’s 45, 37, and Anaesthetist-Editor 17*) who may have limited knowledge of “*operational matters*” (*Surgeon-Author 66*). This could be detrimental to reporting if issues that are important to those in management positions are given priority over what is important to practitioners. Thus, the participants reported that it is important to establish ownership of a QI project (by clinicians or joint ownership between clinicians and managers) so that it can be conducted and then reported in a manner which is capable of translating into maximum benefit for patients. Highlighting these conflicts and choices adds depth to a growing picture of why reporting can be so challenging for many.

Two participants said that financial support might be required to use intellectual property before a ‘scaling-up’ project can begin and that this can influence reporting. Intellectual property (IP) (including copyrights and patents) refers to legal ownership of ideas or other intellectual content.³⁶⁹ A QI author might be reluctant to report that IP was an issue, even if it was felt to be a contextual feature of some significance to the study (*Anaesthetist-Author 55*). IP may also cause reporting items to be omitted when authors are unsure about IP and the ownership of information such as the bespoke materials which could be provided in supplementary resources (*Nurse-Consumer 38*).¹³⁸

Three participants (one editor, one consumer and one influencer) said that an obligation to report completely is increased when financial support for QI work has been received from a hospital, or an external funder of healthcare research.

“The funders need to know what they’re funding, and whether the funding has been worth it” (Academic-Influencer 64).

6.3 Influences on reporting associated with journals

Participants raised three features of journals which can make reporting QI in perioperative care hard: rules and regulations, the perceptions and expectations of editors and authors, and locating QI literature.

6.3.1 Rules and regulations

Journals expect individuals who submit a piece of QI work for publication to meet a set of requirements. Participants in reported that some of these requirements give rise to specific reporting challenges. This section will present the participants’ views relating to word count restrictions, reporting guideline use and manuscript format.

Word count restriction is raised in Chapter 7, section 7.5 because 11 participants said it can complicate reporting about context. In addition to context, participants said that word count can restrict the reporting of the QI method (6), rationale (2), intervention (2) and findings (1). Word count was associated with feeling “*confined*” (*Academic-Influencer 21*), “*squashed*” (*Nurse-Consumer 10*) and “*restricted*” (*Nurse-Consumer 38*).

However, participants recognised that some information may always be missing and six participants (three consumers, two authors, and one gatekeeper) used the analogy of a book to complain that if everything was included, the QI report might become too big to wade through and unmanageable.

The endorsement of reporting guidelines by healthcare journals was raised as a possible challenge to good reporting. Journal editors are required to promote a high writing standard and this includes ensuring manuscripts are complete, transparent and as free from bias as possible before they can be accepted for publication.³⁷⁰ Journals have tried to encourage this by endorsing reporting guidelines.¹³⁴ Although 10 participants said that using reporting guidelines led to a higher writing standard, they also suggested that reporting can become more difficult in QI when reporting guidelines are not widely adopted (4) or not yet fit for purpose (6). Some were not sure whether guidelines for QI existed (6).

Four participants (three authors and one editor) were concerned that the SQUIRE guidelines were not yet widely adopted by journals, peer reviewers or authors. Participants did not use the term endorsement, but they talked about journals 'adopting' or 'recognising' the SQUIRE guidelines.

"One of the reviewers said we hadn't used any guidelines, even though we'd used SQUIRE, but he had never come across it before"
(Physiotherapist-Author 05).

Studies examining the CONSORT³⁷¹ and PRISMA²⁴⁹ checklists suggest that journal endorsement does improve completeness of reporting, but another recent study examining nine checklists (including PRISMA) could not link journal endorsement to completeness of reporting.¹³⁴ Participants reported that encouraging the adoption of a recognised set of standards could improve reporting by reducing the variability of writing quality between authors, and the variability of *"personal differences"* (Academic-Author 48) about what good reporting looks like between editors. An additional three editors and two authors suggested that the SQUIRE guidelines could be used to guide the design and implementation of a QI study.

“Once you've done it and you're writing it up you should not be looking at this guideline for the first time. They often get billed as reporting guidelines, but they are also execution guidelines” (Physician-Editor 53).

Six participants criticised the SQUIRE guidelines, stating that they are *“unwieldy” (Academic-Author 48)* and may still need to *“evolve” (Surgeon-Editor 28)*. They suggested that the SQUIRE guidelines needed *“more detail” (Surgeon-Editor 16, Surgeon-Author 66)* such as including examples of good reporting *(Academic-Editor 51)*. These participants recommended that SQUIRE should be re-evaluated every few years *“to ensure they are not superfluous” (Surgeon-Editor 28)*. Some participants criticised reporting guidelines in general, suggesting that they stifle writing creativity, and that the volume of available reporting tools can make it time consuming for authors to understand which ones to use and how to use them. The time it takes to use reporting tools could be counterproductive because the purpose of the reporting tool is to aid reporting, not to hinder it.

“More and more reporting guidelines and tools” are available and “if you want to apply those tools to describe just the contextual factors, it would take you a while just to do that!” (Academic-Editor 49)

Six participants did not know about the SQUIRE guidelines, as became clear when some of them suggested that I should develop guidelines for QI reporting as part of my PhD.

“That’s what I’d be looking for from you, is some sort of published guidelines.” (Anaesthetist-Author 47)

Twenty-eight participants discussed the format and layout of journal articles. They suggested that to make reporting easier, healthcare journals should provide: resources for authors to provide supplementary information (18); adaptations to the existing model of reporting (8); a new model of reporting (7); and reassurance that producing multiple publications for a single QI study is acceptable (4).

Eighteen participants wanted journals to make it easier for authors to provide additional information about their interventions. They wanted journals to

encourage authors to provide readers with extra information by using: appendices, online supplementary materials and website URL addresses in their manuscript. To encourage this practice to become more widespread, two participants suggested that journals should make provision of extra information a requirement.

“If journals required a more detailed elaboration using an online supplement; that would motivate people because they wouldn’t get published without it.” (Nurse-Consumer 10)

However, three participants (*Academic-Author 09*, *Anaesthetist-Consumer 37* and *Physician-Author 30*) were concerned that even if additional information is published, it may not always be available for everyone to use because access requires a journal subscription. Thus, QI authors need to weigh up whether the additional work they put into creating additional material will be worthwhile. If their intended audience might not be able to read it, some may never go to the effort of reporting a fully explicit account of what they did. They suggested that journals should “*extend the open access movement*” (*Academic-Author 09*) which is a global effort to provide free online access to research and scholarly literature.

Eight participants wanted to retain the existing ‘IMRaD’ (introduction, methods, results and discussion) model of reporting in healthcare journals.³⁷² They felt that this “*conventional structure*” works, first, because readers “*know what to expect*” (*Anaesthetist-Editor 17*), and second, because it provides a simple way to evaluate how credible a paper is.

“The traditional journal format is established, it has a tremendous amount of weight and is respected and successful and I think if quality improvement can sit in that model it should.” (Radiologist-Author 24)

Fourteen participants used the word ‘traditional’ during their talk about healthcare journals. The journals they referred to as ‘traditional’ included *The Lancet* and the *BMJ*, which they said use an ‘established format’ which is structured and values traditional evaluative methods (*Radiologist-Author 24*, *Anaesthetist-Consumer 69* and *Academic-Author 48*). Although this group of

eight participants wanted to retain the conventional model of reporting, they recognised that it could be adapted because sometimes getting QI work published can feel like *“fitting a square peg in a round hole”* (Surgeon-Author 58).

Participants explained that the trick to getting published can involve using terminology they had been *“forced to use”* by peer reviewers (Surgeon-Author 66) and:

“Learn[ing] ways of presenting a project that you know more orthodox journal editors would find acceptable.” (Academic-Author 48)

The adaptations to the conventional model participants suggested could be helpful included: closer work between editors and the journal’s readers to identify what is deemed to be acceptable (Academic-Author 09); headings additional to IMRaD such as *“what really happened”* (Academic-Influencer 21); sections within surgical journals for QI (Anaesthetist-Consumer 45, Academic-Author 48); and sections within all healthcare journals for negative studies (Physician-Author 30, Anaesthetist-Consumer 37).

Seven participants suggested that journals should deviate completely from the conventional model of reporting. These participants felt ‘switched off’ from the conventional IMRaD model.

“Some of them, quite frankly you might as well just put them straight in the recycling bin because they're so turgid.” (Anaesthetist-Consumer 37)

They suggested that a new model of reporting could speed up access to information and encourage spread of QI. Three participants (one author, one consumer and one editor) described a hypothetical, interactive online journal which presented information in layers. Starting superficially, this journal would gradually reveal more and more information depending on how much the reader demanded.

“You can easily imagine a more interactive paper where I could click on what is interesting to me and it would pop up with the much more detailed supplementary material, and I could actually look at each Plan-

Do- Act cycle. There would be more media and pictures.” (Surgeon-Consumer 58)

Four participants did not suggest a ‘layered’ approach but felt that reporting would be improved by the provision of more visual aids including YouTube videos and webinars. However, they also realised this could be regarded as controversial and added a caveat to their recommendation: *“Maybe I am being too gung ho” (Academic-Influencer 21).*

Two participants (*Surgeon-Consumer 32, Surgeon-Consumer 29*) felt that the existing practice of submitting papers in a standard format should not be changed at all and found it hard to understand why reporting in QI should be difficult:

“The structure should be easy for people to understand. You should have your strategy, your methodology, your outcome. It shouldn’t be difficult to write it up.” (Surgeon-Consumer 32)

Four participants felt that journals should encourage authors to submit multiple publications for a single piece of QI work so that the details, which may not fit into the primary report, (such as the contextual features and modifications which were particularly important) could still be described. This was born from a frustration at restricted word counts, and a frustration that clinical journals may not accept more narrative descriptions of context.

“In a mainstream clinical medical journal, there is no space for this description [of context]. I think it’s quite necessary that there is a sister publication where people go into more depth.” (Academic-Editor 13)

6.3.2 Perceptions and expectations

Thirteen participants discussed whether journal editors perceive QI to be attractive using words like *“fashionable” (Surgeon-Editor 16), “hot” (Surgeon-Consumer 07), “sexy” (Anaesthetist-Consumer 37, Surgeon-Consumer 68) and “exciting” (Anaesthetist-Author 55, Academic-Editor 49).*

Three participants said that QI is not attractive because journal editors think

“We should just bloody well be doing that anyway . . . just pull your socks up chaps.” (Anaesthetist-Author 55)

They also think that QI is not as attractive as other research, and journals would rather publish reports which offer a novel therapeutic approach, a cure for a disease, or something which is more dramatic than QI.

“Saying you’ve recorded how many people have not done a central line change is not the same as saying that you’ve just clamped off somebody’s bleeding aorta on the street.” (Anaesthetist-Consumer 37)

A few participants reported that QI was gaining popularity with *“a little more demand”* (Academic-Editor 51) and some participants named individuals such as Atul Gawande (Anaesthetist-Consumer 37) (a surgeon practicing in America) and Ara Darzi (Academic-Editor 62) (a surgeon practicing in the UK) who have been particularly influential. However, this view was contested.

“Getting a headline statement and wanting everyone else to follow your lead, where everyone across the world knows what you are doing, is missing the point of QI. It should be about addressing a need that you’ve noticed and just putting your findings out there.” (Surgeon-Consumer 68)

Three participants (one editor, one author and one consumer) discussed how trustworthy journal editors perceive their authors to be. They said that when missing information is inevitable (to avoid the manuscript becoming a ‘book’), scientific writing relies upon an element of trust between the reader and the author. This trust can be assumed because an editor is familiar with the prior work of an author, and they may have good reason to trust or mistrust individual authors in the field (Surgeon-Author 26, Manager-Consumer 70). Trust may also be assumed because the responsibility of reporting accurately and honestly is deeply embedded within the practice of science and research (Academic-Editor 13).

“You assume that when they say they randomise patients to groups, they’ve randomised them competently, right? So we always take certain things on face in scientific writing and I think that’s fair enough, I mean that’s the only way we have.” (Academic-Editor 13)

In their talk about why reporting can be hard, editors and authors/consumers defined the expectations they had of each other and what jobs or tasks they thought the other person should be responsible for. The expectations set by editors of what they require from authors included: manuscript submission to an appropriate journal (avoiding a scatter-shot approach, where authors submit their article to lots of different journals with little thought) (2); use of language and terminology appropriate to the journal's audience (1); and a clear and succinct abstract and conclusion (4). Expectations set by authors and consumers of what they required from editors included: a quick peer review process which includes written feedback (3) and clarity on article length, minimum reporting standards and terminology (6). These participants felt that editors should take a leading role in improving reporting in QI.

"The editors should be the ones who need to really drive this to make sure enough detail is included in the papers." (Cardiologist-Author 25)

Seven participants said that an account of patient and public involvement (PPI) is often expected to be provided in published reports, and they agreed that PPI is a mechanism by which reporting could be improved. They said that patients are clear stakeholders in QI and can contribute helpfully to a QI report to provide a fully complete account of what actually happened during the project, and why the work was relevant to patient care.

"Why would you not ask the patients to be involved? That would be crazy because the patients are there practically more than the staff are." (Academic-Influencer 21)

The participants' comments showing how and why patients should be involved throughout a QI project are shown in Table 11.

Table 11 Patient and public involvement in QI

Phase of QI study	Reason for patient involvement
Conception of a QI study	Identifying research aims (<i>Physician-Editor 64</i>) and ascertaining the likely impact of the project on patient care (<i>Academic-Influencer 21</i>).
Intervention design	Reason not given (<i>Physician-Editor 65</i>).
During a QI project	Patients can aid the staff during implementation of an intervention. For example, “a <i>patient involvement arm</i> ” within a QI study could supplement an electronic reminder system for DVT prophylaxis “ <i>I remind patients to tell us that they should receive a ‘poke’ every day they are in hospital</i> ” (<i>Surgeon-Author 58</i>).
Data analysis	<p>“We are collecting feedback from patients about safety and that gets fed back to teams. We have been really keen to make sure that patients help us with the process of classifying those events, coming up with categorisations” (<i>Academic-Editor 60</i>).</p> <p>Patients may also be “<i>part of audit teams. Their task is to identify how far the department is patient orientated or not</i>” (<i>Physician-Editor 65</i>).</p>
Writing up	Ensuring the written output is understandable by everybody (<i>Anaesthetist-Consumer 37</i>), including patients. We should be aware that “ <i>most patients don’t read academic papers</i> ” (<i>Surgeon-Consumer 32</i>).
Dissemination	To ease relationships between clinicians and patients: “ <i>If the patients understood a bit more about some of the difficulties that we face on a day to day basis, they would then understand what we’re aiming to do, rather than just being annoyed</i> ” (<i>Anaesthetist-Consumer 37</i>).

Although participants felt that PPI can often be expected, and they aspired towards reporting PPI as part of their QI approach, they also knew that conducting PPI properly can be time consuming, perhaps even more so in the emerging field of QI where QI experts cannot agree what to report.

*“It’s hard enough to get two authors to agree about writing something.”
(Academic-Influencer 64)*

If the experts have not yet reached a consensus on how to report QI work, patients may not have a proper understanding of the subject either and adding a patient to the QI writing team may further complicate a difficult task.

“The involvement of patients can be very useful, but it can also be utterly useless and positively harmful if they do not have a proper understanding of the subject [and] cannot detach their own biases and experiences from the view they give.” (Surgeon-Author 66)

Editors and authors had clear expectations of each other, but they also empathised with each other. Editors found it frustrating that *“peer review takes so long”* (Academic-Editor 51) and that confusion can exist at the basic level of understanding what words mean (Academic-Editor 56) (see also section 5.5.1).

Poor consensus between editors and authors on terminology and what distinguishes QI from related fields could affect whether the work is published at all. QI may remain unpublished if it does not ‘fit’ into the journal editors expectations about what is suitable for their journal.

“There’s potentially good work that is not being published because people are making basic mistakes . . . we can only accept a small minority.” (Physician-Editor 53)

Editors empathised with authors by recognising that reporting QI methods can be difficult. QI methods such as Lean or PDSA cycles can be implemented by researchers in many different ways and by different people, and to different levels of quality in the same study, and this can be difficult to record.

“Getting the methodology right from a QI perspective is hard in the first place, but they need to be able to translate that onto the page.”
(Surgeon-Editor 28)

Eleven consumers, authors and influencers empathised with editors and peer reviewers, reporting that it can be *“hard for them to make a judgement”* (Cardiologist-Author 25) because journal editors (particularly of clinical journals) might be very unused to publishing quality improvement.

“We’re only just beginning to get there in trying to get the editors of biomedical journals, particularly the clinical journals to have a clue what QI is about, and they have much less of a clue about what really high quality social science reporting is.” (Academic-Influencer 64)

Participants (Anaesthetist-Editor 17, Academic-Editor 13, Physician-Editor 53, Academic-Editor 56) wanted to improve the quality of submitted manuscripts so that burden on editors and peer reviewers could be reduced. This is especially important when finding peer reviewers can be a time-consuming effort, requiring a lot of resources to co-ordinate (Academic-Editor 56).

Participants (two authors, one consumer and one influencer) thought that the burden on editors and peer reviewers could be reduced by providing them with education to *“demystify the woolliness”* (Anaesthetist-Consumer 37) and equip them with the *“sophisticated”* skills they need to review QI manuscripts (Radiologist-Author 24). Conversely, this could add to their burden, rather than diminish it. Adding extra tasks for editors to complete may not be feasible, when much of their work load is completed during evenings or weekends (Academic-Editor 49). Adding extra work may not be desirable, as editors may understand what QI is but reject manuscripts because of a high manuscript submission rate (Physician-Editor 53).

6.3.3 Locating QI literature

Tasks such as systematic review²⁴ and learning new skills (such as how to report in QI), demand the accurate retrieval of existing QI literature, and participants suggested that the ability to retrieve QI literature was important.

“I’m finding papers now that I wish I’d found a year ago ... I personally quite like to sit and read a couple of examples of how things should be done, just so you’ve got some idea in your head where the goalposts are.” (Anaesthetist-Consumer 69)

QI evidence²⁴ is scattered across a variety of publication types (such as QI research and QI reports) and journal types (such as QI journals, surgical journals, or management journals).²⁸⁸ Eleven participants (five consumers, four authors, and two gatekeepers) echoed this concern, reporting that QI literature can be difficult to find sometimes because terminology can be inconsistent and confusing (section 5.5.2) but also because they did not know which journals published QI work. The ‘brand’ of a journal and whether it can be distinguished as a ‘QI journal,’ can influence how likely it is that a participant would choose to select it to look for QI articles. Participants identified BMJ Quality and Safety as a journal orientated towards QI work (*Physician-Consumer 30, Anaesthetist-Consumer 69, Academic-Editor 49, Academic-Editor 56, Academic-Editor 62 and Physician-Author 55*).

In addition to searching for the literature, these eleven participants suggested that understanding which journals were likely to publish QI work could affect where they chose to submit their manuscripts. They realised that there are not many ‘QI journals’ available and because QI is a hybrid of social, behavioural and clinical scientists, QI authors may have restricted options of where to publish.

“It’s difficult to know where to place that sort of thing.” (Anaesthetist-Consumer 45)

As the field of QI is emerging, it is not easy to recognise ‘QI journals’ but despite this, participants felt that the existence of journals which contain predominantly QI literature, was a positive strategy to collect QI literature in one place. However, other participants disagreed, recommending that QI work relating to surgery should be published in perioperative journals where it will be more easily accessed by a surgical audience (*Cardiologist-Author 25, Academic-Author 48, and Surgeon-Author 58*). These participants wanted QI to be an

intrinsic component of wider (non-QI) perioperative journals to make the QI literature more easily accessible to everyone who might benefit from it.

Participants suggested several solutions to ease the retrieval of QI literature: simplifying terminology^{24,185} (see Chapter 5, section 5.5) and creating a central repository of QI work (7). Repositories of literature in healthcare include services such as PubMed (provided by the US National Library of Medicine (NLM) to enable free access to MEDLINE and NLM indexed papers), EMBASE, SCOPUS, and ProQuest. These repositories contain a comprehensive index of scientific journal articles. Organisations which index or register QI work are important to journals because it helps them to fulfil their function of presenting scholarly material to a wide audience. The participants suggested that the absence of a repository to index QI literature in one place could add to the problems of reporting. A scattered pool of literature could mean that it is hard to find, making it difficult for authors to learn from a field which is rapidly developing. The scatter of literature may also present QI as a discipline which is disjointed across many fields of science.

Six participants (three consumers and three authors) recommended the use of regional (such as QI collaboratives) or national (such as the UK's Health Foundation) organisations. This could improve an individual's ability to locate QI literature by: raising awareness of QI work; providing QI training programmes to bring people together; or providing contact information for editors and authors with similar interests.

6.4 Discussion: Influences on reporting

This chapter has explored the views of QI stakeholders to identify how their own motives, the external influences imposed upon them, and the influences they may need to impose upon others, can affect QI reporting. Participants valued QI reporting for three reasons: academic reward, patient benefit and reduction of research waste. Features of hospitals and journals which had a compelling effect on the actual conduct of QI reporting included: organisational resources (time and money); organisational culture; rules and regulations; mechanisms to locate the QI literature; and the perceptions and expectations that stakeholders have of each other. However, the interests and concerns of QI stakeholder groups are very different, and in order to produce a publishable account of QI work, they have to juggle competing priorities and sometimes make choices between them. This does not simply highlight the problem that there are not enough hours in the day to create and polish a piece of writing, or that work load makes the physical act of reporting hard. The pressures participants face and the choices they have to make are a further visible manifestation of the problem that QI is not yet established as a scientific field in its own right.

The problem that QI projects do not enjoy full scientific status has pervaded the findings presented in this chapter, as well as Chapter 5. Participants identified that for QI work to be well reported, healthcare and academic staff may need to work together more frequently, perhaps to the extent that joint clinical academic roles or QI 'units' are created, which operate as clinical trials units do. The motivations of academics and those who are doing the improvement work may sometimes, but not invariably, conflict. For example, many clinicians were keen to acknowledge their own drive to publish in high-impact journals, and all QI stakeholders agreed that the purpose of QI work should be to make care better for patients. However, QI work (as opposed to research on quality improvement) is often carried out by healthcare staff who are not driven by the currency of publishing and face a day-to-day struggle of balancing conflicting demands, for example, setting study aims that they see as important, and adopting goals set by their organisation; or satisfying the demands of their clinical work at the same time as completing polished accounts of their QI activities. The relationship between clinicians who are quality improvers and

academic staff might be difficult to establish because in the field of QI there is an uncertain boarder between science and service delivery. It is also unclear exactly how QI should operate as a science along a continuum from the hard science of health services research through to local QI projects, or to what extent the values a science like QI should be melded with the values of related scientific fields.

There are challenges associated with supporting purposeful relationships between academia and QI, but much headway has already been made. A broad infrastructure exists globally to support the development of applied healthcare research³⁷³ and the National Institute for Health Research (NIHR) provides infrastructure to support health research in the UK.³⁷⁴ A 'researcher-in-residence' model has also been proposed to stimulate on-site generation and co-creation of knowledge between academics and clinicians,³⁷⁵ to yield publications of greater relevance and utility.²⁶ Similarly, boundary-spanning roles can allow knowledge to flow between different professional groups.³⁷⁶ Although my participants suggest that this type of infrastructure could aid reporting, gaining access to these resources could be hard, especially if QI must first be regarded as research. For example, if a research nurse was allocated time to work with a QI team, additional obstacles, such as the payment of additional treatment costs, would need to be managed.³⁷⁷ The integration of patient involvement which is often expected in research (section 6.3.2) can also be hard to enact.³⁷⁸ Even if obstacles can be overcome, the contribution of PPI itself is difficult to report,²⁷¹ and this could add to the challenges authors already face.

Participants also raised questions about what a scientific frame for QI reporting should look like during their discussions about QI journals and manuscript format. Reporting is hard if the field of QI does not quite fit into the types of journal and manuscript formats which are currently available. Participants separated QI journals from specialist clinical journals, suggesting that QI is a 'hybrid' science which does not fully embrace the values of any one field. Cross-pollination of ideas from different fields such as the social sciences, IS and conventional medical reporting can be seen as a blessing and a curse. Encouraging the flow of knowledge between different disciplines can serve as a

breeding ground for generating new ideas.³⁷⁹ For example, perioperative research has benefited from cross-pollination from engineering (nuclear power and oil rig engineering) and aviation,^{97,380} and this could suggest that perioperative researchers will be receptive to QI, and helpfully contribute to its development. However, the 'hybrid' science of QI may also be problematic because some terminology from the social sciences,⁹² or even QI itself (like the term 'value-added') might be unfamiliar to QI stakeholders. Thus, authors need to make choices. They might aspire to report features which characterise QI, but may have originated in other scientific fields, or to report features which are typical of a conventional medical reporting framework. How authors make this choice could relate to how publishable they think the finished account will be, and how acceptable it will be to their peers.

The process of publishing is influenced by deeply engrained issues such as the degree to which an author is trusted by their peers, and academic reputation. Authors who are driven by the currency of publishing may have much riding on these issues, and so the question of how to strengthen the scientific character of QI is pressing. This raises the question that a trade-off may need to be made. If QI becomes cast in a brand new mould that is unique, and does not fit with the values of traditional clinical journals, it could be a difficult pill to swallow for QI authors who want to make their work known to their surgical peers. Conversely, if QI is cast in a mould which tries to conform to the norms of other fields, the spark which is unique to QI and may strengthen its ability to change care for the better, could be concealed.

In Chapter 5 I suggested that QI may need to develop its own standards of scientific reporting, or carve out a recognisable place for itself which sits 'between' other related fields. This chapter shows that when this choice is made further decisions will need to be taken, such as whether to create 'QI journals', and how to resolve the uncertain boarder between science and service delivery. Each chapter in this thesis feeds into the next, to build a complete picture of why reporting in QI can be hard. Therefore, some findings reported in this chapter, such as the contextual features of organisations such as resources and relationships, will be described in more depth in Chapter 7.

7 THE PROBLEM OF HOW TO REPORT CONTEXT

7.1 Introduction

In this chapter, the problem of how to report context in QI will be presented. Reporting the context of a QI intervention involves 'understanding what happens when a particular QI intervention is joined together with a team, organisation, or health system, through multiple interacting contextual layers' (p.2).³⁸¹ Participants in this study positioned context as an important feature of QI reporting, but they presented several reasons why context is a difficult concept to write about clearly, completely and scientifically.

First, the participants themselves gave varied definitions of context, suggesting that many QI stakeholders find it difficult to work out what context is. Second, they reported that it can be very difficult to notice which contextual features are important during a study, or whether some contextual features exist at all. Third, participants raised concerns that some QI authors might not feel able to be completely honest about what happened during a QI project in a published QI report. Fourth, they suggested that the scientific community may have failed to value the qualitative approaches which may be especially helpful for reporting contextual features of QI studies. Finally, participants perceived scientific uncertainty about how context should be reported, and they explained why it is important to get this right so that successful interventions might be more easily transferred from one setting to another. Understanding why these problems exist helps in explaining why reporting is hard in QI.

7.2 Defining context and understanding what to report

All 42 participants spontaneously raised context as a problem when asked: 'What is difficult about reporting in QI?' The majority (38) of participants used the word context or a close synonym for it such as 'setting.'

Twenty-eight participants raised the problem that authors may not understand what context is and that QI authors, editors and peer reviewers may lack the knowledge which is required for this type of writing. Knowledge is defined as how well an individual understands a concept.^{382,383} Participants reported that attempting to define contextual features and describe them in detail is difficult because the contextual features of a study tend to be 'amorphous', 'dynamic',

‘ethereal’, ‘nuanced’, ‘intangible’, ‘subjective’, ‘fuzzy’, ‘mushy’, ‘implicit’ (not explicit), ‘nebulous’, ‘qualitative’, ‘hard to describe’, and ‘softer’. Contextual features, being difficult to describe, are often relegated to a ‘black box’.

"There are a lot of factors that impact on the outcomes of those processes and a lot of these factors people put in this big black box we call context." (Academic-Editor 13)

Most of these 28 participants said that authors should report the contextual features of a QI study so their readers can *“know what was happening”* (Academic-Influencer 21). They reported that ‘knowing what was happening’ during a QI project requires descriptions of how success was facilitated but also what might have caused barriers. The contextual features which can impact success or barriers included: leadership, buy-in, culture, relationships, teamwork, resources, expertise, management, communication, environment or setting, and behaviour. Participants described some of these features in further detail, for example, *Academic-Editor 49* described buy-in as:

“Where the spark was, what motivated people to do it, what really happened, why was it accepted and how were they able to keep it pushing along.” (Academic-Editor 49)

Four participants (two editors and two surgeons) talked directly about scaling up QI projects in a new setting. During ‘scaling up’, modifications to interventions which have previously been tested elsewhere and ideas on how to implement them are assembled in a new place.³⁸⁴ These participants suggested that to allow scaling up to happen, the cost and quantity of resources should be reported alongside (or as part of) the context:

"So context and resources – these would be two takeaways to report on, if you want to inform scalability of a quality-improvement project." (Academic-Editor 49)

This initial portrayal of context as a black box reflects the problem that QI is a young field: it is not yet clear to stakeholders how features which are important in this field (such as context) should be defined.

Three academics (who were also clinicians) referred to a QI intervention as a 'behaviour change intervention' during a conversation about reporting context, suggesting that strategies to improve buy-in are deliberately and thoughtfully introduced. The perception that QI often involves a thoughtful introduction of behaviour change highlights the issue (raised previously in Chapter 5, section 5.5.2) that QI is a field which draws from many other disciplines such as behavioural science.

Participants said that reporting context should include a description of barriers in the setting in which the QI intervention is operating, or specific events which stimulated a helpful (or unhelpful) influence on the QI work. This could include barriers related to the organisation:

"It might say this was done in a rural hospital, but then it doesn't say under threat of closure in the next six months." (Academic-Influencer 21)

It could also include barriers related to the people working within an organisation:

"I know a person who would repeatedly tear off the wall a poster related to a particular QI project because he didn't believe in it. This person was a consultant surgeon. A senior consultant surgeon. So it was very unprofessional thing to do and, of course, massively undermined the project – it completely sort of disrupted the whole thing." (Anaesthetist-Author 47)

The diverse manner in which context is described in the literature is a known barrier to learning about context and being able to describe it adequately³⁸⁵ and the need for common language and terminology was raised as a major theme in Chapter 5. A set of domains (the MUSIQ tool) which can aid description of context already exists³⁹ but they were discussed by only three people (two authors and one editor). Participants said that while the use of CONSORT for reporting RCTs has become well established, the reporting of context is less intuitive, and it can be hard to remember exactly what should be included. Nine participants suggested that establishing reporting guidelines for contextual features would be helpful to act as an aid memoir, and to provide structure and

an element of control, reducing the temptation to provide too much detail on some items at the expense of others.

“Using those guidelines are helpful because it actually pushes you to make sure you think about things that you might not otherwise have written in the paper.” (Academic-Editor 62)

Conversely, some participants warned that using reporting tools could stifle creativity by “straightjacketing” people into one way of thinking (Academic-Editor 13). Participants were also concerned that trying to use new reporting guidelines on top of navigating a sea of literature about context, would be hard and could exacerbate poor reporting.

“The implications of asking authors to do this – use more and more reporting tools, is not really consistent with the requirement to summarise your research in 2,000 words.” (Academic-Editor 49)

Six participants (two editors, two authors and two consumers) also said that context can be hard to report because it is not usually presented as a taught subject during clinical training, where emphasis is typically placed upon experimental science (such as RCTs).

“Although I was very interested in organisational behaviour and culture, I didn’t really have the kind of understanding that allowed me to put it as an observation on a piece of paper. Most people don’t have that skill.” (Physician-Consumer 30)

Participants suggested that when authors have been trained predominantly in classical experimental science, they tend to “roll their eyes” (Surgeon-Author 26) when they are asked to report contextual features such as teamwork or communication because they cannot easily measure or summarise them.

“As a clinician, if you’re reading something and there’s a measurable change at the end: that makes it easy to digest.” (Anaesthetist-Consumer 37)

Eight participants (four editors, two authors and two consumers, of whom six were clinicians) reported that a method of measuring or quantifying the

contextual features of an intervention would improve reporting. Quantifying context could simplify the data, encourage brevity, and aid the process of analysing contextual data, which could otherwise be unwieldy. However, participants realised that not all contextual data could be measured, and one participant paraphrased a famous quotation of the physicist Albert Einstein to illustrate this point:

“Not everything that can be measured is important; not everything that’s important can be measured.” (Surgeon-Author 26)

7.3 Noticing contextual features

Five gatekeepers, three consumers and two authors identified that context may not be reported at all (poor reporting by omission) because the author may not realise the contextual feature was important, or that it existed. Corridor conversations, chance meetings of charismatic personalities, a changed team member, simultaneous unrelated work in other departments, arguments with different departments, or board level decisions can contribute towards the success of the QI study, but these factors may be evade the attention of even the most diligent QI researcher. Some participants regarded context as something that *“goes on around the intervention” (Academic-Editor 49)* as opposed to being a deliberate ingredient of the intervention itself.

“Everything that actually happened in the sort of underneath might not be as obvious, so you might not report that.” (Academic-Author 09)

Four participants recognised that teams leading QI research can involve a lot of people. Large teams can mean that *“a lot can be lost in translation” (Surgeon-Consumer 33)*, especially if contextual features are hard to identify in the first place. It may also be difficult to notice contextual features if:

“You’re too ‘in it’ to realise that it’s something about the place you work in . . . you’re almost too close.” (Surgeon-Influencer 11)

Contextual factors may be missed because searching for all possible contextual features with a vigilant and reflexive mind set is hard work.³⁸⁶ Fourteen participants talked about the burden of their hectic clinical schedules (see Chapter 6, section 6.2.1). Four of these participants said that their clinical

workload made it more difficult for them to put time aside to focus on a task they found hard – writing about contextual features.

“I should have kept an accurate record of what was happening, but it’s quite a lot of work when you’re trying to work a busy job at the same time.”
(Physician-Consumer 72)

7.4 Reluctance to write candidly about context

The participants recognised that it is important to report the aspects of a study which did not go well, but they felt reticent about doing this. Almost a third of participants (four authors, four gatekeepers and five consumers) raised concerns that they would feel uncomfortable writing about how individuals or institutions had negatively influenced their study. In this chapter these influences are called ‘negative contextual features.’ Participants gave accounts from their own experience as examples.

“It can be a sentence that you had a clinical lead that’s strong, an active clinical leadership blah, blah, blah. But we both know how that meeting is delivered can either be really charismatic and energising for the team or it can be on the other extreme, tantamount to bullying and finger pointing and a completely destructive sort of culture.” (Physiotherapist-Author 05)

The negative contextual features raised included: bullying, finger pointing, senior executives feeling powerless to enact a change, QI interventions being tampered with (e.g. posters torn down) and organisational pressure to avoid ‘bad marketing.’ Two participants (one surgeon and one academic), talked about surgeons they knew personally to identify negative contextual features. They suggested that for an individual to flourish as a surgeon, they might require certain personality traits, but while these personality traits can be helpful in the field of surgery, they can be an obstacle to QI work.

“They don’t want to be told how to do things or what to do.” (Surgeon-Editor 16)

“They’re in control, total control.” (Academic-Influencer 64)

Participants knew that negative contextual features were important because they could influence the implementation of the QI intervention, but they did not know how to write about them in a professional and courteous way. Participants were anxious to protect their reputation and future working relationships with others. Participants also wanted to protect others as well as themselves, and they raised concerns that deliberately seeking out and then publicly reporting this kind of data might feel threatening or uncomfortable for the staff involved in the QI study. Finally, participants said that negative contextual features cannot be reported if they are invisible to the QI lead because staff are not given the opportunity to speak out, or feel they can't speak out because nothing would happen as a result. Thus, many of these negative issues are swept 'under the carpet', creating poor reporting by omission.

"All under the carpet . . . people don't want to say the chief of surgery was an idiot and we had to get the hospital president to sit, make him agree to this." (Surgeon-Editor 16)

Participants felt uneasy about being 'brutally' honest about negative contextual features because they wanted to create the 'best sell' for their work. Creating a positive 'spin' on a study is related to the issue of selective reporting bias and the perception that negative results are harder to write up.

"You're always trying to give the best sell for your work, and so I think that inevitably leads you to frame your ideas in a certain way because you want to get published, so there needs to be some positivity there. I think that may sometimes prevent you from being completely brutally honest about how things actually were." (Nurse-Consumer 10)

7.5 A failure to value qualitative approaches

Participants reported that QI was a 'creative' enterprise which could involve a story telling or narrative writing style (see Chapter 5, section 5.6.2), and this requires the use of words to interpret phenomena. Thus, I use the phrase 'qualitative writing style' to indicate the use of words, as opposed to a quantitative writing style, which may require the use of numbers.³⁸⁷ Participants said that using qualitative methods could aid the reporting of context in QI (see

Chapter 7, section 7.5), but they also recognised that adopting a qualitative writing style might be difficult.

Fifteen participants (seven authors, six gatekeepers and two consumers) used the term 'qualitative' to describe how the experience of another person could be captured, particularly in relation to staff engagement (or buy in) and organisational culture. Other participants described contextual features such as communication, culture and teamwork as:

“Anything that is not easy to measure numerically.” (Surgeon-Author 26, Surgeon-Consumer 33)

Participants proposed that these data could be captured by keeping a QI diary (6), or using ethnography (2). Conversely, two participants said this kind of data cannot be captured, and that instead, the consumers of the QI work should be encouraged to contact the authors, or visit the site where the QI research is operating.

“Visit them, look around, take part of their improvement activities and take that knowledge home with you.” (Physician-Editor 65)

This view was contested by others who said being forced to contact the authors or go and visit could be time consuming and a barrier to learning (*Nurse-Consumer 10, Academic-Influencer 64*).

Participants valued qualitative approaches because it helps them to convey “*the depths of the difficulties in creating change*” (*Physician-Consumer 72*) and to avoid “*thin*” descriptions of what happened (*Academic-Author 01*). Ten participants (authors, consumers and editors) preferred a mixture of qualitative and quantitative approaches:

“These papers are not necessarily qualitative or purely qualitative. You can [also] have very quantitative data on say organisational culture. What I think is important is to get that picture in whichever way the researchers have managed to assess it.” (Academic-Editor 13)

Ten participants (one author, five consumers and four gatekeepers, of whom five were clinicians and five were academics) reported feeling reticent to use qualitative writing styles.

“That probably only gets a sentence in a write-up. I don’t know – I don’t really know what that would gain.” (Surgeon-Consumer 68)

These participants said that authors (particularly those who actively publish in surgery) tend to view qualitative inquiry as an adjunct to quantitative work, rather than a valuable feature of QI reporting. In the field of surgical research qualitative work may be dismissed altogether, because authors perceive it to be “*turgid*” (Nurse-Consumer 10) or “*waffly*” (Anaesthetist-Consumer 69, Physician-Consumer 72) and:

“The kind of woolly stuff that the nurses do.” (Anaesthetist-Consumer 37)

One consumer, one author and two editors explained that editors (particularly of surgical journals) may not feel that qualitative descriptions of context creates useful knowledge (Anaesthetist-Consumer 45, Surgeon-Author 55) because their classical training in experimental research methods teaches them that context is ‘noise’ which should be controlled out, or that variables which are hard to measure (such as context) are less desirable. Three editors were resolute about their reluctance to publish papers which require in-depth descriptions of local contextual features and do not resemble an experimental or quasi-experimental study design. This helps explain why reporting is hard because authors may feel they do not have permission to report context. Thus, contextual details are often the components of a QI report which “*get edited out*” (Academic-Influencer 21).

Finally, both authors and gatekeepers are trying to conform to a set of rules and regulations which include an allocated word count. Authors (Radiologist-Author 24 and Surgeon-Author 25) and Gatekeepers (Academic-Influencer 21, Academic-Influencer 64, and Academic-Editor 49) said that weighing up a desire to include detail about context while keeping to a set of concise statements which fit an allocated word count can be challenging. Participants did not know when a description of context “*is too little and when is too much*”

(*Anaesthetist-Consumer 69*). They said “it’s a fine line to walk” (*Academic-Influencer 64*) because if you provide every contextual detail the writing could be almost infinite, but it is hard to choose which details to include at the expense of others. For example, it is difficult to describe all the details of a system which might be alien to an international reader with a limited amount of space.

"I can report about NSQIP, I would send it over to [anonymised British journal], and people would say well we don't know what that even is. I have to spend most of the time discussing that as opposed to the project." (Surgeon-Editor 16)

Both editors and surgical authors are trying to fit QI reporting into a set of existing values about whether qualitative writing is useful or not, and how long an article should be. This indicates that a degree of incongruence exists between the conventions of medical publishing and the field of QI. The need to report details such as context makes QI different to other types of surgical research and this can make reporting hard.

"QI is difficult to report at the moment because I think it's bucking the trend of most journal preferences in terms of what knowledge is or what science is." (Academic-Author 01)

In contrast to the majority view of participants (26) who explicitly stated that context is hard to report, two individuals (*Anaesthetist-Consumer 45* and *Academic-Author 48*) felt that context should not be hard to describe. Instead, they felt that incomplete reporting of context stems from journal editors, who think that context is not important knowledge, and from QI authors who are not trying hard enough to use literature which already exists to help them improve their writing.

7.6 Encouraging scientific writing about context

Participants reflected on the problem that the reporting of context does not always have a high scientific profile with journal editors, and they drew on their knowledge of traditional scientific approaches to work out how this problem might be alleviated. Twenty-four participants (10/15 authors, 4/12 consumers,

and 10/15 editors) said that context was an important reporting item because it shapes the reader's scientific understanding of generalisability and causality, but they found it difficult to identify how this should be reported in a scientific manner.

7.6.1 Generalisability, replication and scaling up

Chapter 5 introduced the view that QI stakeholders wish to know which ingredients of a QI intervention are generalisable and which are not (section 5.6.3.) This section will present the views of 22 participants who talked directly about the reporting of context and how this affects generalisability, replication and scaling up. I will now draw on the literature to provide a definition of each of these concepts in turn. In order to generalise results, the study population must be representative of the 'real world' population that a tested intervention intends to benefit.²⁹⁴ Measureable features such as participants' age or their number of co-morbidities are often used to help determine whether a sample is representative. Generalisability, replication and scaling up are not synonymous terms, but they all relate to getting the intervention to a population that would benefit from using it. A generalisable study suggests that the characteristics of an intervention and/or population can be the same when an intervention is applied to another setting. Replication is a process by which an intervention, which has been tested elsewhere, is implemented in a new setting.³⁸⁸ Scaling up is also about implementing an intervention in a new setting and it relates to how local issues of infrastructure and context are addressed during this process.³⁸⁴

Journal editors (4) said it is harder to publish studies which place a lot of emphasis on qualitative descriptions of context because although contextual features can help to explain how an intervention works, they are very often unique to each setting. A non-generalisable study is less valuable to an editor because the consumer of the QI work is likely to need to repeat the intervention evaluation all over again somewhere else. QI consumers (*Surgeon-Consumer 29 and Physician-Consumer 72*) and editors (*Physician-Editor 53, Academic Editor 49, Academic-Editor 62 and Academic-Editor 56*) said that when QI studies need to be repeated anyway, the function of a QI publication becomes

more about sharing ideas, rather than writing for exact replication. One view is that when QI reporting involves sharing ideas, details about context are less important because the reader is less reliant on exact descriptions to understand exactly how the intervention could be transferred to another setting. Other participants disagreed (*Nurse-Consumer 10, Academic-Author 05, Cardiologist-Author 25 and Academic-Editor 13*), suggesting that “*more elaborate descriptions of the local setting, local context*” (*Academic-Editor 13*) are important, even when a study is conducted at a single site.

"If you have the detail at least you can see that it is something we could do in our location, or I like the process and maybe we could change it. Knowing the detail can allow for assessment." (*Cardiologist-Author 25*)

Conversely, the editors said that even when a full description of context is provided, it can still only be used to inspire ideas, rather than contribute to understanding.

"There's so many different ways in which even something as basic as a checklist can be done,they're going to essentially have to do the same thing the authors did to check for implementation problems and resolve them." (*Physician-Editor 53*)

This lack of agreement on how much detail should be allocated to descriptions of context, and what these descriptions should be used for (replication or the generation of ideas) creates difficulties. If stakeholders perceive that a description of context cannot aid understanding about the generalisability of an intervention, authors may perceive that descriptions of context will be less publishable and may need to “*go to and fro a couple of times*” (*Academic-Author 60*) between editors and peer reviewers. Then, QI authors may feel less inclined to invest a lot of time in reporting a concept which is not valued, or they may not report context at all. If context is recognised as being a useful aid to generalisability, it may take time for authors and editors alike to reach agreement on the extent to which exact replication can be attempted, and how context should be reported in a scientific manner.

Five clinicians (*Surgeon-Author 66, Surgeon-Editor 28, Surgeon-Consumer 07, Anaesthetist-Editor 17, and Physician-Consumer 30*) said that exact replication should never be attempted.

"Unless you describe the context of where 'A' is being applied and how 'X, Y, Z' (which you can't control for) may or may not be influencing the final effect called B, it is absolutely dangerous to take that and then apply it somewhere else and we do that in the NHS all the time." (Physician-Consumer 30)

Instead, these participants proposed that the terms 'recreate' or 'reproduce' might replace 'replicate'. They felt that context should be described so that the reader can understand how to adapt or modify the intervention to suit the requirements of their own setting.

"I think it's helpful to have that [contextual] information, not necessarily so that you can replicate it exactly, but so that you could see what they did and see how you might have to modify it in your hospital." (Surgeon-Consumer 07)

Other participants said that replication could be attempted.

"Can be replicated as long as we know what the context was where the patient was and what made it work." (Surgeon-Author 14)

Participants suggested terms such as feasibility or portability could replace generalisability to make explicit that some contextual features can be changed.

"The problem with portability is that because of an interaction with context, you might take the same recipe and insert it somewhere else and experience a completely different outcome and that is because of local soft factors, variances in the people involved, local systems, culture." (Academic-Author 48)

The 22 participants who talked about generalisability and replication expressed frustration that QI is not a recipe which can be translated exactly into another setting. This can cause problems between hospitals who want to utilise QI work which was conducted in the same region, but these problems might be

heightened for consumers who are interested to replicate a QI study which was conducted in another country. Three participants specifically raised the problem that the context of healthcare systems in the US, the UK, the Netherlands and countries in Africa (where the WHO surgical safety checklist was implemented) are very different. Thus, the intervention's use is limited if the context described bears little resemblance to their own location:

"Clearly what you are going to do in your hospital in the UK is different to what I am going to do in my hospital here." (Surgeon-Author 58)

A few participants felt that it is almost impossible to improve descriptions of context so that 'recreation' or 'portability' could be facilitated. They said that regardless of how much detail is provided, it is unlikely that the same intervention can ever be translated successfully into another setting.

"I think it's true to say that the same intervention is unlikely to work exactly the same in two contexts whatever you do." (Surgeon-Author 66)

The majority of this group of 22 participants did describe how the reporting of context could be improved so that generalisability, replication or scaling up could be better managed. First, participants hypothesised that many hospitals face similar issues and therefore, some contextual features are likely to be similar between settings. They suggested that a standardised set of 'similar contextual issues' could be encouraged, to enable consumers to recognise the contextual issues they need to focus on more quickly, thus reducing time spent on assessing the feasibility of replicating or scaling up the intervention in their own setting.

"When we read about QI we should have a clinical perspective of what is similar to our health system and our environment and that would speak to the feasibility and translatability of different efforts." (Surgeon-Author 58)

Second, three participants suggested that a separate external or independent evaluation which takes place alongside the main QI study, should be used to identify important contextual features which might otherwise be missed, and facilitate scaling up activities.

"The independent evaluation work that we have donewe got a sense that they are doing all this fantastic stuff: the way that they talk, the leadership that goes on, and they don't know they are doing it."
(Academic-Editor 62)

7.6.2 Establishing causality

Nine participants (three consumers, three authors and three editors) discussed the problem whereby varied contextual factors, which can change in an unpredictable manner over time, can make it difficult to understand which factors '*influenced*' and which factors '*caused*' an intervention's success or failure. Participants talked about a range of study designs that can be used in QI research (including uncontrolled before-and-after studies and RCTs) and QI projects, and some of these study designs are not intended to ascertain whether one variable *caused* the occurrence of another. Yet, across this whole spectrum of study designs, participants felt that establishing how much an intervention caused or influenced an effect was important so that they did not make false assumptions or create "*false hope*" that one QI study "*is going to work for someone else*" (Nurse-Consumer 10). Participants used the word 'influence' to refer to factors which could have a more tentative relationship with the outcome. They said that these factors may be 'associated' with, or 'attributed to' the outcome, rather than "*definitely causing the improved outcomes*" (Surgeon-Consumer 07).

Whether they chose to use the term 'influenced' or 'caused', participants wanted to know what made the intervention work.

"It's important for the people undertaking the QI programme to be able to identify what the barriers are, what the facilitators are and, more importantly, what are the key features that made the QI programme work." (Surgeon-Author 14)

Participants associated the presence of multiple contextual factors with the 'messiness' of QI and they associated this messiness with the presence of bias. Research texts describe bias as a systematic error which can influence the outcome, and six participants (three editors and three authors) said that

teamwork, culture, communication, leadership, resources, funding and buy-in are contextual features which are capable of introducing bias and could systematically influence the direction a QI study takes.

"They did these things and their mortality reduced in a risk-adjusted way by 40%.' Somebody else said, 'well gee, we did all that and nothing happened.' So, this means you either don't believe the results of the QI programme, or you think the institution didn't really do the things that they said, or there may have been these intangible, difficult to measure things like culture, teamwork, communication that weren't included [in the written report] and could be attributed to the results." (Surgeon-Author 26)

While generalisability is often connected with the concept of external validity, minimising the risk of bias is often connected with the concept of internal validity.³⁸⁹ Two participants (*Nurse-Consumer 10, Physician-Consumer 30*) used the term 'internal validity', in connection to their talk about bias and causality.

"Internal validity cannot be addressed unless the details are there of the environment within which the study was carried out." (Physician-Consumer 30)

Participants suggested that balancing the need to write about context, generalisability and causality can be hard because those writing up QI work might not be *"epidemiologists or researcher types, they are led by QI champions"* (*Academic-Author 09*) or they may have come from a quantitative background with their first QI article being a *"change of direction"* which involves more qualitative work (*Surgeon-Author 58*).

Some participants who talked about causality also talked about probability (they referred to Bayesian methods to illustrate this) and falsifiable results (referring to Karl Popper to illustrate this) (*Editor-Physician 53, Anaesthetist-Author 55*). They were interested to explore how contextual features could be examined using a traditional scientific approach. However, others maintained that QI is a *"halfway house between practice and research"* (*Academic-Influencer 21*).

Thus, it is important to combine ‘narrative’ accounts of context with some of the principles of more traditional scientific approaches, but it will be hard to do this because what is emerging to become recognised as QI does not fit neatly with what is expected in conventional medical publishing.

“It’s a question of how you write that down in a way that’s easy for the human brain to take in, in a structured way such as we are used to, while still explaining and talking people through the context of how you got to where you are.” (Anaesthetist-Consumer 69)

While some participants suggested that contextual features could be important to obtain a better understanding of causality, others did not engage in this debate at all. Rather than regarding context as key features of a QI programme which could help to explain how an intervention worked, two surgeons (*Surgeon-Consumer 32, Surgeon-Consumer 68*) described contextual features such as culture and teamwork as ‘natural variation.’ These are factors which are happening anyway, over which the author has no control.³⁴⁵

7.7 Discussion: The problem of how to report context

This chapter has presented some of the challenges associated with reporting the contextual features of QI studies. A sense of frustration emerged among participants and they suggested QI authors may not always feel equipped with adequate knowledge or permission to write fully about all the contextual features they encounter during QI work. Participants felt unsure that they would be able to notice all contextual features, and they raised the problem that when they did notice contextual features which were negative, it is hard to report them in a professional manner. Participants also suggested that reporting contextual features may require a qualitative approach to writing which may prove difficult if it is poorly regarded by some QI stakeholders. Finally, while most participants recognised that context is important to report, authors, consumers and editors alike could not identify an optimal scientific approach for how context should be described. Participants suggested that accounts of context should be used to allow a reader to make judgements about whether to generalise, replicate or scale up successful studies in other settings and whether the intervention had a causal relationship with the outcome. However, more work is needed to clarify the scientific framework for QI reporting to raise reporting standards.

The data reported here raises many easy-to-identify problems, such as the absence of reporting guidelines for the contextual features of QI, and the disagreement between authors and editors about how much depth an author should go into when recording details about context. These problems are also symptoms of the underlying pathology that QI as a distinct scientific field has not yet become established, and it sits between other related disciplines. The data presented here also reflects another characteristic of QI reported in Chapter 5 – that QI reporting can occur on a continuum, from those who are working as quality improvers on the front line, struggling to obtain buy-in for their local projects, to those who are working towards multi-site QI research which is more generalisable.

The new issue that this data raises, is that QI reporting requires a description of some features which are generic to other scientific fields (such as describing the degree of generalisability) and some features which are specific and unique to QI (such as reporting context). Attempting to describe features which are

specific to QI in a way which is relevant and acceptable to medical publishers, and attempting to report them in a way which is aligned with the character of QI, are sometimes conflicting goals. Using the example of context as a specific reporting feature which is particularly important in QI, it became apparent that participants valued many concepts of traditional science – such as identifying sources of bias and understanding causality. However, they recognised that it is not possible to faithfully report context while remaining completely obedient to a scientific approach which originated in the experimental sciences. To overcome this problem, participants drew from their knowledge of more conventional medical reporting, and what they valued about this (generating a study which could be used to benefit patients at other sites), to present a new image of how specific features of QI such as context could be reported.

Participants demonstrated a desire to adapt a conventional reporting style (such as replacing the term generalisability with portability), in order to tackle the specific reporting features of QI (such as context). Participants said that QI should not aspire to satisfy the harder scientific value that an intervention should be extrapolated from a study and universally applied everywhere. Instead, they proposed that with appropriate scientific description of context (which could include the use of an external evaluation team, for example), decision making about how generalisable a study is and which features may have caused the outcome, might be eased. The projection of a *degree* of obedience to the demands of conventional science while allowing this new science to emerge in its own right indicates what type of science QI might become. Stakeholders have proposed that it would be helpful to move away from the perception that context is fluffy or soft, towards the view that descriptions of context can be valuable to the scientific field of QI. This could represent an optimal solution for perioperative QI authors to whom publishing is important because it could eventually reduce the threat that detailed descriptions of context are less publishable, generate less impact, or are less valuable for career progression (Chapter 6, section 6.1.1).

Actually enacting this balance between conventional medical reporting and the new science of QI is not going to be easy. A garden might be a helpful analogy to describe what might happen next. The garden of QI is beginning to become planted out with herbaceous perennials (such as the familiar tasks of using a control group), bulbs which will come out in the spring (such as outcome measures), and roses (such as the intervention materials), but in the midst of it all sits a large banana plant (the contextual features of QI), which has completely different needs to the rest of this garden. It is unfamiliar to most English gardeners and it needs to be taken out of the soil and looked after in a special environment in the winter. However, understanding what is unique about this garden, understanding how the plants exist together in the same environment, and enhancing communication between different planting experts may then lead to a more confident landscaping attempt. The garden of QI will grow to become mature, with paths that interlink carved out between the many types of plants, and a clear map which shows how it can be navigated.

Currently, templates to encourage the description of clinical interventions (such as the TIDieR reporting guideline and journal manuscript formats) do not fit the unique challenges of QI: a QI intervention is not a stable thing because it can rapidly mutate over several iterative cycles, and neither is the context in which it operates. Many researchers, such as those working as part of the SQUIRE group, have already begun to 'landscape' the garden of QI, but further work is needed to: clarify exactly what sets QI apart from other fields; to get a better grasp on elements of context which have eluded quantification and description; and to understand which aspects of context are relevant enough to merit reporting in outputs with tight word limits. In the next chapter, a practical exercise will be used to further illuminate which reporting features are unique to the separate scientific field of QI.

8 QI AUTHORS AND THEIR AUDIENCES

8.1 Introduction

Chapters 5, 6 and 7 have demonstrated that QI reporting can be difficult because there is a lack of consensus about whether QI is a science and what type of science QI should aspire to become. QI reporting can be complicated by challenges in the field such as practical decisions about where to publish and how to format the manuscript. This chapter explains that QI authors must ensure that what they record in their published manuscript can be understood by their reader so that it can be used in practice.

A QI author writes for an audience. This audience may read the perioperative literature for many reasons, but some may wish to apply what they have learned in their own place of work. This application of learning can take many forms, such as using the article to generate ideas about what they could do in their own setting, or to replicate an intervention exactly (Chapter 5, section 5.6.3 and Chapter 7, section 7.6.1). For this to be successful, the audience must be able to understand the intervention in the way the author intended. Therefore, I designed a study to explore whether participants correctly interpreted a report of a perioperative QI intervention and if not, which parts were misunderstood. The report I chose, by McCulloch et al,²⁶⁸ will be referred to as ‘the Lean article’ and the authors of this article (the Lean authors) are referred to as A, B and C. Lean is a QI method taken from industry which aims to reduce waste by simplifying processes and ensuring errors can be easily visible.²⁶⁸ The Lean article described seven interventions which were implemented using a Lean approach (Figure 5).

I asked participants in my qualitative interview study to read excerpts from the Lean article. Each participant was asked to choose one of the interventions described, and explain what the ingredients of the intervention were, as if they were going to deliver it in their own hospital. I analysed their accounts and displayed 23 anonymised quotations from the transcripts of ten participants in Table 12, and 10 methods by which reporting might be improved in Table 13. These ten participants included five QI authors, three consumers and two gatekeepers. A further 10 transcripts (which had not at that point been coded)

and six audio recordings (which had not been transcribed) were also used, to check that both tables remained representative of the wider dataset. During January 2016, when Table 12 was designed, the most frequent interventions described by the participants were those numbered 1, 3 and 6 in Figure 5, and quotations describing these interventions were selected to populate Table 12. I then shared these tables with three authors of the Lean article and asked them to check whether the participants' interpretation of their intervention matched their understanding of how it was conducted.

Table 12 Excerpts from interview transcripts

	Theme	Excerpts from interviews with QI stakeholders who discussed this QI report: <i>McCulloch et al (2010) Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit. Br Med J. 2010;2; 341</i>
1	Word count/Journal requirements	<ol style="list-style-type: none"> 1. “A journal article doesn’t do it any justice. We hamstring them by limiting them to 3000 words” [09] 2. “We have limitations in what we can publish and how we can publish it, because clinical journals are more used to, I mean it doesn’t take long to describe a drug trial, right?” [13] 3. “If you’re writing up a quality improvement project within a randomised control trial for the Lancet you’re going to be following a more traditional approach which would prevent you from explaining things in as much detail as you might like.” [10] 4. “They’re just going through the steps of what they need to do to publish – ‘I feel I’ve done something important that I want to share, I’ve invested loads of time in it, what are the steps I need to follow in order to get it out there?’” [01] 5. “There are a huge number of journals. Some are very focused on quality improvement, but others are not, and yet all of them will possibly publish quality improvement research. The challenge there, is that the reviewers and the editors of journals may not have a sophistication around some of the challenges and methods in quality improvement” [24]
2	The character of QI work	<ol style="list-style-type: none"> 1. “Quality improvement tends to focus on time series work, but the traditional evidence based medicine pyramid doesn’t recognise a time series study design because that wasn’t the paradigm.” [24] 2. “Quality improvement is more real-world projects, it’s not research. Quality improvement inherently has all the biases. It’ll have clinical biases, selection biases, and reporting bias” [14]

	Theme	Excerpts from interviews with QI stakeholders who discussed this QI report: <i>McCulloch et al (2010) Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit. Br Med J. 2010;2; 341</i>
3	Description of the materials used	<ol style="list-style-type: none"> 1. “A nurse-led exercise to improve inter-shift handover was introduced. The nurses presumably led the exercise I am not exactly sure what the exercise was. It says the bedside handover process but I don’t know what that is. The ‘basic care checklist’ doesn’t tell me what kinds of things are on the checklist or what it looked like or, who uses it (I suppose the nurses use it). And a track and trigger chart, I don’t know what that is” [09] 2. “I’m not sure what the education was on. It says other areas that tackles education but I don’t know what the prompt cards are for or what the education was for.” [01] 3. “A simple protocol was used to allocate roles and responsibilities. But they don’t really describe exactly what that protocol is for or how they improved it” [07] 4. “They talk about using the data to drive the analysis and I’m assuming that that means that they’re feeding it back to people” [01]
4	Replication	<ol style="list-style-type: none"> 1. “The fact that there was a table, with a lot of description for a medical journal, of what was done, that’s what I found useful because I could take this and I could think about recreating a <i>version</i> of this” [13] 2. “I couldn’t necessarily reproduce it, but the key thing for me is that with quality improvement I wouldn’t necessarily want to reproduce identically” [05]
Table 12 - Excerpts from interview transcripts		

	Theme	Excerpts from interviews with QI stakeholders who discussed this QI report: <i>McCulloch et al (2010) Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit. Br Med J. 2010;2; 341</i>
5	Context and second publications	<ol style="list-style-type: none"> 1. “I would be more curious as to how they adopted the implementation or improvement strategy that they used and what challenges emerged for them” [01] 2. “I was still curious that whether or not it went smoothly, they don’t talk about any kind of push back” [09] 3. “The way that it was presented seemed like everything was done very smoothly and they didn't really have any issues. But it’s a little bit surprising, because they were working on seven or eight different interventions at once which could be difficult” [07] 4. “I don’t know for example whether this study that we’re looking at here ever had another publication to describe the local context and issues around how, the reality of how these interventions were actually done” [13]
6	QI Method	<ol style="list-style-type: none"> 1. “It doesn't imply that anything was learned or any changes were made as a result of the learning so I guess that’s why I’m saying ‘did you really do Lean?’” [01] 2. “It’s not about DVT prophylaxis. It’s more about the Lean intervention than it is about the DVT” [05] <p>“It sounds like they were interested in implementing not one particular intervention. They wanted to implement the whole Lean model into the surgical unit. They wanted to, kind of, create a culture of continuous quality improvement” [07]</p>
Table 12 - Excerpts from interview transcripts		

	Theme	Excerpts from interviews with QI stakeholders who discussed this QI report: <i>McCulloch et al (2010) Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit. Br Med J. 2010;2; 341</i>
7	Theory underlying the intervention	<ol style="list-style-type: none"> 1. “7 routine care processes were selected. It is always curious to me how they get selected, so that is not something that they did well” [09] 2. “They maybe could have been a little bit more detailed in talking about how they chose the different interventions” [47] 3. “One of the things that I liked about it was that they focused on the things that the staff were interested in improving, which I think is very important. So I think that’s good advice when you're putting something like this into effect.” [21]

Table 12 - Excerpts from interview transcripts

Table 13 How reporting in QI could be improved

	This solution would make it easier for QI authors to produce an account of their QI work which is reproducible in another setting	
1.	Authors should keep a QI diary so that important information is not lost by the time the final report is written	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
2.	Every journal should allow authors to include web links for supplementary material	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
3.	We don't need any new interventions to improve reporting. We just need people to use the guidelines which are available such as SQUIRE, MUSIQ and TIDieR.	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
4.	Journals should allow larger word counts for QI projects	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
5.	YouTube videos for each QI project should be included as a link within the journal article	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>

	This solution would make it easier for QI authors to produce an account of their QI work which is reproducible in another setting	
6.	Journals should only provide ‘bite’ size summaries of a QI project on a single page of A4, with a link to an open access peer reviewed version of the full QI paper	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
7.	QI research is currently published across many different clinical journals and it can be hard to find it. It would be good to have improved indexing in bibliographic databases.	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
8.	Authors should publish one paper giving the overview of the whole project and at least one further paper expanding on the intervention itself and the contextual features of the work to support replication and scaling.	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
9.	There should be a QI specific data base alongside traditional databases like pub-med to make it easier to find QI literature.	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>
10.	QI teams should receive education/guidance to better understand what is important to report and how to plan their study	<div> <div></div> <div>Strongly disagree</div> <div>Strongly Agree</div> </div>

Figure 5 From McCulloch et al: Table of interventions (p.1044)

Table 1 Study measure and intervention used for each safety relevant process studied		
Safety relevant care process	Study measure	Intervention
1. Direct verbal communication between medical and nursing teams on daily rounds	All ward rounds for general surgical inpatients on the unit were observed at different times of the day. To avoid qualitative problems, the quality of the verbal communication between doctor and nurse on the ward round was recorded in a binary fashion: "Yes" if a nurse was present during the ward round review of a patient; "No" if he or she wasn't.	We conducted a protracted root-cause analysis and a problem identification exercise with representatives of all users. Improvements were made in visual communications on the ward to facilitate staff identification, and a simple protocol was used to allocate roles and responsibilities.
2. Correct administration of prophylaxis for deep vein thrombosis	Use of compression stockings (thrombo-embolus deterrent stockings) and low molecular weight heparin on all general surgical inpatients on the unit was audited against the local protocol.	Fishbone problem analysis was used, followed by solution design. Stickers were put in drug charts to ensure compression stockings and dalteparin sodium (fragmin) were prescribed, and nurses checked that patients were complying. Information was distributed to patients. Four "plan, do, check, act" cycles were used, with refinement and improvement at each stage. Intervention culminated in pre-printed drug chart and self audit.
3. Reduction of drug prescribing errors	An experienced ward pharmacist used a bespoke form to record all prescribing errors identified during routine daily drug chart review in all general surgical inpatients on the unit.	Data driven analysis of prescribing errors identified that a few medications were being incorrectly prescribed frequently. Prompt cards kept on the staff ID card fob were distributed. Other areas were tackled with education alone—with no improvement.
4. Use of alcohol gel for hand hygiene on entering ward	An observer recorded use of alcohol gel by visitors and staff entering the ward.	Simple attempt to improve compliance was made using better visual cues and availability of alcohol gel. No formal "plan, do, check, act" process was done.
5. Correct use of venous site infection protocol	Compliance with local venflon usage and infection monitoring guidelines was audited by ward staff against hospital guidelines for all general surgical inpatients on the unit with a venflon.	A pre-printed check box on drug charts was used to improve reliability, and a checklist handover regimen was introduced to promote compliance.
6. Adequate monitoring of patients' vital signs and recording of their risk scores	Observation charts were evaluated for the preceding 24 hour period and audited against the local protocol. The quality of completion of the "track and trigger" component of the observation chart was scored on a four point Likert scale for all general surgical inpatients on the unit.	A nurse led exercise to improve inter-shift handover was introduced. The bedside handover process was implemented in association with a basic care checklist, which included a track and trigger chart.
7. Adequate completion of fluid balance chart	The quality of the written documentation in the fluid balance chart over the most recent 24 hour period was scored on a five point Likert scale (excellent to poor) for all general surgical patients receiving intravenous fluids or with a documented request for fluid balance monitoring.	Given the high level of pre-intervention compliance, the staff did not consider this process a high enough priority to address.

8.2 Does an audience of a QI article understand the authors' description of an intervention?

The first Lean author interview was completed on 26 February 2016, when 36 participant interviews had been completed. A chart which documents a timeline of each milestone of the SIQINS study is provided in Appendix 10. A total of three Lean author interviews and 39 participant interviews were completed. Five participants completed a shortened version of the task (to read a section of the Lean article and describe the intervention), four due to time constraints and one because they did not provide consent for extracts of their transcript to be shared with the Lean authors.²⁶⁸ During the interviews the participants could not name the journal or the Lean authors, suggesting that the anonymisation was successful.

During the task of describing an intervention, the participants' most frequently expressed concern was that they did not understand the procedures involved in the intervention's delivery, which corresponds to item 4 on the TIDieR checklist¹⁵⁰ (for a full list of TIDieR checklist items see Table 5). This relates to how the materials used were used and what was done.¹⁵⁰ Twenty-three participants found it hard to describe the procedures involved with the following interventions: education; improvements in direct verbal communication; better visual cues for hand gel; nurse checks to ensure compliance with prophylaxis; a nurse led exercise to improve inter-shift handover; and fluid balance. Participants used phrases such as *"I presume"* (Physiotherapist-Author 05) and *"it sounds like"* (Anaesthetist-Consumer 69) to describe what was done.

Descriptions differed between participants. For example, when describing 'better visual cues for hand gel' participants said:

"That means more bottles of gel on the ward that were better signposted." (Anaesthetist-Editor 17)

"That means somebody watched people when they used it or not." (Anaesthetist-Author 47)

Another participant could not attempt the intervention description at all:

“I don’t know what happened, because all they say is we put up visual cues.” (Academic-Editor 51)

In contrast, two participants reported that the description of the intervention procedures *“sounded very clear” (Anaesthetist-Author 47)* and that they *“would do exactly the same” (Academic-Editor 13)*.

The next most frequent concern was that participants disagreed about how well theory, study aims and rationale were reported. Twelve participants said they wanted more information about how and why the safety relevant care processes were selected and a clearer statement of aims (Table 12, row 7). Participants (*Surgeon-Consumer 07, Academic-Author 09, and Nurse-Consumer 10*) recognised that *“staff were encouraged to address the problems they felt to be most important”²⁶⁸ (p.1045)*, but they wanted more information about how the priority setting exercise with the staff was conducted. The participants were looking for deeper explanations about why the intervention was felt to work in order to use the article in practice.

Participants also felt that full details of the *materials* required to deliver the intervention were not available (item 3 on the TIDieR checklist). Eleven participants wanted the materials used to deliver the intervention to be included as an appendix or described in more detail. These materials included: the protocol, the fob cards, the educational material, the basic care checklist, the information distributed to patients, the track and trigger chart, the fluid balance chart, and the fishbone diagram. Participants knew that a fishbone diagram is a type of flow chart used to explore the causes of an effect,²⁰⁴ but they weren’t sure whether the fishbone analysis should be described as part of the intervention or not. Participants also described some difficulty in understanding terminology such as ‘safety relevant processes’, ‘track and trigger’ and ‘data driven analysis.’

Finally, seven participants could not always completely understand who had delivered the intervention (item 5 on the TIDieR checklist), and whom the intervention would be delivered to. They wanted more information to understand *“who did what” (Academic-Influencer 21)*. For example, one of the interventions involved a problem identification exercise with ‘representatives of all users’

(Figure 5, p.1044²⁶⁸), but the participants did not know who those representatives were.

“Presumably that is doctors and nurses, but were there any other people, physiotherapists, healthcare assistants?” (Academic-Editor 62)

When the Lean authors clarified whether the participants had correctly understood their interventions, five major problems arose. First, all three authors said it was difficult to remember what happened. They could obtain details about what happened from memory or stored paper files, but this might be hard.

“Erm... right, well . . . I scratch my head. Did we do that, I can’t remember doing that... I suppose we did. I do have quite copious notes and files and I could find something that would be relevant, but it would be a bit of a struggle.” (Author B)

Second, the Lean authors acknowledged that terminology was problematic for their readers, but they reported this could not have been avoided. They said that when they wrote the Lean article, it was much harder to publish QI. They were “forced” to use a generalised terminology (Author B) because Lean was only just emerging in surgery. Even now QI is regarded to be a youthful field with a wide spread of terminology in use, and work is still ongoing to obtain consensus on how interventions should be described.¹⁸⁵ The problem that some terminology, such as track and trigger (which is now known as MEWS), may change over time was also a barrier the Lean authors could not overcome.

“Back then it was a comparatively new field and it was just starting to break into mainstream clinical journals.” (Author C)

Third, lack of funding was problematic. This could frustrate attempts to write fully because the task of writing demands much time and effort to execute many competing tasks at the same time.

“You are trying to work with the staff, do the science, get published in a reasonable journal, get funding, and now, you are having to think about

the people who might read your work in five or 10 years' time and what it is going to mean for them then.” (Author A)

Fourth, the Lean authors raised the problem that it can be hard to implement QI in healthcare and this could add to the burden of reporting. The Lean authors recognised that a QI intervention is not a stable thing. It is ‘messy’ because involves *“dealing with imperfection i.e. sick people” (Author B)*. The act of conducting QI and then reporting it can also be difficult when the motivations of healthcare managers are not aligned with the motivations of the healthcare staff.

“The hospital board (who often come from fields completely outside of healthcare) deal with overall strategy and what that comes down to is keeping the hospital out of the red and out of the newspapers.” (Author B)

Finally, all three authors reported that they were evaluating how Lean could be implemented in their clinical setting, with the intention of encouraging others to use it. Thus, the purpose of their publication was to enable their audience to learn something more generally about Lean, rather than use it to repeat exactly what they did, or gain an in-depth understanding of theory.

“What the nurses did [in intervention 6 – figure 5] doesn't really matter . . . they locally identified a problem, they developed a local solution and they implemented it using Lean principles. I wasn't writing a paper on how to improve handover. I was writing a paper on how Lean can be applied.” (Author C)

The Lean article was never intended to be reproduced; thus, asking the participants to describe the intervention as if they were going to repeat it was seen to be unfair. However, 11 participants drew parallels with their own place of work and believed they could either replicate exactly or do something similar. For example, *Surgeon-Editor 16* described what normally happens on his ward, to explain what he thought ‘direct communication between medical and nursing teams on daily rounds’ (p.1044)²⁶⁸ meant:

“When I go on to the floor to round . . . I have to go right to the nurses' station and announce that I'm there, and then the nurse, the charge

nurse and my team will then assemble and we'll go round. So that's a concrete intervention.” (Surgeon-Editor 16)

As described in previous chapters (sections 5.6, 5.7 and 7.6.1), writing explicitly to allow an adequate description of what happened during the study, can enable exact replication, or a ‘modified’ type of replication. Both are legitimate uses of a QI publication, but both require the reader to understand exactly what was done. The Lean article represents the latter group: publications which are intended to stimulate ideas for new QI work in other settings, and the author-checking exercise can be used to learn lessons about how this type of QI article can be used in practice.

8.3 Authors and their audience – disagreements arising from the author checking exercise

The main disagreements which arose during a discussion about the intervention description were about the type of replication being attempted, and the type of QI publication. These will be presented in more depth, below.

8.3.1 A blueprint for replication?

The Lean authors felt their article was intended to inspire others to use Lean. They said that a QI intervention is not a basic science intervention and therefore cannot ever be replicated exactly. Thirteen participants noticed that the article was more about describing Lean than the interventions, and the Lean authors agreed that this view was “spot on” (Author C).

“It was never meant for people to go and replicate precisely what we’ve done. We want people to understand what we have done so they can build on it and learn from it.” (Author C)

The viewpoint of the Lean authors is starkly different to the data presented in previous chapters, which demonstrates that while sharing ideas is important (“we should be able to learn from others, steal ideas . . . shamelessly” (Nurse-Consumer 10), it is better for them to be not just useful, but actually used (see section 7.6.1). To do this, the reader of a QI article would require more detail about what happened, and an indication of which components could be adapted.

Participants said that replication of the interventions described would be difficult, but rather than saying replication is not possible at all, they suggested two levels of replication that could be attempted. First, eight participants explained that a 'near-exact replica' could be created which contained many “*very similar*” features. Interventions such as reorganising the drugs cupboards and developing prompt cards for the staff ID fob were deemed to be easily replicable. Second, eleven participants described a method of replication which involved being able to use the Lean article, not just to generate ideas but also to develop a set of interventions suited to their own setting. This would represent an attempt to create a 'modified replica'.

“This article [the Lean article] contributes a framework for the reproduction of the project. It gives you ideas.” (Academic-Editor 49)

Participants reported that to attempt either one of these types of replication, they would need a better description of the interventions, which was constrained because the authors bundled too many interventions together. However, the authors intended to deliver a multi-faceted intervention (not a singular one), and they suggested that their audience should use the article to generate something of their own which was entirely new, rather than an exact replica, or modified replica. Therefore, a QI article can be used to create an almost exact replica; a 'modified' replica; or something almost entirely new. These very different representations of what an article can be used for will be explored further during Chapter 8.

8.3.2 The purpose and type of QI publication

The Lean authors recognised that there are two possible outlets for QI work: a QI journal and a clinical journal. Some of their talk suggested that they held clinical journals in much higher regard. The high esteem of a clinical journal was a good motivator for Lean authors to publish there.

“Times have changed a bit - the QI community now is much bigger and the readership is probably more open-minded . . . you would probably still only get it into a dedicated QI journal.” (Author A)

The Lean authors gave three reasons why their article is well suited to the arena of traditional academic publications. First, they perceived that journal editors, as well as the physicians who would be reading the article, would prefer a concise report which avoids over-elaborate or:

“Slightly random, quirky, opinion based, woolly narratives” (Author C).

Furthermore, the Lean authors felt that a detailed description of intervention elements was not necessary because the Lean article was not intended to facilitate replication. Yet, the participants valued more information over brevity so that some form of replication could be attempted. Some editors also valued the idea that replication should be possible, suggesting that a concise report which is not replicable has a dubious purpose and is less publishable. Thus, the ideal format for a piece of QI research, which satisfies all stakeholders’ requirements, becomes a hard solution to find.

“If it’s something that you could have thought of yourself and you still have to replicate what the authors did . . . because of your local issues, then it’s not clear what service is being performed. “ (Physician-Editor 53)

Second, the Lean authors felt their article was well suited to publication in a clinical journal because it made a novel contribution to knowledge at a time when the use of Lean in surgery was very new. Third, the Lean authors used a proper suite of statistical tools to implement their time-series method, which therefore merits publication in a *“decent journal” (Author B).*

8.4 Participants’ views of how reporting can be improved

Participants suggested how the description of a complex set of interventions, such as the seven described by McCulloch et al,²⁶⁸ could be improved. The participants’ ideas to improve reporting (summarised in Tables 12 and 13) were ‘task-orientated’ i.e. generated by an experience of completing a difficult task which involved describing a complex QI intervention. Consequently, the purpose of any interventions they suggested to improve reporting was likely to be well matched to the situation they intended to improve (the complete reporting of QI).³⁹⁰ During the Lean author interviews, Table 13 was used only

as a stimulus for discussion, to encourage the flow of conversation about reporting. It was not intended to be used as a formal scoring system.

The participants and the Lean authors agreed that a QI diary (Table 13, row 1) and supplementary material (Table 13, row 2) should be used to enhance reporting. The Lean authors expressed regret that they did not keep a diary during this particular project (McCulloch et al),²⁶⁸ because it had worked well during work conducted later. They related the use of a diary to hard laboratory science. Thus, the influence of conventional medical reporting is very important, and can shape how authors choose to describe QI interventions, as well as the solutions that are put into place to improve reporting.

“A QI diary could be similar to a ‘lab book’ – one of the practices of a good scientist is to have a lab book – it’s a basic science skill to keep that diary and to keep the generation of ideas and themes running through.” (Author C)

The three Lean authors also agreed with the participants that supplementary material is important, but they reminded me *“going back 10 years [when this work was conducted], it was quite rare” (Author B)*. They had all used supplementary material in more recent publications and said that this is a *“modern sensible answer” (Author C)*.

Eighteen participants felt that word count could have restricted how much detail could be provided (Table 12, row 1). However, all three Lean authors, and a group of nine participants, reported that that increasing a word count restriction (Table 13, row 4) may not be useful. First, because authors who have been trained to write within the confines of a traditional academic reporting model for such a long time may not be able to write in any other way, and second, because the esteem of the academic journal carries with it the assumption that a competent author should be able to explain exactly what they did within the word count. Thus, increasing the word limit of an article may not be well accepted because it could carry with it some negative connotations about the capability of the author. Third, removing a word count restriction may not be useful because it will always be difficult to use words adequately to portray a full description of a QI intervention, regardless of how many words are available.

Thus, reporting problems are not related to the constraints of word count, but rather to the *“constraints of being a human being”* (Author A) – an author cannot report everything. Participants and Lean authors both referred to the example of context to illustrate this point because context is hard to portray fully using only words.

Twelve participants wanted more detail about five contextual features: the setting (including type of hospital and type of admissions); buy-in (how did they get frontline staff to participate); leadership (*“who drives the process”* (Surgeon-Author 14)); *“senior support”* (Academic-Influencer 21); and barriers or *“push back”* (Academic-Author 09). All three Lean authors reported that many contextual features influenced the conduct of their study, but that a full description of context was not included because they preferred to write a concise account (see also section 8.3.2); they did not have *“an adequate language to describe it”* (Author B); and that authors who are not well versed in qualitative or ethnographic research may feel ill equipped to report it (Author A). Also, while adding details about context could make the article more useful by allowing the audience to consider how they would translate the intervention into their own setting, it could simultaneously make the article *“woolly”* (Author C) and less readable. While discussing whether the Lean article provided sufficient detail about context, a small number of participants changed their mind about word count restrictions, and this suggests that the question of how to improve reporting may not be straightforward to answer.

“Maybe I was wrong at first to say we don’t need more words, because yes it is complicated and I would be perfectly happy to admit that these inadequacies of reporting were down to the fact that we were squeezing into a word count.” (Author B)

Adding ‘supplementary’ material to the main article was regarded to be beneficial by the participants and authors alike, but the participants could not agree on whether other methods of providing additional information should be encouraged. For example, some participants suggested that the use of ‘bite-size’ summaries (Table 13, row 6) and second articles (Table 13, row 8) could be used to elaborate on some especially interesting aspect of a previously

published paper. A bite-size summary could be especially helpful when time is short, and a second publication when more detail is needed for replication. However, Authors A and B said that promoting the use of 'bite-size' summaries or second articles would be self-indulgent and greedy: *"That's like getting two bites of the cherry"* (Author B).

When using words as a sole medium of relaying a message to an audience, some discrepancies between the author and their audience may always arise because words may not be an adequate medium through which QI can be fully described. For example, the wording of a publication can be chosen by a QI author who is unfamiliar with, or trained in, the rigours of academic writing (Chapter 6, section 6.2.1). Some participants suggested the Lean article was written by authors who had QI expertise and therefore, the audience would need a lot of knowledge about Lean to be able to use it (*Anaesthetist-Consumer 45*). Lean authors admitted that their version of Lean was *"a bit of a quasi Lean"* (Author C) which could have been problematic, but they felt that their article, like most QI research, tends to be written to be used by front-line staff who do not have specific academic training.

"People who describe their work as quality improvement tend to take the approach that amateurs – people who have mainly clinical experience and no academic professional training as an improver – can do a perfectly adequate job. QI papers need some expertise, but the emphasis is much more strongly on the front-line staff designing things." (Author B).

8.5 Discussion: Authors and their audience

The author-checking exercise highlights some of the reporting difficulties which have been presented in previous chapters. First, QI is a field which suffers from inconsistent terminology and terminology which is not always well understood (section 8.2). Second, QI is often conducted by healthcare staff who face many competing demands, such as satisfying the goal of 'keeping the hospital out of the red' (section 8.2) while selecting project goals which suit their own clinical priorities (see also Chapter 6, section 6.2.1). It may not be easy to balance the conflicting requirements of a QI improver working on the front line, who needs a practical usable resource, with those of a journal editor who needs a concise, scientific account. In this chapter, the main disagreements between authors and participants were about whether QI should be replicated and how (section 8.3.1), and the purpose and type of QI publication (section 8.3.2). Analysis of this data has shaped a developing argument about what type of science QI is starting to become and how it should fit in with the existing landscape of scientific fields which are important in perioperative care.

The author checking exercise raised some fundamental differences in opinion about what the purpose of a QI article might be. QI articles may be published in order to generate ideas for use in practice, or to allow others to create an (almost) exact replica, or a modified replica of an intervention. Each of these different aims may require a different style of writing, and could be published in a 'QI journal' or a 'traditional' journal. However, as demonstrated by this author checking exercise, an article can be read by different stakeholders, who may wish to use the article in a different way from that originally intended, and this can cause a mismatch of views between the author and the reader. If an article is concise, it might be less useful for those wishing to apply it in practice, but helpful for editors and authors who value brevity in the traditional model of reporting. If an article is completely novel, it may not be as relevant for clinicians who see QI as being routine good practice. If QI authors publish in the journals they hold in highest esteem and select some research designs over others, some research questions might remain unanswered. If an article is light on theory, readers may not fully understand how and why the intervention is thought to work. If these opposing views cannot be resolved tensions between

editors and authors about what is deemed to be acceptable to publish may result in some QI work being more visible than others.

The data presented in this chapter suggests that a model for reporting which can satisfy the needs of all QI stakeholders does not yet exist. There is tension between what participants perceive to be the *proper* content of scientific accounts (in terms of style and content) and what might actually be useful and relevant for those who want to use QI work to improve care. An account which is written to meet the needs and expectations of a front-line QI audience in terms of what is included and how it is presented, may not be acceptable to a 'traditional' journal such as the *BMJ*. However, many QI authors who have a heavy investment in the currency of academic reporting may not be prepared to sacrifice academic esteem for the needs of a QI audience. Authors may not wish to move away from the reporting outlets that they hold in highest regard because it may require them to invest a lot of time and effort into changing their writing style (which they have perfected over many years) to gain a reward that seems dubious.

The choices that the authors and their audience made about which interventions could helpfully resolve reporting problems offers interesting insights into the type of field they envisage QI might become. Participants and Lean authors saw value in the use of supplementary material and a QI diary. The use of supplementary material is already endorsed by many traditional journals, and lab books (which participants likened to the QI diary) are a feature of the hard laboratory sciences. Conversely, the idea of longer word counts and additional 'second' publications were rejected, just as they are often rejected by conventional medical reporting models which favours brevity. Thus, the authors expressed a desire to retain what is valued about a traditional medical reporting model. To be useful, the interventions selected must be meaningful for the QI authors, journal editors, and those who are consuming the work in practice.

QI is an applied science but the science can be used for different purposes (exact replication, modified replication, or the generation of a new idea), and this adds to a picture which is being built up in this thesis about what is unique about QI. The interventions designed to improve reporting must improve each

type of QI publication, while also being valued by front-line improvers and hard-core health services researchers who both have an equal stake in QI. The challenge now will be for QI stakeholders to embrace what is unique about QI, while also retaining what they value in traditional research.

9 DISCUSSION

The discussion will open with my personal reflection on the process of completing this PhD. A discussion of the main learning points will be presented, with a view to integrating and distilling my findings about why reporting is hard, and whether the field of perioperative care is likely to involve problems which are markedly different from those faced in other healthcare specialties. The limitations of the qualitative study, and how the findings of this PhD may inform recommendations for practice for all those who work in the field of perioperative QI, will also be presented. Finally, this chapter will outline how this PhD might be taken forward.

9.1 Personal reflection

At the very beginning of undertaking this programme of research I completed a statement of personal motives (Chapter 3, section 3.2), and I reviewed the literature (Chapter 2, section 2.1) to develop a research question about why reporting QI is so hard. In order to understand what QI involves, I also conducted a small QI project involving PDSA cycles, and discovered for myself some of the challenges of intervention specification, modification and reporting. This highlighted why it is important for QI authors to explain their perceptions, because although they do not always know for certain which intervention elements were important, attempting to explain to their readers what worked best for them can be helpful.^{39,381} This led me to believe that QI can be pursued in a manner which gives value to both logic and reason (an outsider view) *and* individuals' perceptions (an insider view), which are two views proposed by Blaikie about how knowledge can be acquired.⁹¹ Making this explicit allowed me to begin to understand that QI can entail an important mix of qualitative and quantitative approaches to data collection and synthesis.

It is my duty as a healthcare professional and a researcher to reflect upon my own performance, and reflecting on how I have conducted myself during this PhD will help me to focus on how I should progress towards a future leadership role. This process of reflection has generated four categories of personal attributes: curiosity, reliability, humour and patience. As an applied researcher I do not situate myself as an external actor in society, but rather, I recognise myself as a member of it. I am curious to understand an individual's perception

of their reality and this has translated into a high regard for well-executed qualitative research. I sought to be a reliable researcher during the process of obtaining ethics. I consistently provided prompt responses to queries; I kept all correspondence correctly filed so that I could later explain which decisions were made, in which order and for which reasons. Humour and warmth enabled me to build a rapport with my interview participants, as well as with people across multiple organisations such as hospitals, universities and journals. Finally, patience helped me to complete the work during sometimes challenging circumstances. Through understanding patience, I have also understood the value of caution and I avoided rushing to action at the outset of the qualitative work, (see section 3.4). This decision was instrumental in enabling me to shine a light on what really makes reporting difficult in perioperative QI, and what I should focus on to improve it during the next phase of my academic career.

9.2 Youth, aspiration and maturity

To fully explain why reporting QI is so hard, this discussion chapter is divided into three sections: the youthful field of QI, the aspiring field of QI, and the maturing field of QI.

My overall view of the data presented in this thesis is that reporting QI is a 'messy' business because we cannot yet answer the question: 'What is this thing called QI?' QI is a young field. It involves concepts which are poorly defined, and it exists on a continuum from health services research (which can be easily recognisable as science) through to local QI projects. Beneath this overview of the youthful, messy world of QI, QI stakeholders operationalise it, making decisions day to day about how QI should be reported. QI is reported and published by many different organisations and individuals who are tasked with the job of deciding what kind of science QI should aspire to become. QI stakeholders have different (and sometimes competing) priorities about what they value about scientific reporting and what QI should be used for. It may not be possible to remain completely faithful to what is valued most about scientific reporting, while allowing QI to emerge as a separate scientific field, and some trade-offs need to be made. From the aspirations stakeholders have for what QI reporting should look like, a set of new norms and values are re-shaping the field of QI. Now, reporting items can be identified which represent what is

unique about QI, and what makes this field distinct and different. This delicate, somewhat embryonic nomenclature represents a move away from the messiness and ambiguity which has characterised QI as a youthful field. Now, QI is being cast in a more mature mould which is acceptable to a wide variety of stakeholders who wish to use QI work to benefit patients who experience perioperative care.

9.2.1 The youthful field of QI

The problem that QI is an emerging field which is poorly understood as a science (if it should be regarded as a science at all) has pervaded the findings chapters of this thesis. Participants found it hard to explain what QI is, because they struggled to answer questions such as:

- Which interventions are QI interventions?
- How does QI fit in with the landscape of ‘traditional’ science that many QI stakeholders are used to?
- What is context and how should its influence on the study outcome be reported?
- How can a QI researcher know which ingredients of an intervention have the most effect on the study outcome?
- How can a QI researcher know which ingredients need to stay the same and which can be changed?

Inadequate reporting of QI in the perioperative literature was confirmed by conducting a systematic review (Chapter 4, section 4.3). The 100 included perioperative QI papers scored an average (arithmetic mean) score of 6.31 (of a total of 11 TIDieR checklist items which guide intervention description reporting – Table 5). This indicates that on average, the papers were missing 43% of the elements they should have been reporting. This could be a symptom of the problem that QI is a youthful field, whose followers are yet to gain a full consensus on what QI is and how it should be described (Chapter 5, section 5.5). As a team of systematic reviewers (MDW, GPM, NJL and I), we too fell afoul of the lack of agreement in the literature about what constitutes QI, and disagreements arose between us about which literature should count as QI (section 3.3.1). QI is a term that covers a heterogeneous set of interventions

and methods. Some of the terms used in QI have definitions which are imprecise because they can be understood in different ways by different people. Two key issues underlie the symptom of fuzzy terminology in QI reporting, and characterise QI as a youthful field. First, QI has poorly defined boundaries with scientific fields, and between QI research and service delivery. Second, agreement is poor about what makes QI research, research, and what constitutes proper scientific reporting.

A boundary which separates one set of ideas from another can be important in establishing the identity of a particular approach (such as clinical trials or QI). The qualitative work undertaken during this PhD highlights a lack of clarity in how QI should be distinguished from (and overlap with) a number of other scientific fields including implementation science (IS), human factors, the social sciences and traditional clinical sciences (section 5.5.2). For example, the term 'implementation' is used in the field of IS and QI, but may mean different things. The term implementation could signify a type of science called IS, or it could indicate that something is applied in practice. IS can require extensive searching of the literature to identify an appropriate model of implementation³⁹¹ or suite of 'implementation interventions',³⁹² but during QI projects an intervention is often implemented relatively spontaneously, because it is thought to work well in that setting. The distinction between IS and QI has evolved over time; literature from 2008⁸² and 2009¹⁶⁹ was less clear about how IS should be distinct from and augment QI efforts than it is now. The emergence of the journal *Implementation Science* is helpful because when IS is reported here, the way in which it supports the QI work tends to be carefully illustrated.

A lack of agreement about how QI should be distinguished from other scientific fields creates easy-to-identify issues such as inconsistent terminology as well as imprecision regarding where QI work is published (Chapter 6, section 6.3.2). 'Outsiders' to IS and QI who might not be academically trained (section 6.1.1) could struggle to understand the terminology and, as a result, submit their work to a wide range of journals in an effort to achieve publication. Conversely, those on the inside can feel like 'pioneers' (p.1539)⁸² staking out a territory, taking care to distinguish what is and is not worthy of publication in a specific journal. Systematic review (Chapter 4, section 4.3) confirmed that perioperative QI work

does tend to be published in journals across a scatter of category types which can be specific to a surgical specialty, such as 'urology and nephrology', or across healthcare more generally, such as 'healthcare policy.' One approach to resolving this could involve pursuing a trade-off between positioning the perioperative QI literature only in a core set of 'QI journals', and encouraging acceptance of QI across a wider set of publication types, including clinical journals (Chapter 6, sections 6.3.3 and 6.4).

A specific set of QI journals could ease the problem of identifying relevant literature, and could also promote the integrity (as well as the 'brand') of QI as a specialised field of science. This approach is responsive to general trends in publication whereby people are more likely to find an article by searching through a multi-journal, online database than by leafing through the pages of an individual journal.³⁹³ Thus, for QI work to be accessible to everyone, perhaps it matters less which journal it is published in, and more what key words are used to ensure it is accessible. Making improvements to search terms and indexing,¹²⁰ and/or building a specific repository (or database) for QI³⁹⁴ could generate a better portal to QI work. The individual identities of journals are important, but optimising the selection and retrieval of relevant literature which is published in a range of journals, could then translate into the production of publications that are understood more consistently between QI stakeholders. In other words, the creation of such a portal could itself contribute to the cohesiveness of this youthful field, and the coherence of terminology that has previously caused confusion or division. However, promoting recognisable QI journals (and perhaps a QI database or repository) does not come without challenges. Perhaps most importantly, my study does not provide enough data to suggest that this is the best approach.

Some participants expressed a preference for publishing in places visible to peers in their clinical field rather than a broader, cross-disciplinary QI audience (section 6.1.1 and 6.4). Also, QI authors who publish for career progression might be afraid that QI journals may not be held with the same esteem as alternative clinical journals (Chapter 6, section 6.1.1, and Chapter 8 section 8.3.1). A more tempered approach could involve encouraging such a repository or portal to target both 'QI journals' and clinical journals, so that a range of

literature could be accessible to readers. However, opening a debate about the value of 'QI journals' or even a composite repository which combines specialist clinical journals into a single portal, could be met with mixed feelings. Some stakeholders may feel this is a legitimate and useful debate, but others may suggest time should be invested in making other choices to optimise perioperative QI reporting. Manipulating where perioperative QI literature should be placed is not an easy task and cannot simply be engineered: a journal's reputation takes many years to evolve and is the product of multiple factors, including the impact factor (Chapter 6, section 6.1.1) and the esteem with which it is held by its audience.³⁹³

Through examining another poorly understood term in QI – 'intervention' – another important characteristic of QI can be identified: it is conducted and reported by multiple stakeholder groups. The question of 'did the intervention make the improvement, or was it was something else?' ^{30,392} is an important one, but to answer this question, a QI author first needs know what the intervention comprised. For clinical staff, a QI intervention (or quality intervention – see section 3.3.1)²⁸⁸ can include either clinical or non-clinical activities, because any effort to improve quality counts towards the overall effect of a whole QI programme (Chapter 5, section 5.5.1). Conversely, most editors reported that only non-clinical interventions can be QI interventions. Participants also found it difficult to ascertain whether QI tools (such as fishbone) were part of the intervention (Chapter 8, section 8.2). The literature is not clear about this either, and fishbone has been described as an improvement tool,^{395,55} a QI method³²⁷ or an intervention.³ During attempts to unravel what a QI intervention involves, both trained researchers (including ourselves – see section 3.3.1) and clinical staff (who may or may not have received formal training in research or QI) navigate a sea of literature including taxonomies. However, clinical staff often reported that their learning was self-directed and required a heuristic approach because it needed to be balanced with the sometimes conflicting demand of their clinical workload (Chapter 6, section 6.2.1).

Examining this range of beliefs about what interventions are, and how demands are juggled to accommodate reporting, enables the observer to notice that QI can be reported in two different ways. Some QI stakeholders may wish to

publish research, and others are interested to improve the delivery of a local service and publish a QI report. Research and service evaluation or improvement projects require different reporting styles because the aims of the work, as well as the risks and level of intrusion on participants, are very different.³⁹⁶ However, both types of report require a detailed description of the intervention to promote the application of QI work in practice. My systematic review demonstrated that the nine articles from the journal *BMJ Quality Improvement Reports* (which publishes QI projects – section 5.6), achieved a higher mean TIDieR score of 8.2, compared to the research papers, which achieved 6.31 (Chapter 4, section 4.3.1). It is possible that these scores occurred by chance, but the observed difference does offer a point for discussion.

Several features could contribute towards the more optimal level of intervention reporting achieved in *BMJ Improvement Reports*. This journal was established in 2012, which could have enabled authors to gain more learning from recently available publications, such as the SQUIRE guidelines.¹⁴⁹ Also, this journal provides a list of resources including the SQUIRE guidelines¹⁴⁹ and a structured online template to aid intervention description. Finally, individuals publishing in this journal are expected to comply with one set of expectations: they are writing a QI project, rather than writing QI research. To provide complete accounts of their QI work, authors need to have a set of strong, clear expectations,¹⁵⁷ and this debate raises the question of whether some of the expectations set out for QI project authors, could also be useful for authors of traditional research publications. The authors of QI research may need to balance the expectations of an audience who are orientated towards both QI *and* research, and a lack of clarity about what constitutes proper scientific writing in QI could result in sub-optimal reporting.

Research and healthcare service delivery are two systems whose actors interact with each other³⁷⁴ and my qualitative study of QI reporting provides insight into how a continuum between publications of different types actually operates in practice. QI stakeholder groups include: healthcare staff who are not trained scientists and are interested in local service improvement; academics who are driven by the currency of publishing; and those who have

both identities. Thus, some stakeholders may aspire to fulfil a dual aim: publishing a piece of QI work which adds to a pool of published research literature, *and* fits the mould of a QI project because the intervention was conducted at a single site (Chapter 5, section 5.6.3). The question of how lessons can be extrapolated from single site studies to be applied more broadly was discussed at length (Chapter 7, section 7.6.1), and stakeholders may have many reasons for wanting to publish this type of work in more traditional research orientated literature (Chapter 6, section 6.1.1). However, this creates the thorny issue that QI may not in fact operate on a linear track from health services research, through to local QI projects and service delivery. Instead, this track may have a series of junction boxes which can allow some melding to occur between what can be recognised as the values of hard sciences (such as generalisability) and the values of QI (such as explaining what really happened) (Chapter 5 section 5.6.2 and 5.7.6, Chapter 6, section 6.3.1).

Several existing theoretically informed publications helpfully indicate how QI operates in this way, spanning the institutional fields of academia and clinical practice. Martin et al³⁹⁷ explain that an institutionalised field incorporates a set of institutional expectations and demands which are made of the people who inhabit that field. A direct examination of the everyday lives of the QI stakeholders who inhabit the fields of academia and service delivery shows, that their own personal preferences, beliefs and values can generate various tensions and conflicts (Chapter 6, sections 6.1 and 6.2, Chapter 7, section 7.5). Resolving these tensions could involve a ‘melding’ of scientific values, but this does not demonstrate a move to transform the norms of conventional publishing. Instead, QI stakeholders are attempting to ‘forge a legitimate path’ (p.4)³⁹⁷ between fields that have meaning and value in QI – research and service delivery. This could create a route through which QI might be better understood, and respond to a recent call to increase the range of underpinning disciplines that evidence- based medicine can draw upon.¹¹¹ However, these stakeholders face many challenges. For example, the reputation of an author or a journal, and confidence in a particular style of writing, is developed over many years and has become deeply engrained in institutional fields. Thus, in keeping with Martin et al’s³⁹⁷ theoretically informed understanding of institutional fields,

any new attempt to forge a new path will require a very co-ordinated effort (and perhaps even a degree of entrepreneurship) to create any lasting change.

To fully explain how a continuum between QI, and research can operate, it is necessary to understand what makes QI research, research. Many authors have pondered similar questions. Greenhalgh³⁹⁸ for example, recently asked – ‘What is knowledge?’ In this view, which is informed by Van de Ven and Johnson’s ‘knowledge for theory and practice’,³⁹⁹ what one person can know is directly linked to what collective groups know (such as a profession), and knowledge formed through a convergence of academics and practitioners can be very helpful. Participants in this study understood this, suggesting multidisciplinary writing teams would be a great enhancement to QI publications. Data from this thesis cannot suggest a precise method by which QI reporting can embrace MDT working, but highlighting a need for it is a good place to start because it can begin to illuminate exactly what type of science QI might become. While participants knew what would count as useful *knowledge*, they could not clarify what counts as *research*.

Miller⁴⁰⁰ points out that an important function of research should be to identify causality. Unfortunately, as Miller⁴⁰⁰ explains, according to Poncaré,⁴⁰¹ the ability to make causal predictions (with certainty) requires a low threshold of complexity (because it is difficult to trace causality when there are multiple interactions between a complex range of variables), and in healthcare this is often almost impossible to achieve. This could frustrate many authors, not just in the field of QI, for many systems (healthcare and biological) can be too complex to maintain control over many interacting variables. Conversely, in clinical practice a degree of creativity and uncertainty has been celebrated because it can create freedom for thinking to evolve.⁴⁰² If such low levels of complexity were sought, research in healthcare might never be conducted at all and so inevitably, compromises need to be made.¹¹² QI is embedded in clinical practice, and QI researchers find value in transplanting traditional research practices – such as the concept of replication,¹¹² but also research practices from other fields, such as the ability to understand how complex systems work.⁴⁰⁰ In the young field of QI, researchers are grappling with the issue of how much and what type of compromises should be made. In the field of QI, values

associated with both research and practice are important, and it may be necessary to adjudicate between uncertainty which is inherent (in practice) and uncertainty which should be reduced (through research), especially when findings are translated into practice for the public good.²¹²

The concept of novelty has been linked to understanding what makes research, research. Some literature published relatively early in the evolution of QI in healthcare clearly states that QI does not set out to generate new scientific knowledge: instead, QI evaluates and improves existing ways of working.⁴⁰³ Margolis et al⁸⁴ maintain that QI usually combines existing practical knowledge (from clinical experience) and academic knowledge (from interventions which have already been shown to work) to improve clinical practice. However, in the same document they also suggest that QI can be used to generate and adapt new knowledge to create improvements. The participants in my study reflected this absence of a clear consensus about whether QI can generate new knowledge or not (Chapter 5, section 5.5.1), and it is clear that 'novelty' is not a universally accepted means of differentiating between QI research and 'non-research'.

The question of what makes research, research, thus becomes connected to another important question for QI reporting: what is valued in QI research? While novelty is clearly held in high esteem, the value of the more mundane task of steadily chipping away at refinement, adjustment and careful replication of something someone else has done has been raised in the literature. Replication can be regarded an important scientific principle.⁴⁰⁴ Participants valued the principle of replication and the ability to extrapolate what was learned and apply it elsewhere. In addition, the original QI publication itself should not be a search for glory, headlines or a truly novel intervention (Chapter 6, section 6.2). Instead, it should be about making care better on the ground. Participants drew upon their existing knowledge of healthcare research to explain that in QI research they valued concepts such as validity, generalisability, historical controls, contemporaneous controls, cause and effect, narrative, and qualitative data. Validity and generalisability can be contentious terms in the research community, especially in relation to qualitative research,²⁸⁹ and my participants noticed that the application of these terms might also be contentious in QI. For

example, a lengthy discussion about generalisability (Chapter 7, section 7.6.1) generated alternative approaches to applying the term in QI. Thus, participants were generally clear that they felt that ‘yes’, QI is scientific, but they struggled to explain exactly: ‘What type of science is QI?’

In their work on IS, Sobo et al⁸² (p.1530) use a quote from Geertz (1973) to explain to their reader:

“If you want to understand what science is, you should look not at its theories or findings; you should look at what the practitioners of it do.”

A helpful insight from the quotation from Geertz (1973) is the word ‘practitioner,’ which directs attention towards the many different stakeholders who may be involved in the execution of science. A large number of people from different professional groups are involved in designing, implementing and reporting the QI interventions (Chapter 6, section 6.1). Clinical staff interviewed in this study were especially keen that those conducting and reporting QI should have access to expertise, which could include QI experts, methodologists, epidemiologists, clinicians, research nurses and patients. This support could improve the relevance, utility, and scientific quality of the QI work, so that it can be used most effectively to benefit patients (section 6.1.2). These discussions, which were about supporting clinicians to report QI research as part of their everyday work, align with the idea of a ‘health research system’, which is: “a systemic, co-ordinated approach to the generation and use of research in healthcare organisations” (p.1).³⁷⁴ Therefore, the next section will discuss how organisations, and the individuals working in them, attempt to remain faithful to what they value most about scientific reporting, while allowing the QI to emerge as a separate scientific field.

9.2.2 The aspiring field of QI

The context of QI as a youthful scientific field which lacks agreement about how QI should be understood is clearly problematic for reporting, but this does not mean that the task of conducting and reporting QI work has been halted or stunted. A pressing international agenda to improve quality of care (Chapter 1, section 1.2) has meant that QI is actively encouraged across global healthcare

systems to facilitate an effort to continually improve the care that perioperative patients receive.^{204,261} Thus, a group of individual people and organisations, who operationalise QI reporting and publishing on a day-to-day basis, are making decisions about what a QI article should look like and what QI should aspire to become. Together, the decisions made by this group come to shape the field, through the reproduction of the norms and rules that govern the field on a day-to-day basis, and the deliberate, strategically oriented choices intended to produce new norms, rules and practices.

An individual QI author or team of authors make decisions about how a QI study should be presented, what should be included, and where it should be submitted for publication. However, this decision making is influenced by the many priorities of different QI stakeholders, which can sometimes conflict with each other. For example, QI authors may not know how to write in a style which is appropriate for a QI project or QI research (Chapter 5, section 5.6), and QI editors (and peer reviewers) may apply different parameters to decide whether QI work could be published as original research or not (Chapter 6, section 6.3). The scatter of articles submitted for publication can cause tension between authors, who do not understand where QI fits among different types of journals, and editors who feel irritated because it seems that articles are submitted with little thought. Conversely, authors may feel that they know what is expected of them, but it is the editors and peer reviewers who are not aware of how QI should be reported. These disagreements can create many problems, including the issue that some QI authors who have little patience for the process of publication, may never have their work published at all. At the root of these disagreements lies an incongruence between QI and the conventions of medical publishing, but a manuscript format that is faithful to the values of both fields might be hard to find.

Conventional scientific publishing often promotes the use of the introduction, methods, results and discussion (IMRaD) structure, which began to be used in the 1940s.³⁷² IMRaD is retained by the SQUIRE guidelines as a preferred reporting structure and it is familiar to most researchers (Chapter 6, section 6.3.1), yet some QI stakeholders struggle to use it. Conventional formats of scientific reporting may not enable QI authors to satisfy what they personally

value about QI reporting while simultaneously satisfying what they perceive editors to want (section 6.3.2). For example, some participants valued the use of qualitative data and a narrative reporting style to convey information about the contextual features of their QI interventions. Others disagreed, preferring to objectify contextual data, because this is a style of writing they are already familiar with, or because they deem research which does not require an in-depth description of context at one site to be more generalisable and publishable (section 7.6.1 and 7.5). To manage this, section headings such as 'what really happened' could be added to embrace what is valued about QI while retaining the IMRaD format, yet this can conflict with a requirement for brevity which is important in conventional medical reporting.¹¹¹

Manuscript format is perhaps a mundane problem, but nevertheless it is important: as Greenhalgh explains, '*real evidence based medicine*' (p.4)¹¹¹ presents information in a format that the audience can understand. In QI the 'form' of what reporting characteristics are valued in the field, should match the 'function' of the publication – to relay knowledge to an audience. Participants suggested that a simultaneous need for brevity and more detail could be accommodated, but they went to great lengths to produce solutions which allowed reporting to retain what is valued about a traditional medical reporting model. In the spirit of maintaining a close tie with existing styles of reporting, they suggested that supplementary material or videos could be added, or new methods of publishing could be established such as 'layering' an article so that interested readers could delve deeper and deeper, obtaining more information as needed. The internet means some of these changes are already occurring and can benefit all study design types, not just QI. For example, guidance for social scientists,³⁹³ systematic reviewers¹¹⁷ and healthcare researchers^{113,150} encourages authors to reference web links or online supplements.

The emergence of a series of practical suggestions about how a QI manuscript could accommodate detail indicates that stakeholders aspire to the idea that contextual features could be completely reported, yet this ideal could be very difficult to enact in practice. Knowing what to report is not easy when 25 separate characteristics of context (under the headings: external environment, organisation, QI support and capacity, microsystem, and QI team) have been

described.³⁹ To complicate this, other authors use different terms to describe context, such as resources, readiness for an intervention, and the socio-political climate.²³⁹ These contextual factors can be hard to describe because they change in an unpredictable manner over time, they evade quantification (or even clear description), and they can vary significantly between clinical settings.⁹⁰ Thus, it is easy to understand why participants described context as 'hard to describe' and 'nebulous'. Guidance such as the MUSIQ tool³⁹ might help, but despite encouragement for authors to read between the lines,⁴⁰⁵ items which do not quite fall under the categories presented in the tool could still be neglected. Further, tools like the MUSIQ model and the SQUIRE guidelines were published relatively recently (in 2012 and 2008 respectively), and it can take many years for research to be applied in practice.⁴⁰⁶

Enacting the task of reporting context may be facilitated by encouraging the use of qualitative methods to collect data about context, and a narrative writing style to report it. Qualitative approaches to gathering data have become more prominent in surgical research^{101,407,408} and their use in QI could also be valuable, for example, with the use of a QI diary. QI research has included a recent focus on learning from positively deviant environments, instead of focusing exclusively on environments where quality is poor,⁴⁰⁹ and this could encourage researchers to notice the contextual features that influenced an intervention. External evaluators could add value to data collection (Chapter 5, section 5.7.6) to ease the problem of noticing contextual features in the first place. Noticing contextual features relies on feedback mechanisms and a sense of collaboration between the QI team,⁴¹⁰ but even when such a team is in place, a QI researcher might be 'too close' to their own work to be alert to what is really happening (Chapter 7, section 7.3). There is not a simple solution to enable authors to tease out which of the vast array of contextual features should be reported, and to understand how this should be done. The methods used to recognise and collect contextual data are likely to be different for each QI team, so I do not offer a prescription. Instead, my data provides a structured overview of what might be needed to improve reporting, to allow a process of further discussion and consensus building to begin.

To reconcile the incongruence between QI and the conventions of medical publishing, and establish what should be accepted as scientific writing in QI, tensions need to be resolved. Tensions might be resolved with the provision of training for authors and editors alike (Chapter 6, section 6.2), but others suggested that it should be the editors who should take the lead and encourage authors to submit articles which combine quantitative and qualitative approaches. Tension may also be eased by increasing the available word count to accommodate detailed narrative accounts, but this may not be sufficient to fully portray the twists and turns of what happened during a QI study (Chapter 8, section 8.4). Reaching a compromise on reporting style will be hard because beliefs about what type of reporting is held in the highest esteem are entrenched. QI authors may be unwilling to deviate too far from existing academic reporting styles because investing a lot of effort to learn a new reporting format might offer dubious academic reward, and could even be detrimental if authors cannot publish in a format (and in a journal) which is held in the highest esteem by their peers (Chapter 6, section 6.1.1). Discussions (which are already ongoing)⁴¹¹ between journal editors and authors about what an appropriate balance between qualitative and quantitative writing could look like, will take time to resolve and, at present, an optimal format for reporting which might satisfy all QI stakeholders remains elusive.

A considered response to the challenge of presenting QI in a scientific manner is important, but a response must be made to the problem that QI authors struggle to juggle competing demands. My study identified several practical solutions to manage this, such as the availability of QI training and multi-disciplinary writing teams (Chapter 6, section 6.2.1). However, some demands might be easier to resolve than others. For example, some QI authors struggle to fulfil what they themselves would wish to investigate and what hospital boards or journal editors perceived to be most important (Chapter 6, section 6.2.1, 6.3.2, and Chapter 8, section 8.2). This could affect what eventually becomes reported, as well as the pool of available literature. In my systematic review (Chapter 4, section 4.3) the majority of articles targeted clinical questions about reducing surgical complications. Much less literature examined the discharge process, patient information and handover to primary care on

discharge, and coordination within and between specialties in emergency care, even though all of these are known to be problematic.^{70,72,412} Patient and public involvement (PPI) can be used as a strategy to ensure that research can span a large number of topics which are relevant to healthcare staff and patients,²⁷¹ but my systematic review only identified two papers which reported on PPI,^{231,333} despite encouragement for QI work to include patients.⁴¹³ This is primarily a discussion about the actual conduct of QI work, rather than how it is reported. However, systematic review²⁸⁸ (Chapter 4) reveals that many opportunities remain to be addressed in the perioperative QI literature, and as Øvretveit et al⁴¹⁴ points out, if only a few specific contexts of QI work are reported, learning will be stunted: it will be hard to become more accomplished in reporting if only a small repertoire of topics is available.

Another opportunity that participants felt is yet to be fully embraced in the field of QI is the publication of negative data. Judging the extent of selective reporting is not easy because protocols are rarely published in QI, making it difficult to ascertain whether any outcomes in final reports are actually missing.¹²⁴ While it is difficult to examine the true extent of the problem, participants in this study reported that it would be good practice to address it. They suggested that reporting of null or negative findings may be encouraged by hospitals which celebrate success and failure (section 6.2.1), or by journal editors who provide a section for negative studies in their journals (Chapter 6, section 6.3.1). However, while learning from both positive and negative studies was proposed as a laudable aspiration about what could be good about the future of QI reporting, the risks (that professional reputation might be threatened, for example) may not always outweigh the benefits (that consumers can learn from studies that did not work) (section 6.1.1). This should not be ignored and efforts to create a new institutional norm in QI should not diverge from what individual actors in the field need to do to advance their own interests. Currie et al⁴¹⁵ shed some light on how a balance capable of satisfying all stakeholders might be achieved. They show that institutional norms are shaped by elite professionals who, also acting in partnership with more peripheral actors, work as ‘arbiters’ of risk. They ‘buttress’⁴¹⁵ (p.957) the existing institutional norm with rules and monitoring so that they can maintain and

enhance what they value while reducing any threat to their professional status. Thus, creating a new norm of reporting and learning from negative QI research will take time, and might be easier said than done.

A failure to report negative contextual features was also identified to be problematic, and something which could be an aspirational reporting feature in QI. Almost a third of my participants said that they would feel uncomfortable writing about how individuals or institutions had negatively influenced their study. Literature encourages the documentation of contextual features such as the complexity of the surgical team⁴¹⁶ and negative contextual features such as a loss of professional autonomy¹⁷² because this can impact the design, conduct and analysis of a surgical trial. Encouragement to publish unintended negative contextual features in QI is not explicitly provided,^{23,186} but it is beginning to surface as an important issue. Brewster et al⁴¹⁰ present an account of tensions which arose between the designers and implementers of a intervention during one QI study. This is helpful because it indicates that negative contextual features not only exist, but also that they can be recognised and tackled using a co-ordinated effort across the whole team. This might afford an opening for openly discussing negative contextual features in the academic literature.

There is no 'one-size-fits-all' solution. Each QI writing team is likely to encounter different reporting problems which will require a different combination of the solutions presented here, including: the IMRaD structure, a modified IMRaD structure; supplementary material; support for reporting qualitative data; education and MDT teams. In addition, these strategies only form part of the solution to improving the completeness with which QI is reported. As gaps in understanding what distinguishes QI as a unique scientific field are identified and explored, a fuller resolution to reporting problems may be generated. The next section will discuss what has been learned from my PhD study about reporting features which are unique to QI, and how they indicate what a more established, mature field of QI might look like.

9.2.3 The maturing field of QI

A model of reporting is still emerging in QI, in part because the balance between regarding QI as a science and regarding QI as a group of keen

clinicians wanting to make a local change, is in flux. Despite the messiness of QI, individual stakeholders have begun to make decisions about what QI reporting should look like. Some participants referred to a very narrow notion that research should be fairly positivistic, quantitative and broadly generalisable when discussing what scientific QI reporting should look like. However, not everyone agreed with the view that QI should fit in with existing scientific norms.

Consequently, QI reporting can be regarded to be a frustrating task because it can feel like fitting a square peg in a round hole (Chapter 6, section 6.3.1). Yet, from these discussions participants identified different choices and rationalised what rules and norms need to be applied and which need to be dispensed with or modified. This allowed me to extract a new set of indicators (or reporting features), which require particular attention in the field of QI, and could eventually populate a framework to define how QI should be reported in a scientific manner. The indicators include: how QI publications are used; contextual features; and the identification of active ingredients. Thus, a view of what scientific reporting could look like in a more mature field of QI is presented, with a view to creating a square hole for a square peg instead.

9.2.3.1 How QI publications are used

As Sobo et al⁸² (p.1530) indicate by referencing the work of Geertz (1973) in understanding what must be done to report QI more fully, we must first look more carefully at the act of doing it. QI is generally regarded to be applied, practical and used on the front line. It has been described by my participants and in the literature as a 'halfway house between practice and research'⁸⁶ (section 5.6). Therefore an important clinical function of this research is for it to be portable and used in other settings, and a scientific framework to support QI reporting must be responsive to this need for practical application. Chapter 8 analysed data generated from a practical exercise which involved reflecting on how a specific QI article could actually be used. From this, some stakeholders suggested that an 'outline' of a piece of research is all that is necessary to stimulate useful ideas for others, but others suggested that for a piece of academic writing to have a purpose it must be replicable to some degree. Thus,

different styles of reporting are used for different types of QI article. QI articles can be:

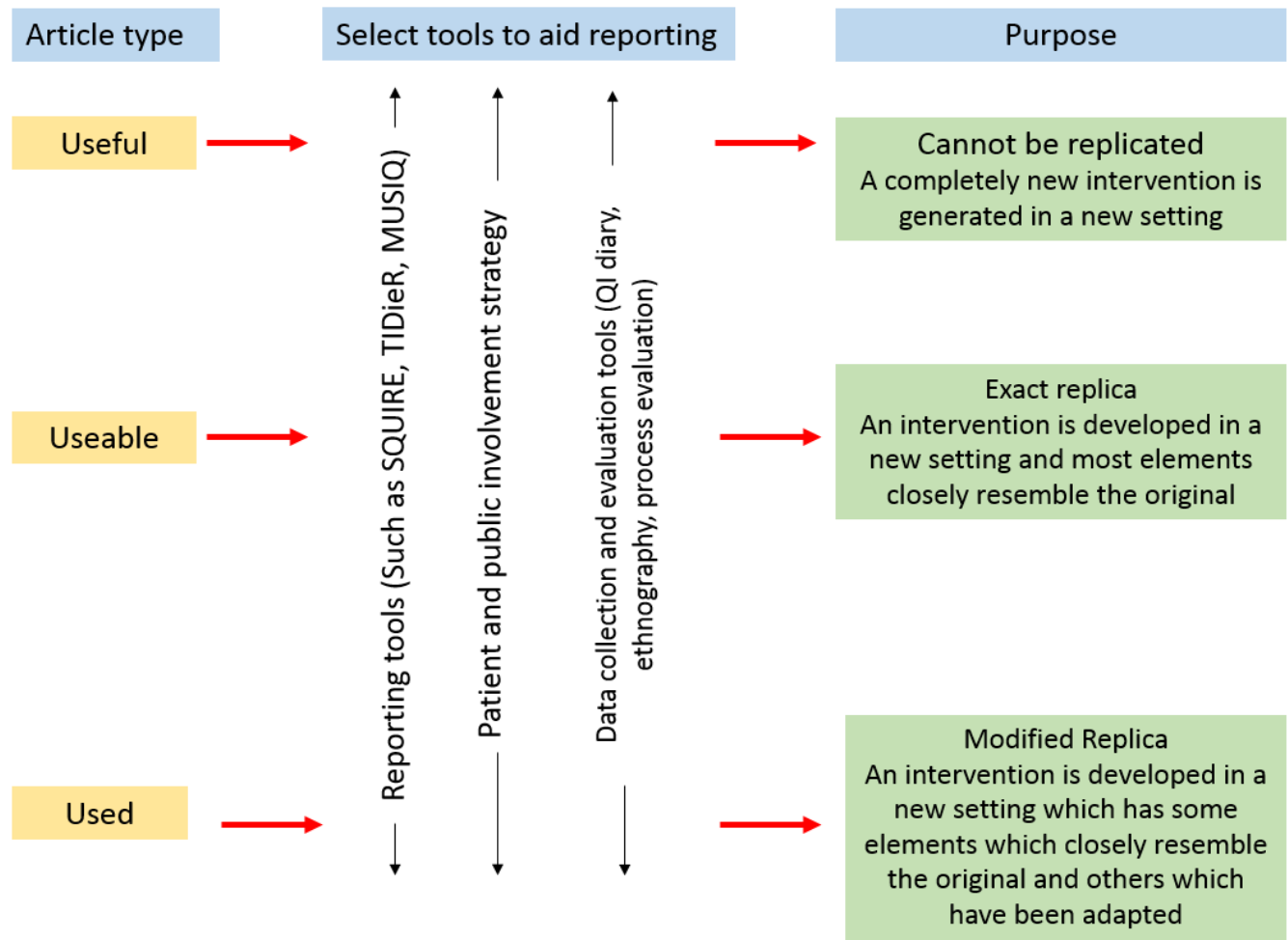
- *Useful* to others by generating ideas which will not stimulate replication of the intervention, but instead the development of an entirely new intervention to suit a new setting.
- *Usable* by others, by providing a description of an intervention which can be replicated almost exactly, with subtle changes to some elements in a new setting.
- Actually *used* by others, by providing a description of each intervention element, which will allow a modified replica to be generated in a new setting.

Figure 6 depicts the distinction between each type of QI article (useful, usable and used) and shows that for each one, the author can select a variety of tools to aid reporting, which have all been discussed by the 42 QI stakeholders in this thesis. The terms useful, usable, and used indicate subtle differences in the interaction between the consumer and the article. For an article to be useful, components do not need to be as fully described. For an article to be usable, the reader has a greater degree of reliance on accurate and complete descriptions of each component. Usable is distinguished from 'used' because 'used' indicates that the reader needs to impose a greater degree of manipulation to actually use some of the elements in their own setting.

To clarify the distinction between each category from a QI author's point of view, an example is provided for each one. For a QI author who intends their QI publication to be *useful* for the generation of completely new ideas, thick, detailed descriptions of context would become superfluous and frustrating. For a QI publication to be *usable*, an author will need to indicate clearly (perhaps through using supplementary materials) which set of materials or procedures could be kept the same. For a QI publication to be *used* it will be important for an author to make contact details explicit (to enable a reader to obtain additional information which may not be available in the written text) and to allow their own (or an external evaluator's) view to pervade the text to indicate which ingredients were felt to be most important and why.

A need to recognise the distinction between *replicating* QI work and *emulating* QI work has been raised by Mosher and Ogrinc,⁴⁰⁵ and along with others such as Øvretveit et al,²¹⁸ they emphasise that it can be unrealistic to expect QI to be used for exact replication. Instead, those using QI work should execute a degree of latitude during implementation in a new setting. Mosher and Ogrinc⁴⁰⁵ proffer that the ‘exhaustive cataloguing of minutiae’ (p.2) is not synonymous with a comprehensive useful report, but QI authors do need to rise to the challenge of getting the balance right. The suggestion that ‘emulation’ can be further understood by the three categories proposed here (useful, usable and actually used) may therefore provide some relief to QI authors because it provides a rough guide to the level (and type) of ‘minutiae’ to be included for each one.

Figure 6 Types of QI article



9.2.3.2 Contextual features

Another reporting feature of specific importance in the field of QI is context. QI interventions tend to operate in many different organisational contexts; they can be implemented by many different people (who have different values and behaviours), and they tend to mutate or change in response to new data as the study evolves. Therefore, imposing a template which is suited to reporting a biomedical intervention such as a pill or even a complex intervention like a surgical procedure, may not be suited to this unique feature of QI.

This thesis demonstrated a high level of agreement among participants that context should be reported in QI studies. For four participants, reporting context was less important because their classical training in experimental research methods taught them that context is 'noise' which should be controlled out (Chapter 7, section 7.5). This view is aligned with literature which suggests contextual features can be termed *effect modifiers*.²⁰⁰ On the other hand, the majority of participants placed a higher value on reporting context, stating that in QI, contextual features can help to explain how much any observed effect (or outcome) can be attributed to the intervention itself.

Two main contrasting views about the influence of context pervade my findings and the QI literature. The first, summarised by Taylor et al,²¹⁰ Øvretveit et al,⁴¹⁴ Øvretveit¹⁹⁹ and Kaplan et al,¹⁹¹ is that contextual features can exist 'around' a QI activity (section 7.3), and while they are not part of the intervention itself, they are capable of influencing it. The second, alluded to by Kringos et al,¹⁰² is that contextual features can be regarded as part of the intervention, such as a behaviour change intervention which is imposed deliberately to modify the values and beliefs of healthcare staff during a QI study. In this latter view, contextual features be defined as one of the active ingredients that caused an intervention's success or failure and brought about change (Chapter 5, section 5.7.6).¹⁸⁵ Exposing this ambiguity about whether the contextual feature is part of the intervention or not, is a helpful step towards extrapolating the finer detail about what is really important in scientific reporting of QI. The question of whether contextual features are part of the intervention or not may not be important to all stakeholders, but posing it has sharpened the concept that

studying and reporting the interactions between interventions and context³⁰ has particular importance in QI.

The relationship between contextual features and QI outcome has been evaluated,^{39,102} and the clinical microsystem (a group of staff working together to provide care)²⁴⁴ and the functioning of the QI team can influence the outcome of QI most directly. Mosher and Ogrinc⁴⁰⁵ suggest that these contextual features could be even more powerful than what they call the ‘hardware’ of the intervention itself – the procedures or materials required to carry it out. This analogy is helpful because just as computer hardware is useless without software, the intervention hardware in QI is useless without the contextual software to deploy it. In this sense, contextual software is a collection of ‘active ingredients’ (or perhaps actions or instructions) that make the hardware usable. However, if the purpose of a QI article is for the reader to be able to deliver some kind of replication (whether exact replication, or a modified type of replication), the hardware requires a scientific framework through which the contextual software can be described to make the hardware not just usable, but actually used in practice.

QI stakeholders agreed that contextual features can help to explain how generalisable a study might be to other settings. QI articles can be of different types (Figure 6), and some QI interventions are intended to be generalised while others are not. The ability to generalise is usually greater if QI interventions are evaluated using a randomised design and a contemporaneous control across multiple sites.²⁴⁵ When a randomised design is used, variables (or mediators) which might explain generalisability can be described, but less detail is required for variables which needed to be ‘controlled out.’ However, in QI it is less usual for a randomised design to be used, and while the intervention may not be intended to be fully generalisable, authors may wish their QI interventions to be ‘actioned’, ‘scaled up’,⁴⁰⁵ or ‘portable.’ In this case, portability could be used to denote the extent to which an intervention can be translated from one setting into another and still achieve equivalent results.

While the concepts of generalisability and external validity originated in the field of traditional experimental science, in QI these ideas may require some further

thought. The data presented in this thesis suggests that rather than generalising to a wider population of interest when the study is completed, in QI studies translation should occur more iteratively, on a case-by-case basis. 'Porting' an intervention to one population, one site at a time, is more time consuming than generalising an intervention to populations at many sites simultaneously. However, it is a sensible and appropriate option when the chosen study design was not one that might be expected to allow unproblematic generalisability. Avoiding the temptation to generalise all intervention components in a blanket approach could then avoid wasted resources, because it is likely that some ingredients will be more likely to fail in a new setting than others. If the idea of 'porting' QI interventions could be accepted as a 'more modest' use for QI work than generalisability—but an equally valuable one—then QI could become more useful than simply developing 'ideas', because it will become a field which can develop interventions which can be used.

Before a QI intervention can be 'ported' from one setting to another, the researchers must be confident that a QI intervention can *cause* an effect. Causality indicates that a relationship between one variable and another can be demonstrated and is not the result of confounding⁹¹ (a variable which renders data difficult to interpret because it covaries with the variable of interest).²⁴⁵ RCTs are regarded as the gold standard for establishing a causal relationship between one variable and another,⁴¹⁷ but quasi-experimental studies (such as the pre-post designs which are popular in QI (Chapter 4, section 4.2) must also attempt to explain causal links between the intervention and any observed outcome.^{245,418} Thus, while generalisability and causality have been raised as topics which tend to preoccupy quantitative researchers,⁹¹ they are also important in QI. During my qualitative work, some editors made it clear that QI authors rarely manage to provide a convincing account of what caused an observed effect. On the other hand, authors wanted to explain to their readers how much their results could be believed, but they understood that establishing a dependable relationship between cause and effect in QI is difficult.³⁸¹

The task of explaining whether a complex QI intervention, which is influenced by many contextual features, caused an effect could be eased by the use of an appropriate methodology. Methods relate to the techniques that are used to

gather data, but methodologies relate to how the research questions will be answered⁹¹ and they can involve a practical discussion of exactly how a particular research method can answer the research questions.⁴¹⁹ The data obtained from this study are not sufficient to feed into a discussion about the role and adequacy of methodologies used in QI, but the participants did provide much dialogue on another important part of the process of conducting research and understanding causality: the study design.⁹¹ The participants used terms such as population, intervention, comparison and outcome (PICO), in addition to terms such as causality, qualitative, single site or multiple sites and fidelity, which can all be used to illuminate what type of study the QI research is.^{245,420} However, assigning a study design to QI work and applying it in a sophisticated manner is challenging.²⁰

The SQUIRE guidelines state that QI techniques can be implemented either with or without an experimental, quasi-experimental, observational or qualitative design.⁸⁸ The QI literature is also clear that any one of a number of study designs can be used in QI research.³⁰ However, despite this seemingly clear depiction of how study designs can be used in QI, the literature suggests that study designs can be hard to describe in QI. QI techniques can be described as a 'design framework' (p.148),²⁰ or a 'study design',³⁵⁴ and reporting both the QI technique²⁰ and study design may require a sophisticated approach which is not yet firm.²⁶

An examination of the importance and variance of contextual features across the many settings in which QI activities take place illuminates QI research as 'practice orientated' and applied in nature. As such, an exploration of the literature on pragmatic trials could help to increase understanding about study design in the field of QI. Trials are distinct from QI research because they typically aim to test an intervention across several settings,⁷⁸ whereas QI interventions can be implemented in a single site, or several. However, in both QI⁴²¹ (across the whole spectrum of study designs) and in pragmatic (effectiveness) trials, it is of great clinical concern to find out exactly what makes perioperative interventions work (or not work) in usual circumstances. Also, both study types require an explanation of the key contextual features of the work^{23,389} so others can see which characteristics are similar to their own

setting.²¹⁸ Yet, the features of QI studies and pragmatic clinical trials are not always clear cut.

Both QI and pragmatic trials can include a blend of features which are strong in pragmatic studies (such as an increased flexibility with intervention delivery) and features which are strong in explanatory studies (such as tight eligibility criteria). However, in order for this blend of features to be helpful, it must depict a balance which is generally orientated towards one or the other so that the study purpose (which involves answering an explanatory or pragmatic question) can be achieved.⁷⁸ The PRECIS⁷⁸ tool has been developed to help researchers strike the right balance between pragmatic and explanatory designs. It is a visual aid used to score study design features on a scale of 1 (which is very explanatory) to 5 (which is very pragmatic). This tool has raised the profile of different types of study design, and recognises their value as well as the distinction between them. Thus, a similar approach to enable QI editors and authors to recognise the value and the differences between different types of QI study design could serve a valuable purpose. This would involve reconciling stakeholders who hold opposing views, and help QI researchers to hone their study design reporting.

9.2.3.3 Active ingredients

A final feature of reporting that requires careful consideration in QI is the task of isolating active ingredients. Once a decision is made that an intervention can be generalised or ported to another setting, the consumers of the work must apply careful judgement to decide which intervention components should be revised in a new setting and which can be kept the same (Chapter 5, section 5.7.6). The literature is sympathetic to the view that implementing a new intervention by ‘throwing everything at it to see if it sticks’ (p.2) is not the right approach.⁴⁰⁵ Instead, QI researchers need to build up pieces of evidence which can eventually lead to extrapolation of some elements which can be universally replicated (generalised), and some elements which need to be modified. Following the distinction put forward above (Chapter 9, section 9.2.3.1), a QI article could then be written for an intervention to be usable, setting out elements of the intervention in the expectation that they will be replicated almost

exactly, or for an intervention to be used with much more modification in a new setting (Figure 6). This process of extrapolation requires an understanding of which intervention ingredients are the most important, or most active, so that care can be taken to describe how these particularly active ingredients should be translated into a new setting. This could facilitate a more parsimonious use of resources by avoiding a blanket approach which involves generalising all of the intervention components, whether they are likely to work or not. Also, focusing attention on the most important ingredients in the principal publication, while providing detail about other intervention elements elsewhere (such as online supplements), might promote the practice of concise reporting which is deemed to be valuable (Chapter 6, section 6.3.1).

The question of how to isolate active ingredients has been the topic of much debate for researchers of all types of complex interventions, not just QI.^{87,117,252} The literature warns that it is probably much easier to identify which of the *technical* elements of an intervention are the most active ingredients (such as the materials required, for example), than it is to identify the *contextual* aspects which influence effectiveness. Chopra and Shojania⁴²² are doubtful that a researcher could state with confidence that the contextual feature is an active ingredient, not a secondary ‘epiphenomenon.’ However, Campbell et al²¹⁹ propose several solutions, suggesting that reviewing the theoretical basis for an intervention, and conducting preliminary (or pilot) work,²¹⁹ can be mechanisms by which active ingredients might be identified and specified. Campbell et al’s²¹⁹ work was published to enhance the design of clinical trials, but data generated in this PhD suggest that the idea of performing preliminary work could also be used to enhance the design (and later reporting) of QI. In this case the ‘preliminary’ work could involve the continual modification of interventions during iterative cycles during the early stages of a QI study (Chapter 5, section 5.7.4). This resembles a process of conducting many small repeated tests.

During the process of repeated testing in QI, alternative methods of data collection, such as a QI diary and the involvement of staff and patients, could be used as the intervention design is modified and honed. As previously suggested, these methods are highly valued but not unproblematic. To instil a meaningful working relationship with patient partners which adds value to the QI

research requires careful planning to answer questions such as: 'Who to involve?' and 'At what organisational level?'⁴²³ The SQUIRE guidelines suggest that reporting should include a full description of the people who were involved in the QI effort,¹⁴⁹ but they do not explicitly describe the possible involvement of patients. Thus, the emergence of a new field of literature which asks how PPI can be enacted in QI⁴²³ illustrates that QI as a science is maturing, and as it does, new demands will be made of QI reports. Process evaluation was also proposed as a method of adding information to identify what made the intervention work (or not work) (see section 5.6.4), and this is described in the literature as tracking the 'implementation, receipt and setting of an intervention'⁴²⁴ (p.332). It has been regarded as 'unconventional' to publish process evaluation,⁴²⁴ but practice may now be shifting as recent literature explains a method by which second publications, like process evaluations, might be identified.³³⁹

Fidelity, or the degree to which each component (or ingredient) of the intervention was carried out as planned, is known to promote generalisability²³⁰ and the ability to isolate the active ingredients.⁴²⁵ For example, it would be much more difficult to specify an ingredient as being particularly important if it was delivered in very different ways by different people throughout the study. However, modification is an important feature of QI and interventions are often delivered in slightly different ways at different sites.¹⁹² In QI, fidelity can be regarded as being in tension with a requirement to create constant change. Thus, just as a clinical trial may leave discretion to the participating clinicians about the type and route of a procedure for each patient,¹⁶¹ a practical degree of leeway can be given to a QI researcher to choose the method of QI implementation which suits their setting. At the same time, maintaining high fidelity in certain aspects of a study (for example, with the basic tenets of PDSA cycles) can also be very important.⁸¹ Indeed, reporting fidelity (and the features which could moderate it) has helped QI researchers to see which elements of care bundles were key to success.³⁹²

Fidelity and modification were frequently poorly reported in 100 perioperative QI studies (Chapter 4, section 4.3). Given that the participants (Chapter 5, section 5.7.4) and the literature^{81,162} both readily accept the importance of both of these

features in QI, it is unlikely that this omission occurred through lack of memory, or demoted importance of the reporting item, but instead due to uncertainty. An uncertainty about the reporting of fidelity and modifications is indicative of the mal-alignment between conventional medical reporting and QI reporting which has pervaded this study. Understanding how fidelity and modifications might be helpfully reported in QI is not fully developed yet, but opening discussion about this could bear fruit for a maturing field with recognisable unique features.

Finally, a QI article which intends to demonstrate which ingredients should be changed or kept the same should also demonstrate how feasible it will be to do this in another setting. The participants in this study suggested that the deployment of each element of an intervention may require varying amounts and types of resources (including staffing, materials and time), and each element may have differing levels of acceptability for staff and patients. These are often areas of focus within feasibility studies,⁴²⁶ but instead of evaluating feasibility for the onward testing of efficacy in a trial, a feasibility evaluation can be built into QI work to understand how the intervention could be ported to another setting for onward testing or implementation. There are cautionary notes to add to this suggestion. Understanding which interventions (or elements of an intervention) are likely to produce returns may require a heavier burden on data collectors to track costs. Also, it may not be desirable to evaluate the feasibility of implementing the 'socio-adaptive' elements of QI, such as team work, at all, because these elements will inherently always be less feasible to replicate than others.⁴²²

9.2.4 Conclusion – Youth, aspiration and maturity

In proposing the key learning points that QI can simultaneously be a youthful field, an aspiring field and a maturing field, this discussion seeks to move beyond simple explanations of the main problems in reporting, towards a more coherent view of the whole scientific field. QI is characterised as a young field, which is poorly defined and struggles to remain loyal to some reporting features of traditional science while introducing new ones which are more sympathetic to its distinctive features. Identifying the trade-offs that the people and organisations in the field are making on a day-to-day basis exposed the type of

science that stakeholders aspire QI might become. From these discussions, a set of key reporting features were identified which could require special attention in the field of QI and could better distinguish QI as a separate scientific field.

Features which are particularly important in QI include: understanding context; whether an intervention can be generalised or 'ported'; which elements are most likely to bring about change (active ingredients);²¹⁹ and how they function (causal mechanisms).¹⁸⁵ In a maturing field of QI, participants recognised each of these features and began to consider how they could shape scientific reporting. This indicates that establishing a scientific framework within which QI can be reported could be useful, but also, and perhaps most importantly, that QI stakeholders are interested to do this, and to overcome the constraints which have made reporting poor.

Some elements of this scientific framework have already been developed, such as the valuable work which has developed the SQUIRE,^{149,23} MUSIQ³⁹ and TIDieR¹⁵⁰ reporting tools, and broad literature on theory,³⁵⁸ process evaluation⁴²⁷ and qualitative methods.^{309,314} The unique contribution that this PhD seeks to make is that it recognises the value of existing work, and it has identified where the gaps in understanding are. It has embraced what is unique about QI and furthermore, exposed where it might fit amongst existing scientific disciplines. Before final recommendations for practice and research are made, a short discussion will be provided to examine whether data has revealed any particular peculiarities which may be unique to the field of perioperative QI reporting. The limitations of the qualitative approach will also be discussed.

9.3 Peculiarities unique to the field of perioperative QI

During the qualitative work, two peculiarities unique to the field of perioperative care were raised which could exacerbate the problem of reporting in QI: the notion that those who actively publish in surgery tend to view qualitative inquiry as an adjunct to quantitative work, rather than a valuable feature of QI reporting (Chapter 7, section 7.5) and the question of what is deemed to be most attractive to perioperative editors (Chapter 6, section 6.3.2).

There is a long tradition of qualitative research in surgery, but support for the use of qualitative methods in perioperative research has become more emphatic in recent years. Many articles have contributed towards this shift, including work by Donovan et al.¹⁰¹ Donovan et al¹⁰¹ used interviews and audio recording of trial recruitment consultations to improve recruitment in a surgical trial, and the success of this approach has influenced the work of other clinical trialists.^{408,428} The use of qualitative methods in trials has now become more common in surgical research,⁴²⁹ and guidelines have been published which offer practical advice for researchers wishing to embed qualitative research into feasibility studies for RCTs.⁴³⁰ In QI, a recently completed clinical trial which implemented and evaluated QI strategies to improve the emergency laparotomy care pathway, involved qualitative approaches to understanding how the intervention worked.⁴³¹ This study illustrates how surgery and QI are both scientific fields which use different approaches to understanding how interventions work.¹⁹⁹ Thus, while the value of qualitative work was diminished by some participants who saw it as an 'adjunct', other participants regarded it as an 'asset,' and the latter view is becoming more visible in the literature.

The problem that some types of surgical research may be regarded as more attractive than others has been considered in the literature. In 2006, topics related to QI including medical and surgical errors, risk analysis and performance of health systems were regarded as 'promising research topics' that are 'not yet global priorities.'⁴³² QI was placed more in the spotlight following the work to introduce the WHO Surgical Safety Checklist, but persistent problems such as poor data capture, as well as a lack of standardised measurement about contextual features and process of care,⁹³ may perhaps give some the view that other fields of research are easier or more appealing. Policy documents are beginning to reflect a growth of QI work in the literature. For example, in the UK, the Royal College of Surgeons' (RCS) research report of 2015/16 celebrates the growth of clinical trials, but also showcases work aimed at evaluating quality improvement initiatives.¹⁷ Also, a national perioperative quality improvement programme (led by the UK's National Institute for Academic Anaesthesia's Health Services Research

Centre) has been developed to encourage surgical teams to act upon research data and data collected through local QI projects.²¹

9.4 Limitations of the qualitative study

There are some limitations to the qualitative work and some of these were recorded in the study diary I kept throughout the PhD. I recorded my frustration that a participant provided some insightful information about what personally drove him to conduct QI work, while the audio recorder was switched off. Capturing potentially useful data after the closure of an interview encounter is not unusual and further information may be offered because participants feel more able to relax or because they have had a little more time to reflect on what has been said.⁴³³ This particular encounter added value to this thesis because my learning as a researcher, like improvement, can be enriched from recognising when things do not always go according to a perfect plan.

Numbers and counting were used throughout the qualitative data analysis. The purpose of the qualitative work was to identify why reporting QI is hard, and this is most conducive to using rich, in-depth textual data to explore a person's experience of reporting. Yet through qualitative inquiry, it is helpful to understand which views were held most commonly and which were the most frequently occurring reasons for poor reporting. There are mixed views within the qualitative research community about the legitimacy of quantifying qualitative data, but Seale and Silverman³¹⁴ suggest that qualitative research can be more rigorous if a researcher counts specific events or items within transcripts to understand how widespread each feature was across the dataset. Some qualitative research (such as that which uses a phenomenological approach) places a high value on the viewpoint of a smaller number of individuals,⁴³⁴ but this study aimed to examine data across a variety of QI stakeholders and it is reassuring to know that conclusions based on the opinion of one or two people were not drawn.

The qualitative work only targeted the recruitment of surgeons who were actively publishing or interested in consuming QI work. A cohort of surgeons who objected to the conduct and publication of QI work could have added some interesting data about the problems of reporting. However, it was possible to

obtain wide-ranging data with a balance of both positive and negative views about QI reporting because the semi-structured interviews were relatively unobtrusive and gave the participants many opportunities to talk freely.

There were limitations specifically associated with the author checking exercise. The issue of recall was raised by two of the Lean authors because the Lean article was published six years ago. Poor recall could threaten the ability of the authors to fully ascertain whether the participants' views were aligned with what happened. However, recall problems represent a real-world problem, that when research consumers contact authors for information they may not always remember it, and an attempt to include authors with a more perfect recall may have produced an unrealistic view of what happens in practice. Second, my systematic review demonstrates that the five years between 2008 and 2013 represented the period of the most growth in the publication perioperative QI research, with a total of 68 articles published during this period. Choosing an article published in 2010 represents part of a large pool of research which QI consumers are likely to use.

9.5 Recommendations for practice

By attending closely to the needs of those who might use the research, and the needs of those who produce it, I have produced recommendations which can be used by a variety of QI stakeholders.

This PhD has presented three key learning points to explain why reporting QI is hard in perioperative care:

- QI is a young scientific field.
- QI stakeholders do not yet hold a shared aspiration about what type of science QI should become.
- QI is characterised by features which are unique to this field.

From these learning points, practical insights have been generated into what is required in practice to create an environment in which improved reporting might flourish. These insights have been discussed in the context of the literature and from this, a set of recommendations for practice has been developed which is

presented in Table 14. Each recommendation relates to the article format, organisational infrastructure and scientific outputs.

These recommendations relate to the interconnecting needs of the individual QI authors, QI teams, healthcare journals and hospitals. Some of these recommendations are very practical, and may already be in use; for example, a great many authors use the SQUIRE guidelines. Other recommendations derived from my data are a form of 'blue sky thinking' and may be more challenging to enact. These recommendations relate to a desire to bring about reconciliation between groups of authors, editors and consumers who do not agree on how QI should be presented in a scientific manner. Just as organisational culture can be challenging during a QI project (Chapter 7, section 7.4), it can also pose challenges to the process of improving reporting. Moving away from the perception that getting QI published requires a 'neat trick', towards building an accepted and recognised scientific framework for QI is likely to require a culture change. Culture change can require:

"Intensive, strenuous, stressful, and sometimes coercive training and the immersive experience of a more subtle but equally strong social pressures over a considerable period" (p.563).⁴³⁵

Thus, I do not suppose that a simple mix of education and reporting checklists can overcome this problem, but rather it requires a prolonged and co-ordinated effort from a team of QI stakeholders acting together.

Table 14 Recommendations for practice

	Recommendations for authors of QI research	Recommendations for journal editors who publish QI work	Recommendations for healthcare organisations who deliver QI work
Article format	Use existing reporting guidelines and taxonomies to guide the structure of your QI report.	Ensure familiarity of editorial staff and peer reviewers with QI reporting tools.	
	Know your audience. Do you want your reader to use it to generate ideas for a new intervention? To replicate your intervention exactly? To modify your intervention?	Provide a clear statement about whether qualitative approaches to data collection and writing are acceptable.	
	Use supplementary materials, embed URLs (web links) into the article where possible.	Provide a clear statement of which additional resources are available to authors (e.g. on-line supplements).	
	Be available to speak to your readers, add your contact details to your published article.	Support the open access movement to encourage connection between authors and consumers.	

	Recommendations for authors of QI research	Recommendations for journal editors who publish QI work	Recommendations for healthcare organisations who deliver QI work
Organisational infrastructure		Sustain open communication channels with QI authors and consumers about what QI is and how it should be reported.	Build internal support and capacity for QI (such as protected time to conduct QI and more formal relationships between clinical QI teams and research nurses).
	Consider: using a multi-disciplinary writing team; how to support patient involvement; and external evaluation.		Build networks with external academic organisations (such as universities) and patients.
	Work with hospital management to identify problems which are most relevant to patients (enable a breadth of topics).		Work with QI teams to identify problems which are most relevant to patients (enable a breadth of topics).
	Consider enrolling in an education programme to enhance your QI reporting.	Consider providing some educational material for editors and peer reviewers about QI.	Embed specific training about QI in library training programmes, on-line training programmes or mentorship schemes.

	Recommendations for authors of QI research	Recommendations for journal editors who publish QI work	Recommendations for healthcare organisations who deliver QI work
Scientific outputs	Demonstrate why your intervention was thought to work (for example, consider using theory, process evaluation, a QI diary).	Enable structured conversations with QI stakeholders to consider how QI can be reported and what good reporting in QI looks like.	
	Provide your reader with a realistic view of what is needed and what is feasible.		
	Consider submitting a QI project which did not go well for publication.	Give specific advice on how to write a negative study well.	Support a culture where negative experiences which create learning are celebrated.

9.6 Recommendations for further research

Further research, involving tightening the scientific framework which should be used to report QI work, could be beneficial. The systematic review and qualitative study conducted during this PhD have identified reporting in QI is poor because disagreement exists between QI stakeholders about how QI should be reported as a science. Specifically:

- How much detail should be afforded to context and how should its interaction with the intervention be described?
- What types of qualitative approach to data collection and reporting can be used?
- What can QI research be used for (should it be generalisable, portable or useful for generating ideas?)
- How can active ingredients be identified?

This PhD has also identified agreement between QI stakeholders that much work has already been conducted (such as the production of the SQUIRE guidelines) and existing resources should be used. However, a scientific framework to support QI reporting is only partially complete, and this is strangling the attempts of authors to get their work published and recognised as a useful and rigorous contribution to knowledge.

A new package of work which aims to reach consensus on areas of uncertainty in scientific QI reporting would be beneficial. This work would begin with a literature scoping exercise to capture any new reporting tools for QI that may have been published. For example, a taxonomy of quality-related interventions has been developed by Malloy et al⁴⁸ which distinguishes between ‘people’ and ‘systems’ based interventions; this was not available when the systematic review (Chapter 3, section 3.3) was designed. Then, a Delphi study conducted to classify, prioritise and obtain consensus on areas of uncertainty identified in this thesis could lead to the development of a complete scientific framework which could be used to report QI. A further study could pilot the utility of such a framework with a group of QI stakeholders.

A whole comprehensive, coherent and valid framework for the design and reporting of quality interventions and QI techniques could improve reporting. As

a consequence, in the long term, the contribution that QI studies could make to improving care and reducing harm^{53,149} for all patients who undergo surgery each year might be enhanced.

APPENDICES

Appendix 1: Chapter 2: Poor reporting in QI - a list of 11 systematic reviews

	Reviews on reporting quality in QI	
1.	Levy et al (2013)	What is the quality of reporting of studies of interventions to increase compliance with antibiotic prophylaxis? J Am Coll Surg 2013 Nov;217(5):770–9 ⁷
2.	Taylor et al (2014)	Systematic review of the application of the Plan-Do-Study-Act method to improve quality in healthcare. BMJ Qual Saf. 2014 Apr;23(4):290–8 ⁸
3.	Ivers et al (2013)	Quality improvement needed in quality improvement randomised trials: systematic review of interventions to improve care in diabetes. BMJ Open 2013 Jan;3(4) ¹²⁴
	Reviews on the application and efficacy of QI in surgery	
4.	Nicolay et al (2012)	Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. Br J Surg. 2012 Mar;99(3):324–35 ³
5.	Powell et al (2009)	A systematic narrative review of quality improvement models in health care. NHS Quality Improvement Scotland. 2009 ⁴
6.	Steinman et al (2006)	Improving Antibiotic Selection: A systematic review and quantitative analysis of quality improvement strategies. Med Care. 2006;44(7):617–28 ⁵
7.	Alexander et al (2009)	Review: What Can We Learn From Quality Improvement Research?: A Critical Review of Research Methods. Medical Care Research and Review. 2009 ²⁶

	Reviews on the application and efficacy of QI in surgery	
8.	Treadwell et al (2014)	Surgical checklists: a systematic review of impacts and implementation. BMJ Qual Saf. 2014 Apr;23(4):299–318 ²⁸
9.	Howell et al (2014)	Reducing the Burden of Surgical Harm: A systematic review of the interventions used to reduce adverse events in surgery. Ann Surg 2014;259(4):630–41 ²⁹
10	Koetsier et al (2012)	Control charts in healthcare quality improvement. A systematic review on adherence to methodological criteria. Methods Inf Med. 2012 Jan;51(3):189–98 ¹²⁵
11	Thor et al (2007)	Application of statistical process control in healthcare improvement: systematic review. Qual Saf Health Care 2007 Oct [cited 2014 Nov 3];16(5):387–99 ²⁷

Appendix 2: Standard letter of invitation



**University of
Leicester**

Department of Health Sciences

College of Medicine,

Biological Sciences & Psychology

22-28 Princess Road West

Leicester LE1 6TP

UK

Date

Name and address of person

Dear

I am conducting an interview study which aims to find out why reporting quality improvement (QI) in surgery can be difficult and how it can be improved. This work is being conducted as part of my University of Leicester PhD in collaboration with the Health Foundation, and is supervised by Professors Mary Dixon-Woods and Graham Martin. I am attaching some information about this study. Details are also available on the project website www.le.ac.uk/qi-project-emma-jones.

I aim to speak participants in a recorded telephone interview for around 45 minutes. I hope to speak with:

- Authors of surgical QI papers (published since 2001)
- People in managerial or clinical roles who are using, or have used QI reports to improve their surgical services in the last 24 months
- People who have been involved in decisions about publishing QI reports (e.g. in an editorial or reviewer role), or in the design of reporting guidance in the last 24 months.

If you feel you are one of these people, I would be very keen to speak with you and would be delighted if you were happy to take part in my study. If you would be willing to be interviewed please let me know by emailing me (elj20@le.ac.uk) or phoning (0116 2523648). You can also register your interest using the on-line registration form at the project website www.le.ac.uk/qi-project-emma-jones

Many thanks for considering this request. It would really help me if I could speak with you and I look forward to hearing from you.

Yours sincerely,

Emma Jones

Health Foundation Doctoral Fellow, SAPPHIRE Group

Appendix 3: Participant information sheet



Page 1: Strengthening Implementation of Quality Improvement Interventions and Methods in Surgery

Strengthening Implementation of Quality Improvement Interventions and Methods in Surgery **SIQINS** Participant Information Leaflet

A study to explore what is needed to improve how we communicate about surgical projects which aim to improve patient care.

You are being invited to take part in an interview study. Before you decide, it is important for you to understand why the study is being undertaken and what it means for you. Please take time to read this and feel free to discuss it with others. If you would like more information please ask Emma Jones: elj20@le.ac.uk, 0116 252 3648.

What is the purpose of this study?

Quality improvement (QI) approaches are used to make changes that can lead to better patient outcomes and better system performance across the entire perioperative journey, such as reducing infection rates, preventing prescribing errors, and improving discharge planning. But we know that when QI projects are published in academic journals, important information may be missing. We want to find out why this happens and to learn about how QI reporting might be improved.



This study aims to explore why you think problems occur in reporting of QI and what could help to improve QI reporting.

The following information describes what will be involved if you decide to take part. Please read the information carefully and discuss it with others if you wish. Please ask us if there is anything that is not clear or if you would like more information (contact details for the researchers are at the end of this information sheet). Take time to decide whether or not you wish to take part.

Why have you been chosen?

You are being asked to participate because you write up quality improvement (QI) projects for publication in the academic literature, you have an editorial role in journals that publish quality improvement reports, you are involved in writing reporting guidelines, or you use quality improvement reports to guide changes in care. Therefore, you may have views about what is important in reporting, why reporting might be difficult, and how it might be improved.

Where is the study being conducted and what does it involve?

This study is being conducted across universities and healthcare organisations with an interest in QI. You will be asked to take part in one telephone interview for a maximum of one hour. Emma will ask you why you feel that reporting might be difficult. During the interview we will also talk about excerpts taken from an anonymised published QI report. You will be asked to describe whether the QI report would provide you with sufficient information to recreate the intervention and the conditions that made it work. The excerpt from the paper will be given to you two weeks before your interview. If you have time to read it in advance this would be good, but if you can't, don't worry, you can do this on the day. Your interview will be conducted at a time which is convenient to you. The interview will be recorded and then transcribed word for word. The transcript will be anonymised so that you cannot be identified. Your interview transcript may then be used during a separate interview with the author of the QI report. This will help us to assess congruence between your interpretation of the paper and the authors' understanding of what was done. Emma will ask your specific consent to use your interview transcript in this way. Participation is completely voluntary and if you do not wish to take part, that's fine. If you do decide to take part you are also free to withdraw at any time and you do not need to give a reason.

What are the possible risks and benefits of taking part?

The interview will involve a small time commitment. No sensitive information will be involved. The information collected in this research will be used to increase knowledge of intervention reporting in QI and could help to influence future best practice.

Confidentiality

All the information collected from you will be kept confidential. Only Emma will have access to data which can identify you or the organisation you work for and this will be kept on a secure password protected University computer. All recording equipment will be encrypted and it will be kept in a locked drawer inside the University of Leicester's Health Sciences Department. In any written report you will be identified by number, not name.

Representatives from the Sponsor (University of Leicester) and / or NHS Trusts may access your research data for monitoring and audit purposes.

Who has reviewed the project?

This research has been reviewed through the University of Leicester's ethics system.

What if there is a problem?

If you have any concerns please contact any member of the project team. We will do our best to answer your questions. If you remain unhappy and wish to complain formally, you can do this through the University of Leicester Complaints procedure.

What will happen to the results of the research study?

The findings from this research will be written up as a report within the PhD thesis of Emma Jones, and may be used in papers submitted to professional journals and conferences. We may also place the report and summaries on the internet. None of these verbal or written outputs will be able to identify you personally.

Research Team



Professor
Mary Dixon-Woods



Professor
Graham Martin



Miss Emma Jones

Mary Dixon-Woods, Professor of Medical Sociology and director of the SAPHIRE group, Department of Health Sciences, University of Leicester. md11@le.ac.uk

Graham Martin, Professor of Health Organisation and Policy,, Department of Health Sciences, University of Leicester gpm7@le.ac.uk

Emma Jones, Doctoral Research candidate, Department of Health Sciences, University of Leicester, Practicing Extended Scope Orthopaedic Physiotherapist, Surgical directorate, University Hospitals Leicester NHS Trust. elj20@le.ac.uk, 0116 2523648

Thank you for taking the time to read this information leaflet.

Appendix 4: Participant interview schedule

The primary objective of the interview is to explore:

“Why is the reporting of QI in perioperative care so hard?”

The following questions need not be covered in this particular order but rather the interview should flow as freely and naturally as possible. The interviewer will prompt with phrases such as ‘can you tell me a little more about that’, ‘can you give me an example’

Topic	Questions
OPENER	Thank you for participating within this PhD study. My name is Emma Jones. I am a PhD student at the Department of Health Sciences at the University of Leicester and I am also a practicing orthopaedic physiotherapist. I will be asking you questions for about 45 minutes but we'll try to keep it shorter than this if we can. I will ask about what you think about QI reporting in surgery. I value your views so please speak as freely as you wish. Everything is important; even if you think it's a little thing, don't be shy about telling me those little things! Please be assured that all of your responses will be anonymised and you will not be able to be identified by any written material relating to this interview. You are free to stop the interview at any time.
BEFORE WE START	<ul style="list-style-type: none">• Have you read the information sheet?• Is it OK if I record the interview?• I'd just like to check that you consent for your anonymised transcript to be shared with the author of the QI report I have given you?

	<ul style="list-style-type: none"> • Have you got any questions about the project?
TELL ME A BIT ABOUT YOURSELF (demographic info)	<ul style="list-style-type: none"> • Please confirm your job title • Which organisation are you working for? • If you have a clinical background which professional group do you belong to and can you confirm your clinical speciality? • How much experience do you have in QI? If you have any publications which focus on QI could you estimate how many?
UNDERSTANDING OF WHAT COUNTS AS QI, WHAT IS IMPORTANT TO THEM AND WHAT IS THE IMPACT OF GETTING REPORTING RIGHT?	<ul style="list-style-type: none"> • To allow me to understand more fully what you mean when we discuss quality improvement, can you tell me what the word quality improvement means to you? • What makes an intervention a 'QI' intervention as opposed to a different kind of intervention? • What about PDSA, Lean, or SPC - what are they? • Can you think of a really good QI report you read recently, what did you like about it? <i>(probe this)</i> • How important is it to you that a QI report is written in such a way that someone reading it can re-create it in another setting? Why is replication important? • We talked about replication, but is there anything else that is important to you about how QI is reported? • How have you used published QI reports in your organisation? <i>(to change care? to realise they are doing well and celebrate their own success? CPD? Design of their own QI studies?)</i>

TASK: Instructions to participants.

OPENER	<p>“We want to check how people understand reported QI interventions as they are presented in a journal article. To do this I have given you a piece of writing which has been highlighted. I’d like you to have a look at this paper. You only need to look at the parts highlighted in yellow, though you may look at the rest of it if you wish. I am going to ask you to describe to me what you would do if you needed to implement the intervention they describe in the paper. Don’t worry about criticising this paper, it’s OK to do that. I will anonymise your responses, and you will not be able to be identified from any written reports or transcripts which are produced. Have you had an opportunity to look at the paper I sent you?”</p>
UNDERSTANDING	<ul style="list-style-type: none">• What was your overall view of this paper?• If you needed to use one of the interventions described in this report, do you feel you could do what they did?• Tell me what the ingredients of this QI intervention are? (as if you are going to deliver it in your own hospital)• What were the most helpful things about how QI was reported which might enable you to replicate what they did?• What were the main problems with how QI was reported?• Did the authors give you a main message about what they felt made their QI intervention work?• Do you think the authors aimed to say “It is possible to use Lean in emergency surgical

	care”, or did they aim to say “Lean can be used in emergency care and this is how you do it?”
EVAUATION OF QUALITY	<ul style="list-style-type: none"> • There might be both good and bad things about this paper, but if a journal editor needs to evaluate if this QI paper should be published or not, which criteria should they use to make that judgement? What were the good things about it? What were the main problems?
HOW COULD THIS BE IMPROVED?	<ul style="list-style-type: none"> • Perhaps this isn’t a perfect way for the authors to get their message across and disseminating their information, is there a better way of doing it?

Putting this paper to one side now, and thinking more widely about your general experience of reading QI articles....

WHY IS REPORTING POOR	<ul style="list-style-type: none"> • In your general experience which bits of QI projects tend to be better reported than others? • In your experience, what makes reporting harder? What is most problematic? How do you deal with that? • Is there anything specific to QI which can’t be reported fully? • Which are <i>the most important</i> things to report to ensure we have a clear and accurate picture of what happened during the QI project? • There are many possible ingredients that could make up the nuts and bolts of a QI intervention, how will you know which of these ingredients is the most <i>active</i> ingredient in making the improvement happen –how can you tell which ingredient really nailed it and made the intervention work?
------------------------------	---

	<ul style="list-style-type: none"> • If you have a reasonable idea of what the intervention was, what extra information do you need to know to decide whether you <i>should or should not</i> use the intervention? • There might be many different people working on a QI project and each of them might have a different perspective on what made it work. How can the lead QI authors ensure they have all of this information? • Some QI projects are conducted as part of a randomised study, and some are conducted as a local project. Does the study design determine how much detail is given about the intervention and the type of information you would expect to see?
INFORM DEBATE ABOUT HOW QI REPORTING SHOULD BE SUPPORTED	<ul style="list-style-type: none"> • What is the most important thing to do to make reporting <i>better or easier</i> for authors? (<i>Encourage them to notice that better and easier are different aims!</i>) • What motivates people to complete good reporting?/How can we influence reporting practice? (<i>Encourage them to find lots of points—financial incentive? Kudos? Integrated into education? A standardised definition to agree what constitutes a QI intervention? A standardised structure or checklist? Expected practice to write a ‘lay’ summary?</i>) • What is the impact of improving reporting? Why is it important? • Number of citations is one type of outcome which is used to measure the success of publications. Are there any other outcomes

	which should be used to judge the impact of QI reporting?
EXIT QUESTION	<ul style="list-style-type: none"> • 'Is there anything else you would like to tell me?' • Do you wish to be informed of any publications relating to the research? • Thank you again for your participation.

Possible question at the end:

Q What will happen to my data?

A I will go away and write it up as part of my thesis, with your permission I will anonymise your entire transcript and take sections of it to share with the authors of the report we discussed to see if your understanding of the report matches what they did. I will also produce publications, within these you will not be able to be identified. Would you like to see any publications which arise as a result of my thesis? Your data will only be stored for 3 – 6 months after the study has ended and it will then be destroyed.

Prompts:

Keep going, Keep talking

Why do you think that?

Yes? Go on? Mmmmm?

Tell me more about that

Appendix 5: The highlighted and anonymised Lean article used during participant interviews

For the full versions of these articles see [REDACTED]

PRACTICE

QUALITY IMPROVEMENT REPORT

Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit

Abstract

Problem Emergency surgical patients are at high risk for harm because of errors in care. Quality improvement methods that involve process redesign, such as “Lean,” appear to improve service reliability and efficiency in healthcare.

Design Interrupted time series.

Setting The emergency general surgery ward of a university hospital in [REDACTED]

Key measures for improvement Seven safety-relevant care processes.

Strategy for change A Lean intervention targeting five of the seven care processes relevant to patient safety.

Effects of change 969 patients were admitted during the four-month study period before the introduction of the Lean intervention (May to August 2007), and 1114 were admitted during the four-month period after completion of the intervention (May to August 2008). Compliance with the five process measures targeted for Lean intervention (but not the two that were not) improved significantly (relative improvement 28% to 149%; $P < 0.007$). Excellent compliance continued at least 10 months after active intervention ceased. The proportion of patients requiring transfer to other wards fell from 27% to 20% ($P < 0.000025$). Rates of adverse events and potential adverse events were unchanged, except for a significant reduction in new safety events after transfer to other wards ($P < 0.028$). Most adverse events and potential adverse events were owing to delays in investigation and treatment caused by factors outside the ward being evaluated.

Lessons learnt Lean can substantially and simultaneously improve compliance with a bundle of safety-related processes. Given the interconnected nature of hospital care, this strategy might not translate into improvements in safety outcomes unless a system-wide approach is adopted to remove barriers to change.

Background

Surgical wards are an area of vulnerability in healthcare systems because patients are much less closely monitored, and patients usually stay on surgical wards during the time when the risk of operative complications is highest. It is now widely accepted that high technology surgical care causes unintentional harm to around 10% of inpatients.¹⁻⁴

Substantial improvement to reduce unintentional harm is likely only if the system of care is redesigned to eliminate

The five principles of “Lean”

- “The five S’s” (sort; set in order; shine; standardise; and sustain): the re-ordering of the working environment to clarify and simplify process and to reduce time wasted in finding supplies and equipment⁶
- Process mapping: identification of problems to facilitate directed measurement and improvement¹⁵
- Error visibility: development of audit methods to make problems instantly obvious^{16,17}
- Elimination of waste: minimising waste and wasted effort to make work easier¹⁸
- “Plan, do, check, act” (PDCA): introducing cycles that allow rapid refinement of interventions,¹⁹ and defining problems and solutions by measurement rather than assumption

small contributory errors across a wide range of care processes, which is relatively common in other industries but rare in healthcare.

The most successful example of system redesign in industrial settings has been the Toyota Production System, or “Lean.”⁵⁻⁸ Defining Lean is difficult: it is in essence the elimination of waste through continuous improvement (box). Identifying waste leads inevitably to the need to define customer value, and reducing waste requires elimination of error. In medical settings, there is extensive evidence of the benefits of Lean in improving efficiency, reducing costs, and improving patient satisfaction.⁹⁻¹² Few studies have directly measured the impact of Lean on patient safety.¹³

The problem

The surgical emergency unit at the [REDACTED] is a 38-bed acute surgical ward. It receives all general surgical emergency admissions to the hospital, as well as a small number of elective surgical patients. The unit is typically fully occupied, with an average turnover of 10 patients a day and a staffing ratio of one trained member of staff to six patients. The mean length of stay is 3.2 days.

The unit is served by 10 consultant teams involved in the general surgery on-call service. Patients are reviewed by a consultant team the morning after admission and a care plan confirmed.

We developed the hypothesis that process redesign using Lean with a focus on improving safety-relevant care processes might significantly reduce the risk of care-related harm to patients on surgical wards.

1043

Appendix 6: Lean authors participant information sheet: Author checking exercise

Strengthening Implementation of Quality Improvement Interventions and Methods in Surgery SIQINS Surgical Team Information Leaflet

A study to explore what is needed to improve how we communicate about surgical projects which aim to improve patient care.

You are being invited to take part in an interview study. Before you decide, it is important for you to understand why the study is being undertaken and what it means for you. Please take time to read this and feel free to discuss it with others. If you would like more information please ask Emma Jones: elj20@le.ac.uk, 0116 252 3648.

What is the purpose of this study?

Quality improvement (QI) approaches are used to make changes that can lead to better patient outcomes and better system performance across the entire perioperative journey, such as reducing infection rates, preventing prescribing errors, and improving discharge planning. But we know that when QI projects are published in academic journals, important information may be missing. As part of a study to find out why reporting in QI is hard, we will ask approximately 30 individuals to read an account of the QI intervention in this paper: McCulloch et al Effect of a "Lean" intervention to improve safety processes and outcomes on a surgical emergency unit. Br Med J. 2010;2; 341.



This part of the study aims to check with you whether our participants *correctly* interpreted your QI paper, *which parts* were misunderstood, and in what ways. This will help us to find out which parts of reporting are the hardest, and what might be needed to improve it.

The following information describes what will be involved if you decide to take part. Please read the information carefully and discuss it with others if you wish. Please ask us if there is anything that is not clear or if you would like more information (contact details for the researchers are at the end of this information sheet). Take time to decide whether or not you wish to take part.

Why have you been chosen?

You are being asked to participate because you were part of the authorship team of the QI paper we asked our participants to read. We anonymised your paper so that you and your institution could not be identified. During individual interviews the participants will be asked to describe whether they feel they could, in their own settings, re-create the intervention described in the paper and the conditions that made it work. We will anonymise selected excerpts from the participant's transcripts, and place them into a grid to show you. We will gain consent from participants to share this information with you and Emma will provide you with this grid in advance of the interview. If you have time to read it in advance this would be good, but if you can't, don't worry, you can do this on the day.

Where is the study being conducted and what does it involve?

Emma will conduct an interview with you, which will last for approximately 45 minutes. The interview can be conducted in person or over the phone, whichever is most convenient for you. Emma will ask you: which parts of the intervention and the QI method did the participants interpret correctly and incorrectly? This will help us to assess congruence between your interpretation of the paper and the participants understanding of what was done. Emma will also discuss with you: whether you think reporting is hard; why this might be; and what is needed to improve it. The interview will be recorded and then transcribed word for word. The transcript will be anonymised so that you cannot be identified. Participation is completely voluntary and if you do not wish to take part, that's fine. If you do decide to take part you are also free to withdraw at any time and you do not need to give a reason.

What are the possible risks and benefits of taking part?

The interview will involve a small time commitment. No sensitive information will be involved. The information collected in this research will be used to increase knowledge of intervention reporting in QI and could help to influence future best practice.

Confidentiality

All the information collected from you will be kept confidential. Only Emma will have access to data which can identify you or the organisation you work for and this will be kept on a secure password protected University computer. All recording equipment will be encrypted and it will be kept in a locked drawer inside the University of Leicester's Health Sciences Department. In any written report you will be identified by number, not name. Representatives from the Sponsor (University of Leicester) and / or NHS Trusts may access your research data for monitoring and audit purposes.

Who has reviewed the project?

This research has been reviewed through the University of Leicester's ethics system.

What if there is a problem?

If you have any concerns please contact any member of the project team. We will do our best to answer your questions. If you remain unhappy and wish to complain formally, you can do this through the University of Leicester Complaints procedure.

What will happen to the results of the research study?

The findings from this research will be written up as a report within the PhD thesis of Emma Jones, and may be used in papers submitted to professional journals and conferences. We may also place the report and summaries on the internet. None of these verbal or written outputs will be able to identify you personally.

Please do not hesitate to ask Emma any questions using the contact details below.

Research Team



Professor

Mary Dixon-Woods



Professor

Graham Martin



Miss Emma Jones

Mary Dixon-Woods, Professor of Medical Sociology and director of the SAPHIRE group, Department of Health Sciences, University of Leicester. md11@le.ac.uk

Graham Martin, Professor of Health Organisation and Policy,, Department of Health Sciences, University of Leicester gpm7@le.ac.uk

Emma Jones, Doctoral Research candidate, Department of Health Sciences, University of Leicester, Practicing Extended Scope Orthopaedic Physiotherapist, Surgical directorate, University Hospitals Leicester NHS Trust. elj20@le.ac.uk, 0116 2523648

Thank you for taking the time to read this information leaflet.

Appendix 7: Lean authors' interview schedule

The primary objective of the interview is to explore:

“Why is the reporting of QI in perioperative care so hard?”

Topic	Questions
OPENER	Thank you for participating within this PhD study. My name is Emma Jones and I will be asking you questions for about 45 minutes about what you think about the way the QI intervention which was described in your paper has been discussed by others. I value your views so please speak as freely as you wish.
BEFORE WE START	<ul style="list-style-type: none">• Is it OK if I record the interview?• Be assured that all of your responses will be anonymised and you will not be able to be identified by any written material relating to this interview. You are free to stop the interview at any time.• Have you got any questions about the project?• If you are happy I'll turn the tape recorder on now.
TELL ME A BIT ABOUT YOURSELF (demographics)	<ul style="list-style-type: none">• Please confirm your job title and a bit about your experience with QI?• Are you a clinician? can you confirm your surgical speciality?• How many years of experience in QI do you have?• Could you tell me what the term 'Quality Improvement' means to you?• What are your thoughts about the word 'replication' – should published accounts of QI be 'replicable'?

CHECKING PARTICIPANTS UNDERSTANDING	<ul style="list-style-type: none"> • We are going to move on to look at the excerpts in table 1 • What was it like to read these excerpts, Interesting? Helpful? • Your paper was published six years ago, but people can still read it. Although time goes by - people may still want to use the paper but the author may find it difficult to recall. Does this matter? Should we try to solve this problem?
--	---

Further Questions relating to table 1 (excerpts form the participant interviews):

	Theme	Excerpts from interviews with QI stakeholders who discussed the QI report: <i>McCulloch et al (2010). Br Med J. 2010;2; 341</i>	
1.	Word count/Journal requirements	<p>1. “A journal article in a way doesn’t do it any justice. We hamstring them by limiting them to 3000 words” [09]</p> <p>1. “We have limitations in what we can publish and how we can publish it, because clinical journals are more used to, I mean it doesn’t take long to describe a drug trial, right?” [13]</p> <p>2. “If you’re writing up a quality improvement project within a randomised control trial for the Lancet you’re going to be following a more traditional approach which would prevent you from explaining things in as much detail as you might like.” [10]</p> <p>3. “They’re just going through the steps of what they need to do to publish – ‘I feel I’ve done something important that I want to share, I’ve invested loads of time in it, what are the steps I need to follow in order to get it out there?’” [01]</p>	<p>[09] These participants describe the possible frustration of limited word count and journal style. Does this frustrate you when you write about QI?</p> <p>[13]/[10]/[01] Allude to the notion that journals require a particular style of writing. Can you tell me what you think about that? Is this a help or a hindrance for QI?</p> <p>[24] What did you think about this? Alludes to the type of knowledge that journal editors might already have – is it a problem?</p>

		4. "There are a huge number of journals. Some are very focused on quality improvement, but others are not, and yet all of them will possibly publish quality improvement research. The challenge there, is that the reviewers and the editors of journals may not have a sophistication around some of the challenges and methods in quality improvement" [24]	
2.	The character of QI work	<p>1 "Quality improvement tends to focus on time series work, but the traditional evidence based medicine pyramid doesn't recognise a time series study design because that wasn't the paradigm." [24]</p> <p>2 "Quality improvement is more real-world projects, it's not research. Quality improvement inherently has all the biases. It'll have clinical biases, selection biases, and reporting bias" [14]</p>	<p>[14] Is QI research?</p> <p>[24] There is something here about how sophisticated the study is perceived to be - should more recognition be given to study designs used in QI?</p> <p>Would this help QI work to be recognised and spread?</p>
3.	Description of the materials used	1 "A nurse-led exercise to improve inter-shift handover was introduced. The nurses presumably led the exercise I am not exactly sure what the exercise was. It says the bedside handover process but I don't know what that is. The 'basic care checklist' doesn't tell me what kinds of things are on the check list or what it looked like or, who uses it (I suppose the nurses use it). And a track and trigger chart, I don't know what that is" [09]	<p>[01] Should the extra detail about what the checklist looked like have been included in the paper... or a mechanism by which these readers can access it?</p> <p>What is the most important thing to do to ensure your work can be used by someone else?</p>

		<p>2 “I’m not sure what the education was on. It says other areas that tackles education but I don’t know what the prompt cards are for or what the education was for.” [01]</p> <p>3 “A simple protocol was used to allocate roles and responsibilities. But they don’t really describe exactly what that protocol is for or how they improved it” [07]</p> <p>“They talk about using the data to drive the analysis and I’m assuming that that means that they’re feeding it back to people“ [01]</p>	
4.	Replication	<p>1 “The fact that there was a table, with a lot of description for a medical journal, of what was done, that’s what I found useful because I could take this and I could think about recreating a <i>version</i> of this” [13]</p> <p>2 “I couldn’t necessarily reproduce it, but the key thing for me is that with quality improvement I wouldn’t necessarily want to reproduce identically” [05]</p>	<p>Should we aim to replicate the study exactly, or a <i>version</i> of it?</p> <p>What was the purpose of your paper (what were you writing it for)?</p> <p>Prompts: To encourage others to replicate? To raise awareness of a problem?</p> <p><i>A lot of description for a medical journal</i>” – are medical journals restrictive? Is this a problem?</p>

5.	Context and second publications	<p>1 “I would be more curious as to how they adopted the implementation or improvement strategy that they used and what challenges emerged for them” [01]</p> <p>2 “I was still curious that whether or not it went smoothly, they don’t talk about any kind of push back” [09]</p> <p>3 “The way that it was presented seemed like everything was done very smoothly and they didn’t really have any issues. But it’s a little bit surprising, because they were working on seven or eight different interventions at once which could be difficult” [07]</p> <p>4 “I don’t know for example whether this study that we’re looking at here ever had another publication to describe the local context and issues around how, the reality of how these interventions were actually done” [13]</p>	<p>Some of our interview findings suggest that the ‘sister’ paper you published three years later is good because context could be described more thoroughly.</p> <p>Should authors of QI work always be encouraged to write a second paper? Prompts: <i>how should this be encouraged? A reporting requirement? on-line as a ‘supplementary piece? Why?</i></p> <p>Should these <i>second</i> articles be published earlier?</p>
6.	QI Method	<p>1 “It doesn’t imply that anything was learned or any changes were made as a result of the learning so I guess that’s why I’m saying ‘did you really do Lean?’” [01]</p> <p>2 “It’s not about DVT prophylaxis. It’s more about the LEAN intervention than it is about the DVT” [05]</p> <p>3 “It sounds like they were interested in implementing not one particular intervention. They wanted to implement the whole LEAN model into the surgical unit. They wanted to,</p>	<p>[01] Readers and authors might have a different understanding of what LEAN entails – is this a barrier to replication? Should attempts be made to overcome it?</p> <p>[05/07]Some participants felt that the paper described LEAN as an overarching intervention. Is this what you wanted the reader to perceive?</p>

		kind of, create a culture of continuous quality improvement” [07]	
7.	Theory underlying the intervention	<p>1 “7 routine care processes were selected. It is always curious to me how they get selected, so that is not something that they did well” [09]</p> <p>2 “They maybe could have been a little bit more detailed in talking about how they chose the different interventions” [07]</p> <p>3 “One of the things that I liked about it was that they focused on the things that the staff were interested in improving, which I think is very important. So I think that’s good advice when you’re putting something like this into effect.” [07]</p>	<p>Some participants thought more detail on rationale to explain <i>why</i> an intervention was selected was needed. Others thought the rationale was linked to what the staff wanted to improve. Did you specifically think about rationale when you were writing this up?</p>

You scored very strongly AGAINST options XYZ – can you tell me more about why you scored them in this way?

You scored very strongly FOR option XYZ – can you tell me more about why you scored them in this way?

OR (if there are no extremes to discuss) – your scores were broadly in the middle for XYZ - **can you tell me more about why you scored them in this way?**

INFORM DEBATE ABOUT HOW QI REPORTING SHOULD BE SUPPORTED	<ul style="list-style-type: none">• What would make it easier for you to report QI projects in surgery?• Do you think anything needs to change about how or where we write about QI in surgery?• What is the most important thing to do to ensure your work can be used by someone else?• Should we write about QI in a way that allows it to be REPLICATED EXACTLY or should we adopt a different approach?• What do you need to be able to portray a clear message to you readers?• What about the timing of when you write your report? Would it make any difference if it was retrospective or contemporaneous (as the project went along)?
EXIT QUESTION	<ul style="list-style-type: none">• What do you feel could be the biggest impact of improving reporting?• ‘Is there anything else you would like to tell me?’• Thank you again for your participation.• Do you wish to be informed of any publications relating to the research?

Appendix 8: Consent form

Participant Email Consent Form For Telephone Interviews

Version 1 August 20 2015

Once you have read the participant information sheet and had a chance to ask questions, please add your initials to each statement and send the e-mail back to elj20@le.ac.uk from your own personal or work email address.

Title of Project: Strengthening Implementation of Quality Improvement Interventions and Methods in Surgery

Name of Researcher: Emma Jones

I confirm that I have read and understand the information leaflet Version 3 dated 24-7-2015 for the above study. I have had the opportunity to consider the information, ask questions and have these answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason, and without my legal rights being affected.

I agree to the interview being digitally recorded for transcription and subsequent analysis.

I agree to anonymised quotations being used in study reports and I understand that I will not be identified in any publications resulting from the study.

I understand that data collected during the study may be looked at by the SIQINS study team at the University of Leicester Health Sciences Department, from regulatory authorities (such as ethics committees), representatives from the sponsor and / or NHS trusts accessing research data for monitoring and audit purposes.

I give my permission for the research team to show sections of my anonymised interview transcript to the authors of the QI paper we will discuss during the interview. *You can say "NO" and still take part in the interview. The interview will then be used in this study only. Please type YES or NO next to this statement as you prefer.*

I agree to take part in the above study.

Full name of participant: _____ Date: _____

Appendix 9: The references of 100 perioperative QI papers included in the systematic review *

Author	Year/ID	Country	Journal	Title
Adams R, et al.	2004;34:140–148	United States	<i>J Nurs Adm.</i>	Decreasing turnaround time between general surgery cases: A Six Sigma initiative.
Aletti GD, et al.	2009;208:614–620	United States	<i>J Am Coll Surg.</i>	Quality improvement in the surgical approach to advanced ovarian cancer: The Mayo Clinic experience.
Arbour R.	2003;34:64–71	United States	<i>J Contin Educ Nurs.</i>	A continuous quality improvement approach to improving clinical practice in the areas of sedation, analgesia, and neuromuscular blockade.
Baker RA, Newland RF.	2008;23:7–16	Australia	<i>Perfusion.</i>	Continuous quality improvement of perfusion practice: The role of electronic data collection and statistical control charts.
Berenholtz SM, et al.	2004;32:2014–2020	United States	<i>Crit Care Med.</i>	Eliminating catheter-related bloodstream infections in the intensive care unit.
Berry SA, et al.	2009;18:360–368	United States	<i>Qual Saf Health Care.</i>	ProvenCare: Quality improvement model for designing highly reliable care in cardiac surgery.
Bertolaccini L, et al.	2011;39:e128–132	Italy	<i>Eur J Cardiothorac Surg.</i>	‘Six sigma approach’—An objective strategy in digital assessment of postoperative air leaks: A prospective randomised study.
Blackmore CC, et al.	2013;39:99–105	United States	<i>Jt Comm J Qual Patient Saf.</i>	Applying Lean methods to improve quality and safety in surgical sterile instrument processing.
Bloos F, et al.	2009;103:232–237	Germany	<i>Br J Anaesth.</i>	Effects of staff training on the care of mechanically ventilated patients: A prospective cohort study.
Bosse G, et al.	2013;8(6):e85428	Africa (Tanzania)	<i>PLoS One.</i>	Immediate outcome indicators in perioperative care: A controlled intervention study on quality improvement in hospitals in Tanzania.
Brackbill ML, et al.	2010;16:329–336	United States	<i>J Manag Care Pharm.</i>	Intervention to increase the proportion of acute myocardial infarction or coronary artery bypass graft patients receiving an order for aspirin at hospital discharge.
Brattebo G, et al.	2004;13:203–205	Norway	<i>Qual Saf Health Care.</i>	Effect of a scoring system and protocol for sedation on duration of patients’ need for ventilator support in a surgical intensive care unit.
Brush JE Jr, et al.	2006;152:379–385	United States	<i>Am Heart J.</i>	Implementation of a continuous quality improvement program for percutaneous coronary intervention and cardiac surgery at a large community hospital.
Burkitt KH, et al.	2009;15:633–642	United States	<i>Am J Manag Care.</i>	Toyota production system quality improvement initiative improves perioperative antibiotic therapy.
Camp SL, et al.	2009;24:414–423	United States	<i>J Card Surg.</i>	Quality improvement program increases early tracheal extubation rate and decreases pulmonary complications and resource utilization after cardiac surgery.
Celik J.	2003;77:737–741	United States	<i>AORN J.</i>	Decreasing preoperative delays—A rapid process improvement project.
Ceppa EP, et al.	2013;15:384–391	United States	<i>HPB (Oxford).</i>	Reducing surgical site infections in hepatopancreatobiliary surgery.

*For a full list refer to the published article Jones et al (2016), How well is quality improvement described in the perioperative care literature? A systematic review, *Jt Comm J Qual Patient Saf* 42 (5) 196-206

Appendix 10: Milestones of SIQINS: Qualitative work

	Month and day milestone achieved	Jan	Jun 16	Jun 25	Jul 2	Jul	Au 24	Sept 7	Sept 18	Sept 30	Oct	Nov 6	Jan 13	Feb 24	Feb 26	Mar 9	Mar 9-16	Mar 18
	Year	2015											2016					
SIQINS Milestone	Protocol design																	
	Submit to UoL REC																	
	Submit to UoL Sponsor																	
	UoL REC approval received																	
	Recruit local collaborators																	
	UoL Sponsor in principle agreement received																	
	R&D approval process commenced																	
	Receive UoL indemnity certificate																	
	UoL agree begin recruitment at non-NHS sites																	
	1st non NHS participant recruited																	
	R&D approval at NHS sites 1,2,3																	
	NHS site 4 R&D approval																	
	1st NHS participant recruited																	
	36 Interviews completed																	
	1st LEAN author interview																	
	2nd LEAN author interview																	
	Last 3 participant interviews																	
	3rd LEAN author interview & study closed																	

Legend for Appendix 10 Milestones of SIQINS: qualitative work

	SIQINS Milestone type
	Local activity by Emma Jones in the SAPPHIRE office
	Process of obtaining approvals: Research Ethics Committee (REC), UoL sponsor, Insurance certificate
	Research and Development (R&D) approvals obtained
	Participant recruitment
	LEAN author recruitment and interviews

Appendix 11: Publications arising from this PhD

Publication 1 (first page only):

Jones et al. *Systematic Reviews* 2014, 3:98
<http://www.systematicreviewsjournal.com/content/3/1/98>



PROTOCOL

Open Access

Describing methods and interventions: a protocol for the systematic analysis of the perioperative quality improvement literature

Emma Jones^{1*}, Nicholas Lees², Graham Martin¹ and Mary Dixon-Woods¹

Abstract

Background: Quality improvement (QI) methods are widely used in surgery in an effort to improve care, often using techniques such as Plan-Do-Study-Act cycles to implement specific interventions. Explicit definition of both the QI method and quality intervention is necessary to enable the accurate replication of effective interventions in practice, facilitate cumulative learning, reduce research waste and optimise benefits to patients. This systematic review aims to assess quality of reporting of QI methods and quality interventions in perioperative care.

Methods: Studies reporting on quality interventions implemented in perioperative care settings will be identified. Searches will be conducted in the Ovid SP version of Medline, Scopus, the Cochrane Central Register of Controlled Trials, the Cochrane Effective Practice and Organisation of Care database and the related articles function of PubMed. The journal *BMJ Quality* will be searched separately. Search strategy terms will relate to (i) surgery, (ii) QI and (iii) evaluation methods. Explicit exclusion and inclusion criteria will be applied. Data from studies will be extracted using a data extraction form. The Template for Intervention Description and Replication (TIDieR) checklist will be used to evaluate quality of reporting, together with additional items aimed at assessing QI methods specifically.

Systematic review registration: PROSPERO CRD42014012845

Keywords: Quality improvement, Perioperative care, Surgery, Interventions, Reporting, Quality of care, Description

Background

Quality Improvement (QI) methods are specially designed efforts and processes aimed at generating improvements in patient care [1]. Such methods include those based on Lean, Six Sigma, Plan-Do-Study-Act (PDSA) cycles, Total Quality Management and Continuous Quality Management, audit and feedback, and many others [2]. Guidance on reporting of QI studies [3] and of intervention delivery in evaluative studies [4-7] has been published. Surgery is an especially important area for quality improvement: an estimated 234 million surgical interventions are performed every year worldwide [8], yet it remains hazardous and prone to error and

complication. An international drive to improve quality of care in surgery is now supported by initiatives such as the Centre for Global Surgery [9] and the Lancet Commission for Global Surgery [10]. Yet the quality of reporting of interventions in QI studies in surgery is unknown. This is an important problem, as it is increasingly recognised that explicit descriptions of interventions are necessary to ensure that successful interventions can be replicated in practice, to avoid research waste, to facilitate cumulative learning and to ensure that patients gain the best possible benefits from any learning from QI studies [11,12]. We seek to adopt and adapt the Template for Intervention Description and Replication (TIDieR) checklist [4] to evaluate the quality and completeness of reporting of studies of quality improvement interventions in perioperative care.

* Correspondence: e.j20@le.ac.uk

¹Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LE1 6PP, UK

Full list of author information is available at the end of the article



© 2014 Jones et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.



Human factors and ergonomics and quality improvement science: integrating approaches for safety in healthcare

Sue Hignett,¹ Emma Leanne Jones,² Duncan Miller,³ Laurie Wolf,⁴ Chetna Modi,⁵ Muhammad Waseem Shahzad,⁵ Peter Buckle,⁶ Jaydip Banerjee,⁷ Ken Catchpole⁸

For numbered affiliations see end of article.

Correspondence to
Professor Sue Hignett,
Loughborough Design School,
Loughborough University, Ashby
Road Loughborough LE11 3TU,
UK;
s.hignett@lboro.ac.uk

Received 26 September 2014
Revised 31 January 2015
Accepted 3 February 2015

INTRODUCTION

In this paper, we will address the important question of how quality improvement science (QIS) and human factors and ergonomics (HFE) can work together to produce safer solutions for healthcare. We suggest that there will be considerable advantages from an integrated approach between the two disciplines and professions which could be achieved in two phases. First, by identifying people trained in HFE and those trained in QIS who understand how to work together and second, by developing opportunities for integrated education and training. To develop this viewpoint we will:

1. Discuss and explore how QIS and HFE could be integrated by building on existing definitions, scope of practice, knowledge, skills, methods, research and expertise in each discipline.
2. Outline opportunities for a longer-term integration through training, and education for healthcare professionals.

HISTORY AND PERSPECTIVES OF QIS AND HFE

The disciplines and professions of QIS and HFE developed from similar origins in the 20th century to engage workers in the identification of problems and development of solutions.^{1–2} They diverged with QIS focussing more on process issues (eg, production quality control) and HFE focussing on wellbeing (occupational health and safety) and performance. Both have been used in healthcare for many years, with several recent papers discussing confusion about jargon in one or both disciplines.^{3–7} We will offer a simple outline of our perspectives

for each before suggesting an approach for integrated working.

We are using the term QIS to include both quality improvement and improvement science.⁸ QIS is used, defined and explained in the literature in many different ways, for example, 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge';⁹ 'better patient experience and outcomes achieved through changing provider behaviour and organisation through using a systematic change method and strategies';⁷ and as 'the combined and unceasing efforts of everyone—healthcare professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development (learning)'.¹⁰

QIS can include any change which improves quality (patient experience and/or clinical outcome), a change that uses a generic (eg, training, setting standards) or specific QIS method or approach.¹¹ Parry¹ outlined a history of QIS by drawing on Deming's System of Profound Knowledge and its influence on the development of Improvement Science. Deming's theory included systems thinking, variation (eg, statistical process control), psychology (or social sciences) and the theory of knowledge.¹² Most QIS practitioners will use a range of improvement methods and tools which may include Plan-Do-Study-Act (PDSA), Model for Improvement, Lean, Six Sigma, Total Quality Improvement and Business Process Reengineering.¹¹

To cite: Hignett S, Jones EL, Miller D, et al. *BMJ Qual Saf* Published Online First: [please include Day Month Year]
doi:10.1136/bmjqs-2014-003623



Copyright Article author (or their employer) 2015. Produced by BMJ Publishing Group Ltd under licence.

Hignett S, et al. *BMJ Qual Saf* 2015;0:1–5. doi:10.1136/bmjqs-2014-003623



Performance Improvement

How Well Is Quality Improvement Described in the Perioperative Care Literature? A Systematic Review

Emma L. Jones, MSc; Nicholas Lees, MBChB, FRCA, FFICM; Graham Martin, MA (Oxon), MSc, PhD;
Mary Dixon-Woods, BA, DipStat, MSc, DPhil

Health care is increasingly the subject of quality improvement (QI),¹ which can be understood as purposeful efforts to make changes that will lead to better patient outcomes, better system performance, and better professional development.² QI efforts often involve a *quality intervention* (specific changes to clinical or organizational systems) and a *QI technique* (a method used to support the implementation of the intervention, such as the Model for Improvement).³ Surgery is a particularly important area for QI. Fourteen record-review studies together indicated that adverse events occurred in 14.4% of 16,424 patients undergoing surgery and that potentially preventable adverse events occurred in 5.2% of them.⁴ For 3.6% of the 16,424 patients, the consequences were fatal, and for around 10.4%, severe.⁴ In the United States, adverse events in surgery account for approximately half (48%) of all adverse events in hospitals.⁵ Given that an estimated 234 million surgical interventions are performed every year worldwide,⁶ improving quality and safety of surgical care is a global priority.⁷

Perioperative care, which encompasses care delivered before, during, and after surgery,⁸ makes an important contribution to the outcomes and experiences of surgery. Systematic reviews of QI efforts in diverse surgical specialties have reported that improvements are possible across the entire perioperative journey.^{1,9–11} However, like randomized controlled trials (RCTs) in surgery,¹² there are indications that important information may be missing from reports of surgical QI studies.¹³ This is not a problem unique to surgery: Notwithstanding relevant guidance,¹⁴ reporting of QI is often weak, lacking, for example, details of implementation context, potential harm from QI activities, intervention components, and the duration of individual Plan-Do-Study-Act (PDSA) cycles.^{15,16}

One challenge in producing full and explicit accounts of interventions—QI or otherwise—has been the absence of clearly articulated expectations about what should be reported.

* National Institute for Health and Care Excellence. Keeping Patients Warm Before, During and After an Operation. Apr 1, 2008. Accessed Mar 31, 2016. <https://www.nice.org.uk/guidance/ig65/resources/keeping-patients-warm-before-during-and-after-an-operation-255053003461>.

Article-at-a-Glance

Background: Quality improvement (QI) approaches are widely used across health care, but how well they are reported in the academic literature is not clear. A systematic review was conducted to assess the completeness of reporting of QI interventions and techniques in the field of perioperative care.

Methods: Searches were conducted using Medline, Scopus, the Cochrane Central Register of Controlled Trials, the Cochrane Effective Practice and Organization of Care database, and PubMed. Two independent reviewers used the Template for Intervention Description and Replication (TIDieR) checklist, which identifies 12 features of interventions that studies should describe (for example, How: the interventions were delivered [e.g., face to face, internet]). When and how much: duration, dose, intensity), to assign scores for each included article. Articles were also scored against a small number of additional criteria relevant to QI.

Results: The search identified 16,103 abstracts from databases and 19 from other sources. Following review, full-text was obtained for 223 articles, 100 of which met the criteria for inclusion. Completeness of reporting of QI in the perioperative care literature was variable. Only one article was judged fully complete against the 11 TIDieR items used. The mean TIDieR score across the 100 included articles was 6.31 (of a maximum 11). More than a third (35%) of the articles scored 5 or lower. Particularly problematic was reporting of fidelity (absent in 74% of articles) and whether any modifications were made to the intervention (absent in 73% of articles).

Conclusions: The standard of reporting of quality interventions and QI techniques in surgery is often suboptimal, making it difficult to determine whether an intervention can be replicated and used to deliver a positive effect in another setting. This suggests a need to explore how reporting practices could be improved.

REFERENCES

1. Batalden PB, Davidoff F. What is “quality improvement” and how can it transform healthcare? *Qual Saf Heal Care*. 2007;16(1):2–3.
2. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*. 2012 Jun 16;379(9833):2252–61.
3. Nicolay CR, Purkayastha S, Greenhalgh A, Benn J, Chaturvedi S, Phillips N, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg*. 2012 Mar;99(3):324–35.
4. Powell AE, Rushmer RK, Davies HT. A systematic narrative review of quality improvement models in health care [Internet]. NHS Quality Improvement Scotland. 2009 [cited 2014 Jan 15]. Available from: <http://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=32011000041>
5. Steinman, M; Sumat, R; Shojania, K; Gonzales R. Improving Antibiotic Selection: A systematic review and quantitative analysis of quality improvement strategies. *Med Care*. 2006;44(7):617–28.
6. National Institute for Health and Clinical Excellence. Keeping patients warm before, during and after an operation (NICE guideline 65) [Internet]. 2008 [cited 2013 Dec 15]. p. 1–7. Available from: <https://www.nice.org.uk/sharedlearning/keeping-patients-warm-nice-guideline-cg65>
7. Levy SM, Phatak UR, Tsao K, Wray CJ, Millas SG, Lally KP, et al. What is the quality of reporting of studies of interventions to increase compliance with antibiotic prophylaxis? *J Am Coll Surg*. 2013 Nov;217(5):770–9.
8. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ Qual Saf*. 2014 Apr;23(4):290–8.

9. Debas H, Gosselin R, McCord C, Thind A. Surgery. In: Jamison, DT; Breman, JG; Measham, AR; Alleyne, G; Claeson, M; Evans, DB; Jha, P; Mills, A; Musgrove P, editor. Disease control priorities in developing countries. 2nd ed. New York: Oxford University Press; 2006.
10. Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, Berry WR, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet*. 2008;372(9633):139–44.
11. Sarker SK, Vincent C. Errors in surgery. *Int J Surg*. 2005;3(1):75–81.
12. Stahel P, Mauffrey C. Patient safety in surgery. London: Springer; 2014.
13. World Health Organization. WHO Guidelines for Safe Surgery 2009: safe surgery saves lives [Internet]. WHO. 2009 [cited 2014 Feb 14]. Available from: http://whqlibdoc.who.int/publications/2009/9789241598552_eng.pdf
14. Anderson O, Davis R, Hanna GB, Vincent CA. Surgical adverse events: A systematic review. *Am J Surg*. 2013;206(2):253–62.
15. Perla RJ, Provost LP, Parry GJ. Seven propositions of the science of improvement: exploring foundations. *Qual Manag Health Care*. 2013;22(3):170–86.
16. The Institute of Medicine Committee on Quality Health Care in America. Crossing the quality chasm: A new health system for the 21st century. Washington DC: The National Academy Press; 2001.
17. The Royal College of Surgeons of England. Surgical Research Report 2015/16. London; 2015.
18. The Health Foundation. Quality Improvement made simple. What every board should know about healthcare quality improvement. London; 2010.
19. Kreckler S, Morgan RD, Catchpole K, New S, Handa A, Collins G, et al. Effective prevention of thromboembolic complications in emergency surgery patients using a quality improvement approach. *BMJ Qual Saf*. 2013 Nov;22(11):916–22.

20. Reed JE, Card AJ. The problem with Plan-Do-Study-Act cycles. *BMJ Qual Saf.* 2016;25(3):147–52.
21. Moonesinghe R. Perioperative Quality Improvement Programme Introduction to Pilot Programme [Internet]. Royal College of Anaesthetists. London, WC1R 4SG; 2015 [cited 2015 Aug 25]. p. 1–12. Available from: <https://www.niaa.org.uk/NIAA-Update-July-2015>
22. Berenguer CM, Ochsner MG, Lord SA, Senkowski CK. Improving surgical site infections: using National Surgical Quality Improvement Program data to institute Surgical Care Improvement Project protocols in improving surgical outcomes. *J Am Coll Surg.* 2010;210(5):737–41, 741–3.
23. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence) : revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* [Internet]. 2015 [cited 2015 Dec 12];(September 14 Epub ahead of print):2015–004411. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/26369893>
24. Hempel S, Rubenstein L V, Shanman RM, Foy R, Golder S, Danz M, et al. Identifying quality improvement intervention publications--a comparison of electronic search strategies. *Implement Sci.* 2011;6(85):1–10.
25. Dixon-Woods, M; Bosk, C; Aveling, E; Goeschel, C; Pronovost P. Explaining Michigan: Developing an Ex Post Theory of a Quality Improvement Program. *Milbank Q.* 2011 Dec;89(2):167–205.
26. Alexander JA, Hearld LR. Review: What Can We Learn From Quality Improvement Research?: A Critical Review of Research Methods. *Med Care Res Rev.* 2009 Jun;66(3):235–71.
27. Thor J, Lundberg J, Ask J, Olsson J, Carli C, Härenstam KP, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care.* 2007 Oct;16(5):387–99.
28. Treadwell JR, Lucas S, Tsou AY. Surgical checklists: a systematic review

- of impacts and implementation. *BMJ Qual Saf.* 2014 Apr;23(4):299–318.
29. Howell A-M, Panesar SS, Burns EM, Donaldson LJ, Darzi A. Reducing the Burden of Surgical Harm: A systematic review of the interventions used to reduce adverse events in surgery. *Ann Surg.* 2014;259(4):630–41.
 30. Portela MC, Pronovost PJ, Woodcock T, Carter P, Dixon-Woods M. How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf.* 2015;0:1–12.
 31. Marjoua Y, Bozic KJ. Brief history of quality movement in US healthcare. *Curr Rev Musculoskelet Med.* 2012;5(4):265–73.
 32. Best M, Neuhauser D. Avedis Donabedian: father of quality assurance and poet. *Qual Saf Health Care.* 2004;13(6):472–3.
 33. Arnstein M. Florence Nightingale's influence on nursing. *Bull N Y Acad Med.* 1956;32(7):540–6.
 34. Nightingale F. Notes on Hospitals being two papers read before the National Association for the Promotion of Social Science, Liverpool, October 1858 with Evidence given to the royal commissioners on the state of the army. London: Forgotten Books; 1857. Reprinted 2015 p.
 35. Sheingold BH, Hahn JA. The history of healthcare quality: The first 100 years 1860-1960. *Int J Africa Nurs Sci.* Elsevier Ltd; 2014;1:18–22.
 36. Donabedian A. Evaluating the quality of medical care. *Milbank Q.* 2005;83(4):691–729 Reprinted from The Milbank Memorial Fund Q.
 37. McDonald, K.M, Sundaram, V, Bravata, D.M, Lewis, R, Lin, N, Kraft, S.A et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (volume 7 - Care Coordination). AHRQ Publ No 04(07)-0051-7 [Internet]. 2007 [cited 2014 Feb 10];7. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK44015/>
 38. Donabedian A. The effectiveness of quality assurance. *Int J Qual Heal Care.* 1996;8(4):401–7.

39. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf.* 2012 Jan;21(1):13–20.
40. Mountford J, Shojania KG. Refocusing quality measurement to best support quality improvement: local ownership of quality measurement by clinicians. *BMJ Qual Saf.* 2012 Jun;21(6):519–23.
41. Moore L, Lavoie A, Bourgeois G, Lapointe J. Donabedian's structure-process-outcome quality of care model: validation in an integrated trauma system. *J Trauma Acute Care Surg.* 2015;78(6):1168–75.
42. Ferlie E, Wood M, Fitzgerald L. Some limits to evidence-based medicine: a case study from elective orthopaedics. *Qual Health Care.* 1999;8(2):99–107.
43. Towill DR. Industrial engineering the Toyota Production System. *J Manag Hist.* 2010;16(3):327–45.
44. Price B. Frank and Lillian Gilbreth and the Manufacturing and Marketing of Motion Study , 1908-1924. *Bus Econ Hist.* 1989;18(c):88–98.
45. Best M, Neuhauser D. Walter A Shewhart, 1924, and the Hawthorne factory. *Qual Saf Health Care.* 2006;15(2):142–3.
46. Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf.* 2011 Jan;20(1):46–51.
47. Provost LP. Analytical studies: a framework for quality improvement design and analysis. *BMJ Qual Saf [Internet].* 2011 Apr [cited 2014 Mar 6];20 Suppl 1(Suppl 1):i92-6. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3066849&tool=pmcentrez&rendertype=abstract>
48. Molloy A, Martin S, Gardner T, Leatherman S. A clear road ahead - Creating a coherent quality strategy for the English NHS. The Health Foundation. London; 2016.

49. Wren, D; Greenwood R. Management Innovators: The People and Ideas That Have Shaped Modern Business. Oxford: Oxford University Press; 1998.
50. Zarbo RJ, D'Angelo R. Transforming to a Quality Culture: The Henry Ford Production System [Internet]. Pathology Patterns Reviews (online), American journal of clinical pathology. 2006 [cited 2016 Mar 10]. p. S21–9. Available from:
[https://www.henryford.com/documents/Pathology/Production System/Transforming_to_a_Quality_Culture_Zarbo_2006.pdf](https://www.henryford.com/documents/Pathology/Production%20System/Transforming_to_a_Quality_Culture_Zarbo_2006.pdf)
51. Ford H. Today and Tomorrow by Henry Ford in collaboration with Samuel Crowther. Garden City, New York: Doubleday, Page & Company; 1926.
52. Langley, G; Moen, R; Nolan, K; Nolan, T; Norman, C; Provost L. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco: Jossey-Bass; 2009.
53. The Health Foundation. Report: Improvement science. Research Scan. [Internet]. London; 2011. Available from:
[http://www.health.org.uk/media_manager/public/75/Research scan_Improvement science.pdf](http://www.health.org.uk/media_manager/public/75/Research%20scan_Improvement%20science.pdf)
54. Parry GJ. Special Series: Quality Care Symposium Perspective - A Brief History of Quality Improvement. J Oncol Pr. 2014;196–9.
55. Walshe K. Pseudoinnovation: The development and spread of healthcare quality improvement methodologies. Int J Qual Heal Care. 2009;21(3):153–9.
56. Day RW, Fielder S, Calhoun J, Kehlet H, Gottumukkala V, Aloia TA. Incomplete reporting of enhanced recovery elements and its impact on achieving quality improvement. Br J Surg. 2015;102(13):1594–602.
57. Kohn, L.T; Corrigan, J.M; Donaldson M. Institute of Medicine (US) Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System. 1st ed. The National Academy Press. Washington (DC): The National Academies Press; 1999.

58. Musgrove P, Creese A, Preker A, Baeza C, Anell A, Prentice T. World Health Report 2000: Health Systems: Improving Performance [Internet]. World Health Organization. 2000 [cited 2016 Aug 20]. Available from: http://www.who.int/whr/2000/en/whr00_en.pdf
59. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The “To Err is Human” report and the patient safety literature. *Qual Saf Heal Care*. 2006;15(3):174–8.
60. Kennedy I. Learning from Bristol: the report of the Public Inquiry into children’s heart surgery at the Bristol Royal Infirmary 1984-1995. London: HMSO; 2001.
61. Martin GP, Finn R. Patients as team members: opportunities, challenges and paradoxes of including patients in multi-professional healthcare teams. *Sociol Health Illn*. 2011 Nov;33(7):1050–65.
62. Walshe K, Offen N. A very public failure: lessons for quality improvement in healthcare organisations from the Bristol Royal Infirmary. *Qual Heal care*. 2001;10(4):250–6.
63. Coulter AEJ. Patient-focused interventions: A review of the evidence [Internet]. The Health Foundation. 2006 [cited 2014 Jan 18]. p. 1–277. Available from: <http://www.health.org.uk/publications/patient-focused-interventions/>
64. Donaldson L. Department of Health: An organisation with a memory - Report of an expert group on learning from adverse events in the NHS. London: HMSO, crown copyright; 2000.
65. Anthony R, Butterworth T. Annual Report and Accounts of the NHS Institute for Innovation and Improvement 2012-13 [Internet]. London: HMSO, crown copyright; 2013. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/246670/0301.pdf
66. Department of Health. High Quality Care For All NHS Next Stage Review Final Report. Darzi A, editor. London: Crown Copyright; 2008.

67. Maybin J; Thorlby R. High Quality Care For All: Briefing on NHS next stage review final report. The Kings Fund. 2008. p. 1–8.
68. Wood, H; Fontain, P; Harvey, D; Catford, G; Hills, A; Holland B et al. Investigation into mid staffordshire NHS foundation trust. Healthcare Commission. London, March; 2009.
69. Francis R. Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust January 2005 – March 2009. London: Crown copyright, Stationery Office; 2010.
70. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. London: Crown copyright, The Stationary office; 2013.
71. Berwick D. A promise to learn – a commitment to act: Improving the Safety of Patients in England [Internet]. National Advisory Group on the Safety of Patients in England. 2013 [cited 2015 Oct 15]. Available from: <https://www.gov.uk/government/publications/berwick-review-into-patient-safety>
72. Keogh B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report [Internet]. London: NHS England; 2013 [cited 2015 Jan 10]. Available from: <http://www.nhs.uk/nhsengland/bruce-keogh-review/documents/outcomes/keogh-review-final-report.pdf>
73. The Health Foundation. Evidence scan: Quality Improvement training for healthcare professionals. London; 2012.
74. Faculty of Pharmaceutical Medicine of the Royal Colleges of Physicians of the United Kingdom. Supporting information for appraisal and revalidation: guidance for pharmaceutical medicine [Internet]. London; 2016 [cited 2016 Oct 20]. p. 1–8. Available from: <https://www.fpm.org.uk/revalidationcpd/revalidation/SupportingInformation>
75. NHS England. NHS Standard Contract 2016/17 Service Conditions (Full Length) [Internet]. London: Crown copyright; 2016. Available from: <https://www.england.nhs.uk/nhs-standard-contract/16-17/>

76. Dixon-Woods M, Baker R, Charles K, Dawson J, Jerzembek G, Martin G, et al. Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study. *BMJ Qual Saf.* 2014 Feb;23(2):106–15.
77. Robertson R, Gregory S, Jabbal J. Background paper: The social care and health systems of nine countries [Internet]. The Kings Fund - Commission on the future of health and social care in England. London; 2014 [cited 2016 Jul 28]. Available from: <http://www.kingsfund.org.uk/sites/files/kf/media/commission-background-paper-social-care-health-system-other-countries.pdf>
78. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman D, et al. A pragmatic – explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Can Med Assoc J.* 2009;180(10):E47–57.
79. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff.* 2005;24(1):138–50.
80. Grol R, Baker R, Moss F. Quality improvement research: understanding the science of change in health care. *Qual Saf Health Care* [Internet]. 2002 Jun;11(2):110–1. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1743590&tool=pmcentrez&rendertype=abstract>
81. Dixon-Woods M, Martin GP. Does quality improvement improve quality? *Futur Hosp J.* 2016;3(3):191–4.
82. Sobo EJ, Bowman C, Gifford AL. Behind the scenes in health care improvement: the complex structures and emergent strategies of Implementation Science. *Soc Sci Med* [Internet]. 2008 Nov [cited 2014 Mar 6];67(10):1530–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18760519>
83. Øvretveit J. Producing useful research about quality improvement. *Int J Health Care Qual Assur.* 2002;15:294–302.

84. Margolis P, Provost LP, Schoettker PJ, Britto MT. Quality improvement, clinical research, and quality improvement research--opportunities for integration. *Pediatr Clin North Am*. 2009 Aug;56(4):831–41.
85. Rycroft-Malone J, Seers K, Titchen A, Harvey G, Kitson A, McCormack B. What counts as evidence in evidence-based practice? *J Adv Nurs* [Internet]. 2004 Jul;47(1):81–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21733101>
86. Ting HH, Shojania KG, Montori VM, Bradley EH. Quality improvement: science and action. *Circulation*. 2009;119(14):1962–74.
87. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337(October):979–83.
88. Davies L, Batalden P, Davidoff F, Stevens D, Ogrinc G. The SQUIRE Guidelines: an evaluation from the field, 5 years post release. *BMJ Qual Saf*. 2015;24(12):769–75.
89. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147(04 May-10 May):1–11.
90. Leviton L. Reconciling complexity and classification in quality improvement research. *BMJ Qual Saf* [Internet]. 2011 Apr [cited 2014 Mar 6];20 Suppl 1(Suppl 1):i28-9. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Reconciling+complexity+and+classification+in+quality+improvement+research>.
91. Blaike N. *Designing Social Research: the logic of anticipation*. 2nd ed. Cambridge: Polity Press; 2009.
92. Aveling E, Martin G, Armstrong N, Banerjee J, Dixon-Woods M. Quality improvement through clinical communities: eight lessons for practice. *J Health Organ Manag*. 2012;26(2):158–74.
93. Cevalco M, Ashley SW. Quality measurement and improvement in

- general surgery. *Perm J* [Internet]. 2011;15(4):48–53. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3267561&tool=pmcentrez&rendertype=abstract>
94. Pickering SP, Robertson ER, Griffin D, Hadi M, Morgan LJ, Catchpole KC, et al. Compliance and use of the World Health Organization checklist in U.K. operating theatres. *Br J Surg*. 2013 Nov;100(12):1664–70.
 95. Ergina, Patrick L; Cook, JA; Blazeby, JM; Boutron, I; Clavien, PA; Reeves, BC; Seiler C. Challenges in evaluating surgical innovation. *Lancet*. 2009;374(9695):1097–104.
 96. Yu A, Flott K, Chainani N, Fontana G, Darzi A. Patient Safety 2030: NIHR Patient Safety Translational Research Centre at Imperial College London and Imperial College Healthcare NHS Trust [Internet]. 2016 [cited 2016 Oct 13]. Available from: <http://www.imperial.ac.uk/media/imperial-college/institute-of-global-health-innovation/centre-for-health-policy/Patient-Safety-2030-Report-VFinal.pdf>
 97. Dankelman, J; Grimbergen, C.A; Stassen HG. Engineering for patient safety Issues in Minimally Invasive Procedures. London: Lawrence Erlbaum Associates; 2005.
 98. Maragh-Bass AC, Appelson JR, Changoor NR, Davis WA, Haider AH, Morris MA. Prioritizing qualitative research in surgery: A synthesis and analysis of publication trends. *Surgery* [Internet]. 2016 [cited 2016 Oct 2];Aug 4(E pub ahead of print):1–9. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Prioritizing+qualitative+research+in+surgery:+A+synthesis+and+analysis+of+publication+trends>
 99. Barlow T, Griffin D, Barlow D, Realpe A. Patients' decision making in total knee arthroplasty: a systematic review of qualitative research. *Bone Joint Res*. 2015;4(10):163–9.
 100. Sibbern T, Bull Sellevold V, Steindal SA, Dale C, Watt-Watson J, Dihle A. Patients' Experiences of Enhanced Recovery after Surgery: A systematic review of qualitative studies. *J Clin Nurs* [Internet]. 2016 [cited 2016 Oct

1];June 27(Epub ahead of print). Available from:
<https://www.ncbi.nlm.nih.gov/pubmed/27345939>

101. Donovan J, Mills N, Smith M, Brindle L, Jacoby A, Peters T, et al. Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *BMJ*. 2002;325(7367):766–70.
102. Kringos DS, Sunol R, Wagner C, Mannion R, Michel P, Klazinga NS, et al. The influence of context on the effectiveness of hospital quality improvement strategies: a review of systematic reviews. *BMC Health Serv Res* [Internet]. 2015 [cited 2015 Aug 25];15(277). Available from: <http://www.biomedcentral.com/1472-6963/15/277>
103. Annandale C. *The New Gresham Dictionary of the English Language*. London: The Gresham Publishing Company Ltd; 1926.
104. Winker, MA; Ferris, LE; WAME, ethics and policy committee; WAME, Board; Aggarwal, R; Barbour V et al. Promoting Global Health: The World Association of Medical Editors position on editors' responsibility. *Int J Occup Env Med*. 2015;6(3):125–7.
105. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [Internet]. The World Medical Association. 2013 [cited 2015 Jan 10]. p. 1–8. Available from: <http://www.wma.net/en/30publications/10policies/b3/index.html>
106. Wright D, Powell J. Adding value in research 4: ensuring full publication of NIHR funded studies [Internet]. 2015 [cited 2016 Aug 12]. p. 1–4. Available from: <http://www.nihr.ac.uk/about/adding-value-in-research.htm>
107. UK Policy Framework for Health and Social Care Research (issued for public consultation) [Internet]. Health Research Agency. London; 2015 [cited 2016 Jul 22]. p. 1–23. Available from: <http://www.hra.nhs.uk/about-the-hra/consultations-calls/uk-policy-framework-health-social-care-research-consultation-active/>
108. Medical Research Council working group. MRC ethics series. Good

- research practice: Principles and guidelines [Internet]. 2012 [cited 2016 Jun 14]. p. 1–27. Available from: <https://www.mrc.ac.uk/research/policies-and-resources-for-researchers/good-research-practice/>
109. European Medicines Agency policy/0070 on publication of clinical data for medicinal products for human use [Internet]. London E14 5EU; 2014 [cited 2015 Nov 1]. p. 1–22. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf
 110. Hale, C; Arrowsmith, J; Bayes, S; Bodmer, H; Garman, A; Hubbard, P; Perrin, N; Sooben P. The concordat to support research integrity [Internet]. Universities UK. 2012 [cited 2014 Oct 3]. Available from: <http://www.universitiesuk.ac.uk/highereducation/Documents/2012/TheConcordatToSupportResearchIntegrity.pdf>
 111. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348(June):1–7.
 112. Ioannidis JPA. How to Make More Published Research True. *PLoS Med* [Internet]. Public Library of Science; 2014 Oct 21 [cited 2014 Oct 21];11(10):e1001747. Available from: <http://dx.doi.org/10.1371/journal.pmed.1001747>
 113. Chan A-W, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, et al. Increasing value and reducing waste: Addressing inaccessible research. *Lancet*. 2014;383(9913):257–66.
 114. National Institute for Health Research Clinical Research Network: Introduction to Good Clinical Practice: A practical guide to quality standards in clinical research consolidated version 1.0 [Internet]. 2011 [cited 2014 Jan 6]. Available from: <http://ichgcp.net/>
 115. NHS England. Transforming urgent and emergency care services in England: Urgent and Emergency Care Review End of Phase 1 Report [Internet]. Leeds, United Kingdom; 2013 [cited 2014 Jan 10]. Available from: <http://www.nhs.uk/NHSEngland/keogh->

116. Dijkers MP. Reporting on interventions: issues and guidelines for rehabilitation researchers. *Arch Phys Med Rehabil*. 2015;96(6):1170–80.
117. York Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. [Internet]. 2009 p. 1–281. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf
118. Rothstein H, Sutton A, Borenstein M. Publication bias in Meta-Analysis. In: Rothstein H, Sutton A, Borenstein M, editors. *Publication Bias in Meta-Analysis Prevention, Assessment and Adjustments*. West Sussex (UK): John Wiley & Sons; 2005. p. 1–9.
119. Shrier I, Boivin J-F, Platt RW, Steele RJ, Brophy JM, Carnevale F, et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* [Internet]. 2008 May [cited 2015 Jan 14];8(19). Available from: <https://www.ncbi.nlm.nih.gov/pubmed/18495019>
120. Lokker C, McKibbin KA, Colquhoun H, Hempel S. A scoping review of classification schemes of interventions to promote and integrate evidence into practice in healthcare. *Implement Sci*. 2015;10(27):1–12.
121. Leistikow IP, Kalkman CJ, de Bruijn H. Why patient safety is such a tough nut to crack. *BMJ*. 2011;342:d3447.
122. West M, Baker R, Dawson J, Woods M, Lilford R, Martin G, et al. Quality and safety in the NHS: evaluating progress, problems and promise. Lancaster, Univ [Internet]. 2013;1–157. Available from: <http://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/lums/cphr/quality-safety-nhs-e.pdf>
123. Conn VS. Unpacking the black box: Countering the problem of inadequate intervention descriptions in research papers. *West J Nurs Res*. 2012;34(4):427–33.
124. Ivers NM, Tricco AC, Taljaard M, Halperin I, Turner L, Moher D, et al.

- Quality improvement needed in quality improvement randomised trials: systematic review of interventions to improve care in diabetes. *BMJ Open* [Internet]. 2013 Jan [cited 2014 Mar 6];3(4). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641459&tool=pmcentrez&rendertype=abstract>
125. Koetsier A, van der Veer SN, Jager KJ, Peek N, de Keizer NF. Control charts in healthcare quality improvement. A systematic review on adherence to methodological criteria. *Methods Inf Med*. 2012 Jan;51(3):189–98.
 126. Shojania, KG; Ranji, SR; Mcdonald KM, Grimshaw JM, Rushakoff RJ, Owens DK. Effects of Quality Improvement Strategies for Type 2 Diabetes on Glycemic Control: A meta-Regression Analysis. 2006;296(4):427–40.
 127. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* [Internet]. 2007 Jan [cited 2015 Jan 5];2(12):e1350. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2131785&tool=pmcentrez&rendertype=abstract>
 128. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* [Internet]. 2007 Feb [cited 2014 Jul 16];7(10):1–7. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/17302989>
 129. Grimshaw J, McAuley LM, Bero LA, Grilli R, Oxman AD, Ramsay C, et al. Systematic reviews of the effectiveness of quality improvement strategies and programmes. *Qual Saf Health Care*. 2003;12(4):298–303.
 130. Altman D, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. [Internet]. *James Lind Library Bulletin: Commentaries on the history of treatment evaluation*. 2015 [cited 2016 Aug 3]. Available from:

<http://www.jameslindlibrary.org/articles/a-history-of-the-evolution-of-guidelines-for-reporting-medical-research-the-long-road-to-the-equator-network/>

131. Altman DG. The scandal of poor medical research. *Bmj*. 1994;308(6924):283–4.
132. Andrew E. A Proposal for Structured Reporting of Randomized Controlled Trials. *JAMA*. 1994;272(24):1926–31.
133. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. *JAMA*. 1996;276(8):637–9.
134. Stevens A, Shamseer L, Weinstein E, Thielman J, Altman DG, Hirst A, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ open access [Internet]*. 2014;3804(June):1–29. Available from: <http://www.bmj.com/content/348/bmj.g3804.long>
135. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *PLOS Med [Internet]*. 2016 [cited 2016 Sep 30];13(6):e1002056. Available from: <http://dx.plos.org/10.1371/journal.pmed.1002056>
136. Brouwers MC, Kerkvliet K, Spithoff K, Consortium ANS. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 2016;352(March):i1152.
137. Agarwal S, Lefevre AE, Lee J, L'engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ*. 2016;352(i1174):1–10.
138. Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *BMJ*. 2013;347:f3755.

139. Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? *Physiotherapy*. The Chartered Society of Physiotherapy; 2016;102(2):121–6.
140. Hoffmann TC, Walker MF, Langhorne P, Eames S, Thomas E, Glasziou P. What's in a name? The challenge of describing interventions in systematic reviews: analysis of a random sample of reviews of non-pharmacological stroke interventions. *BMJ Open* [Internet]. 2015 [cited 2016 Feb 3];5(11):e009051. Available from: <http://bmjopen.bmj.com/cgi/doi/10.1136/bmjopen-2015-009051>
141. Schulz K, Altman DG, Moher D, CONSORT Group. CONSORT (Consolidated Standards of Reporting Trials) checklist of information to include when reporting a randomised trial [Internet]. 2010 [cited 2014 Feb 10]. Available from: <http://www.consort-statement.org/consort-2010>
142. Elm E Von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting. *Ann Intern Med*. 2007;147(8):573–7.
143. Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D. The CARE guidelines: Consensus-based clinical case report guideline development. *J Clin Epidemiol*. 2014;67(1):46–51.
144. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336–41.
145. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32- item checklist for interviews and focus groups. *Int J Qual Heal Care*. 2007;19(6):349–57.
146. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* [Internet]. 2012 Nov [cited 2014 Sep

15];12(181):1–8. Available from:
<https://www.ncbi.nlm.nih.gov/pubmed/23185978>

147. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Paul P, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277(3):826–32.
148. Chan A, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
149. Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care*. 2008 Oct;17 Suppl 1(Suppl I):i13-32.
150. Hoffmann TC, Glasziou PP, Barbour V, Macdonald H. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687(March):1–13.
151. Schulz, R; Czaja, S; McKay, J; Ory, M; Belle S. NIH Public Access Intervention Taxonomy (ITAX): Describing Essential Features of Interventions (HMC). *Am J Heal Behav*. 2010;34(6):811–21.
152. McAuley, L; Ramsay C. Cochrane Effective Practice and Organisation of Care Review Group Data Collection Checklist. Cochrane Effective Practice and Organisation of Care (EPOC), Review Group. Ontario: Institute of Population Health. University of Ottawa; 2002.
153. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. *Ann Intern Med*. 2008;4(148):W60-6.
154. Boutron I, Moher D, Altman D, Kenneth; S, Ravaud P. Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration. *Ann Intern Med*. 2008;148(4):295–309.

155. Cooper H (On behalf of the communications board working group on journal article reporting standards). Reporting standards for research in psychology: why do we need them? What might they be? *Am Psychol*. 2008;63(9):839–51.
156. Pinnock H, Epiphaniou E, Sheikh A, Griffiths C, Eldridge S, Craig P, et al. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. *Implement Sci*. 2015;10(42).
157. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* [Internet]. 2010 Feb [cited 2014 Oct 5];7(2):e1000217. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20169112>
158. Larsen PO, Von Ins M. The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*. 2010;84(3):575–603.
159. Van Bokhoven M A, Kok G, Van der Weijden T. Designing a quality improvement intervention: a systematic approach. *Qual Saf Health Care*. 2003;12(3):215–20.
160. Blencowe NS, Brown JM, Cook JA, Metcalfe C, Morton DG, Nicholl J, et al. Interventions in randomised controlled trials in surgery: issues to consider during trial design. *Trials* [Internet]. 2015 [cited 2016 Jan 17];16(392). Available from: <http://www.trialsjournal.com/content/16/1/392>
161. Cook JA, Mcculloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ*. 2013;346:f2820.
162. Hempel S, Shekelle PG, Liu JL, Sherwood Danz M, Foy R, Lim Y-W, et al. Development of the Quality Improvement Minimum Quality Criteria Set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications. *BMJ Qual Saf*. 2015;12:796–804.

163. Deeks JJ, Dinnes J, D'Amico R, Sowden a J, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess (Rockv)*. 2003;7(27):iii–x, 1-173.
164. Hartling L, Ospina M, Liang Y, Dryden D, Hooton N, Krebs Seida J, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339:1–6.
165. Howell V, Schwartz AE, O'Leary JD, McDonnell C. The effect of the SQUIRE (Standards of QUality Improvement Reporting Excellence) guidelines on reporting standards in the quality improvement literature: a before-and-after study. *BMJ Qual Saf*. 2015;24(6):400–6.
166. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006;185(5):263–7.
167. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005;365(9465):1159–62.
168. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev [Internet]*. 2012 [cited 2014 Jul 14];1(60). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3564748&tool=pmcentrez&rendertype=abstract>
169. Tansella M, Thornicroft G. Implementation science: understanding the translation of evidence into practice. *Br J psychiatry*. 2009 Oct;195(4):283–5.
170. Buus N, Agdal R. Can the use of reporting guidelines in peer-review damage the quality and contribution of qualitative health care research? *Int J Nurs Stud*. 2013;50(10):1289–91.
171. Smyth RM, Jacoby A, Altman DG, Gamble C, Williamson PR. The natural history of conducting and reporting clinical trials: interviews with trialists. *Trials [Internet]*. 2015;16(1):1–12. Available from:

172. Wells M, Williams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. *Trials* [Internet]. 2012 Jan [cited 2014 Sep 18];13(1):95. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22742939>
173. Popham K, Calo WA, Carpentier MY, Chen NE, Kamrudin SA, Le Y-CL, et al. Reporting Guidelines: Optimal Use in Preventive Medicine and Public Health. *Am J Prev Med*. 2012;43(4):e31–42.
174. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349(November):g6501.
175. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340(Feb):c365.
176. Scherer RW, Ugarte-Gil C, Schmucker C, Meerpohl JJ. Authors report lack of time as main reason for unpublished research presented at biomedical conferences: A systematic review. *J Clin Epidemiol*. 2015;68(7):803–10.
177. Gattrell WT, Hopewell S, Young K, Farrow P, White R, Wager E, et al. Professional medical writing support and the quality of randomised controlled trial reporting: a cross-sectional study. *BMJ Open* [Internet]. 2016 [cited 2016 Sep 30];6(2):e010329. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4762118&tool=pmcentrez&rendertype=abstract>
178. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86–9.
179. Stevencon, A; Waite M. *The Concise Oxford English Dictionary*. 12th ed. Oxford OX2 6DP: Oxford University Press; 2011.

180. McCulloch P, Catchpole K. A three-dimensional model of error and safety in surgical health care microsystems. Rationale, development and initial testing. *BMC Surg* [Internet]. 2011 [cited 2015 Jan 21];11(23). Available from: <http://www.biomedcentral.com/1471-2482/11/23>
181. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication Guidelines for Improvement Studies in Health Care : Evolution of the SQUIRE Project. *Ann Intern Med*. 2005;149:670–6.
182. Davidoff F. Improvement Interventions Are Social Treatments, Not Pills. *Ann Intern Med*. 2014;161(7):526–7.
183. Dy SM, Taylor SL, Carr LH, Foy R, Pronovost PJ, Øvretveit J, et al. A framework for classifying patient safety practices: results from an expert consensus process. *BMJ Qual Saf*. 2011;20(7):618–24.
184. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ*. 2007 Mar 3;334(7591):455–9.
185. Colquhoun H, Leeman J, Michie S, Lokker C, Bragge P, Hempel S, et al. Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *Implement Sci*. 2014;9(51).
186. McDonald, Kathryn; Schultz, Ellen; Chang C. Methods Research Report: Through the Quality Kaleidoscope: Reflections on the Science and Practice of Improving Health Care Quality Methods Research Report Through the Quality Kaleidoscope: Reflections on the Science and Practice of Improving Health Care. Agency Healthc Res Qual [Internet]. Rockville, MD, US; 2013 [cited 2014 Mar 14];Feb:AHRQ Publication No. 13-EHC041-EF. Available from: www.EffectiveHealthCare.ahrq.gov/reports/final.cfm.
187. Colquhoun H, Leeman J, Michie S, Lokker C, Bragge P, Hempel S, et al. Additional File 1 . The mapping of : Behaviour Change Wheel, EPOC categories, Leeman taxonomy, and behaviour change techniques. In:

Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practice. *Implement Sci.* 2014;9(1).

188. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci.* 2011;6(42).
189. Shepperd, S; Lewin S, on behalf of the Effective Practice and Organisation of Care Group. The Effective Practice and Organisation of Care (EPOC) taxonomy of health systems interventions. EPOC Resources for review authors. [Internet]. Oslo: Norwegian Knowledge Centre for the Health Services; 2015 [cited 2016 Jan 14]. p. 1–11. Available from: <https://epoc.cochrane.org/epoc-taxonomy>
190. Gianola S, Castellini G, Agostini M, Bolotta R, Corbetta D, Frigerio P, et al. Reporting of Rehabilitation Intervention for Low Back Pain in Randomized Controlled Trials. *Spine*,. 2016;41(5):412–8.
191. Kaplan, H; Brady, M; Hooper, D; Linham, M; Froehle C; Margolis P. The Influence of Context on Quality Improvement Success in Health Care: A Systematic Review of the Literature. *Milbank Q.* 2010;88(4):500–559.
192. Dixon-Woods M, Leslie M, Tarrant C, Bion J. Explaining Matching Michigan: an ethnographic study of a patient safety program. *Implement Sci. Implementation Science*; 2013;8(70).
193. Estabrooks C a, Squires JE, Cummings GG, Birdsell JM, Norton PG. Development and assessment of the Alberta Context Tool. *BMC Health Serv Res* [Internet]. 2009 [cited 2014 Sep 25];9(234). Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20003531>
194. Brennan SE, Bosch M, Buchan H, Green SE. Measuring organizational and individual factors thought to influence the success of quality improvement in primary care: a systematic review of instruments. *Implement Sci.* 2012;7(121).
195. Catchpole K. Spreading human factors expertise in healthcare: untangling

- the knots in people and systems. *BMJ Qual Saf*. 2013 Oct;22(10):793–7.
196. Colquhoun H, Michie S, Sales A, Ivers N, Grimshaw JM, Carroll K, et al. Reporting and design elements of audit and feedback interventions: a secondary review. *BMJ Qual Saf* [Internet]. 2016 [cited 2016 Aug 22];2015–00500(January Epub ahead of print). Available from: <https://www.ncbi.nlm.nih.gov/pubmed/26811541>
 197. Hoffmann T, Eructi C, Thorning S, Glasziou P. The scatter of research: cross sectional comparison of randomised trials and systematic reviews across specialties. *BMJ* [Internet]. 2012;344:e3223. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3354729&tool=pmcentrez&rendertype=abstract>
 198. Rubenstein L V, Hempel S, Farmer MM, Asch SM, Yano EM, Dougherty D, et al. Finding order in heterogeneity: types of quality-improvement intervention publications. *Qual Saf Health Care*. 2008;17(6):403–8.
 199. Øvretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf*. 2011 Apr;20(Suppl 1):i18-23.
 200. Ivers N, Jamtvedt G, Flottorp S, Young JM, French SD, O'Brien MA, et al. Audit and feedback: effects on professional practice and healthcare outcomes (Review). *Cochrane Coll*. 2012;(6):1–227.
 201. Miettinen O. Confounding and effect-modification. 1974 (re-printed). *Am J Epidemiol*. 1995;141(12):350–3.
 202. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ*. 2006;333(August):346–9.
 203. Hanson BP, Bhandari M, Audige L, Helfet D. The need for education in evidence-based orthopedics: an international survey of AO course participants. *Acta Orthop Scand*. 2004;75(3):328–32.
 204. World Health Organization. Patient safety curriculum guide. Multi-

Professional Edition. WHO Library Cataloguing. Malta; 2011.

205. Wong BM, Etchells EE, Kuper A, Levinson W, Shojania KG. Teaching quality improvement and patient safety to trainees: a systematic review. *Acad Med*. 2010;85(9):1425–39.
206. Boonyasai R, Windish D, Chakraborti C, Feldman L, Rubin H, Bass E. Effectiveness of Teaching Quality Improvement to Clinicians A Systematic Review. *JAMA*. 2007;298(9):1023–37.
207. Windish DM, Reed D a, Boonyasai RT, Chakraborti C, Bass EB. Methodological rigor of quality improvement curricula for physician trainees: a systematic review and recommendations for change. *Acad Med*. 2009;84(12):1677–92.
208. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383(9912):156–65.
209. Sheldon TA. It ain't what you do but the way that you do it. *J Health Serv Res Policy*. 2001;6(3):3–5.
210. Taylor SL, Dy S, Foy R, Hempel S, McDonald KM, Øvretveit J, et al. What context features might be important determinants of the effectiveness of patient safety practice interventions? *BMJ Qual Saf*. 2011;20(7):611–7.
211. Dickersin K. Publication Bias: Recognizing the Problem, Understanding Its Origins and Scope, and Preventing Harm. In: Rothstein H, Sutton A, Borenstein M, editors. *Publication Bias in Meta-Analysis—Prevention, Assessment and Adjustments*. West Sussex (UK): John Wiley & Sons; 2005. p. 11–35.
212. Chalmers I. Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. *Ann Am Acad Pol Soc Sci*. 2003;589(1):22–40.
213. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* [Internet].

2007 Mar 27 [cited 2014 Oct 15];4(3):e78. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1831728&tool=pmcentrez&rendertype=abstract>

214. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg.* 2011;39(2):91–2.
215. Wallace J, Byrne C, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. *Int J Evid Based Healthc.* 2012;10(4):338–46.
216. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature.* 2012;483(7391):531–3.
217. Vaux DL, Fidler F, Cumming G. Replicates and repeats—what is the difference and is it significant? *EMBO Rep.* 2012;13(4):291–6.
218. Øvretveit J, Leviton L, Parry G. Increasing the generalisability of improvement research with an improvement replication programme. *BMJ Qual Saf.* 2011;20(Suppl 1):i87-91.
219. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ.* 2000;321(September):694–6.
220. Khan S, Wilson T, Ahmed J, Owais A, MacFie J. Quality of life and patient satisfaction with enhanced recovery protocols. *Color Dis.* 2010 Dec;12(12):1175–82.
221. Jones E, Wainwright T, Foster J, Smith J, Middleton R, Francis N. A systematic review of patient reported outcomes and patient experience in enhanced recovery after orthopaedic surgery. *Ann R Coll Surg Engl.* 2014;96(2):89–94.
222. Hall TC, Dennison a R, Bilku DK, Metcalfe MS, Garcea G. Enhanced recovery programmes in hepatobiliary and pancreatic surgery: a systematic review. *Ann R Coll Surg Engl.* 2012;94(5):318–26.

223. Lv D, Wang X, Shi G. Perioperative enhanced recovery programmes for gynaecological cancer patients (Review). Cochrane database Syst Rev. 2010;(6).
224. Spanjersberg W, Reurings J, Keus F, Van Laarhoven CJHM. Fast track surgery versus conventional recovery strategies for colorectal surgery (Review). Cochrane database Syst Rev. 2011;(2).
225. Smart NJ, White P, Allison a S, Ockrim JB, Kennedy RH, Francis NK. Deviation and failure of enhanced recovery after surgery following laparoscopic colorectal surgery: early prediction model. Colorectal Dis. 2012;14(10):e727-34.
226. Boulind CE, Yeo M, Burkill C, Witt A, James E, Ewings P, et al. Factors predicting deviation from an enhanced recovery programme and delayed discharge after laparoscopic colorectal surgery. Colorectal Dis. 2012;14(3):e103-10.
227. Ioannidis J, Munafo M, Fusar-Poli P, Nosek B, David S. Publication and other reporting biases in cognitive sciences: detection, prevalence and prevention. Trends Cogn Sci. 2014;18(5):235–41.
228. Bion J, Richardson A, Hibbert P, Beer J, Abrusci T, McCutcheon M, et al. “Matching Michigan”: a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. BMJ Qual Saf. 2013;22(2):110–23.
229. Pronovost PJ, Goeschel C a, Colantuoni E, Watson S, Lubomski LH, Berenholtz SM, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. BMJ. 2010;340(c309).
230. DeVito Dabbs A, Song M-K, Hawkins R, Aubrecht J, Kovach K, Terhorst L, et al. An Intervention Fidelity Framework for Technology-Based Behavioral Interventions. Nurs Res. 2011;60(5):340–7.
231. Rycroft-Malone J, Seers K, Crichton N, Chandler J, Hawkes C., Allen C, et al. A pragmatic cluster randomised trial evaluating three

- implementation interventions. *Implement Sci.* 2012 Jan;7(80).
232. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *J Public Heal Dent.* 2011;71(s1):S52–63.
233. Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ.* 2004;328(7455):1561–3.
234. Waife S. Problems of publication. In: Waife SO, Shapiro AP, eds *The Clinical Evaluation of New Drugs*. New York: Hoeber-Harper; 1959. p. 213–216.
235. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet.* 2014;383(9912):176–85.
236. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* 2014;383(9913):267–76.
237. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* 2014;383(9912):166–75.
238. Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A. Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ.* 2014;349(December):g6870.
239. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *Milbank Q.* 2004;82(4):581–629.
240. Treasure T, Monson K, Fiorentino F, Russell C. Operating to remove recurrent colorectal cancer: have we got it right? *BMJ.* 2014;2085(May):1–5.
241. Editorial. Consequences of a competitive research culture. [Internet]. *Lancet.* 2014 [cited 2015 Jan 13]. p. 2084. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/25497186>

242. Montgomery J, Nurse P, Thomas J, Tildesley D, Tooke J. The findings of a series of engagement activities exploring: The culture of scientific research in the UK [Internet]. Nuffield Council on Bioethics. Oxford; 2014 [cited 2015 Jan 8]. Available from: <http://nuffieldbioethics.org/project/research-culture/>
243. Taylor BJ, Francis K. Qualitative research in the health sciences: methodologies, methods, and process. Oxon, OX14 4RN: Routledge; 2013. 297 p.
244. Mohr J, Batalden P, Barach P. Integrating patient safety into the clinical microsystem. *Qual Saf Health Care*. 2004;13 Suppl 2(Suppl II):ii34-i38.
245. Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Prancan K, Baruth K, Wise S, Lee J, editors. Belmont, CA: Wadsworth, Cengage Learning; 2002.
246. Schlosser RW. Appraising the quality of systematic reviews. FOCUS A Publication of the National Center for the Dissemination of Disability Research (NCDDR). 2007.
247. Higgins J, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. [Internet]. 2011 [cited 2014 Jan 11]. Available from: <http://training.cochrane.org/handbook>
248. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP a, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009 Jul;6(7):e1000100.
249. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One*. 2013 Dec;8(12):e83138.
250. Popay J, Rogers A WG. Rationale and standards for the systematic

- review of qualitative literature in health services research. *Qual Heal Res.* 1998;8:341–51.
251. Jones E, Lees N, Martin G, Dixon-Woods M. Describing methods and interventions: a protocol for the systematic analysis of the perioperative quality improvement literature. *BMC Syst Rev* [Internet]. 2014 [cited 2014 Sep 30];3(98). Available from:
<https://www.ncbi.nlm.nih.gov/pubmed/?term=Describing+methods+and+interventions:+a+protocol+for+the+systematic+analysis+of+the+perioperative>
 252. Shepperd S, Lewin S, Straus S, Clarke M, Eccles MP, Fitzpatrick R, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Med* [Internet]. 2009 Aug [cited 2014 Jul 23];6(8):e1000086. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2717209&tool=pmcentrez&rendertype=abstract>
 253. K, Shojania; McDonald, K; Watcher, R; Owens D. Closing the Quality Gap : A Critical Analysis of Quality Improvement Strategies - (Vol 1— Series Overview and Methodology -Technical Review 9). Agency for Healthcare Research and Quality. Rockville, MD, US; 2004.
 254. Ellis, P; Robinson, P; Ciliska, D; Armour, T; Raina, P; Brouwers M. Diffusion and dissemination of evidence based cancer control interventions: summary. Evidence Report/Technology Assessment No. 79. Agency for Healthcare Research and Quality. Rockville, MD, US; 2003.
 255. Grimshaw J, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions . *Med Care.* 2001;39(8 Suppl 2):112–45.
 256. Stone EG, Morton SC, Hulscher ME, Maglione MA, Roth EA, Grimshaw JM, et al. Article Interventions That Increase Use of Adult Interventions that increase use of adult Immunization and Cancer Screening Services : A Meta-Analysis. *Ann Intern Med.* 2002;136(9):641–51.

257. Stafford R. Alternative strategies for controlling rising cesarean section rates. *JAMA*. 1990;263(5):683–7.
258. Weingarten SR, Henning JM, Badamgarav E, Knight K, Hasselblad V, Anacleto G, et al. Interventions used in disease management programmes for patients with chronic illness—which ones work? Meta-analysis of published reports. *BMJ*. 2002;325(7370):925.
259. Hignett S, Jones EL, Miller D, Wolf L, Modi C, Shahzad MW, et al. Human factors and ergonomics and quality improvement science: integrating approaches for safety in healthcare. *BMJ Qual Saf*. 2015;24(4):250–4.
260. Naumann. How to develop a search strategy for a Cochrane Review [Internet]. Cochrane Haematological Malignancies Group. 2007 [cited 2014 Jan 31]. p. 1–10. Available from: <http://hm.cochrane.org/welcome-website-cochrane-haematological-malignancies>
261. Russ S, Rout S, Sevdalis N, Moorthy K, Darzi A, Vincent C. Do safety checklists improve teamwork and communication in the operating room? A systematic review. *Ann Surg*. 2013;258(6):856–71.
262. Chan Z, Kan C, Lee P, Chan I, Lam J. A systematic review of qualitative studies: Patients' experiences of preoperative communication. *J Clin Nurs*. 2012;21(5–6):812–24.
263. Shaw RL, Booth A, Sutton AJ, Miller T, Smith JA, Young B, et al. Finding qualitative research: an evaluation of search strategies. *BMC Med Res Methodol* [Internet]. 2004;4:1–5. Available from: <http://www.biomedcentral.com/1471-2288/4/5>
264. Forster DH, Krause G, Gastmeier P, Ebner W, Rath A, Wischniewski N, et al. Can quality circles improve hospital-acquired infection control? *J Hosp Infect*. 2000 Aug;45(4):302–10.
265. National Libraray of Medicine. PubMed Expert Searching: Using PubMed to Get Advanced Results [Internet]. 2006 [cited 2014 Jan 11]. Available from: <https://nnlm.gov/training/resources/pubmedalt.html>

266. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol*. 2006;59(7):697–703.
267. Rao N, Schilling D, Rice J, Ridenour M, Mook W, Santa E. Prevention of postoperative mediastinitis: a clinical process improvement model. *J Healthc Qual*. 2004;26(1):22–7.
268. McCulloch, P; Kreckler, S; New, S; Sheena, Y; Handa, A; Catchpole K. Effect of a “Lean” intervention to improve safety processes and outcomes on a surgical emergency unit. *Br Med J*. 2010;2; 341(Nov).
269. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney SE. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ*. 2009;338(Feb):402–4.
270. Vale C, Fitzgibbon J, Hanley B, Muir D, Murphy C, Nelson A, et al. INVOLVE, National Institute for Health Research: Public involvement in clinical trials: Supplement to the briefing notes for researchers [Internet]. Haynes H, editor. Eastleigh, Hampshire, UK; 2012. Available from: <http://www.nihr.ac.uk/get-involved/INVOLVEpublicinvolvementinclinicaltrialsBriefingnotes2012.pdf>
271. Jones EL, Williams-Yesson BA, Hackett RC, Staniszewska SH, Evans D, Francis NK. Quality of Reporting on Patient and Public Involvement Within Surgical Research: A Systematic Review. *Ann Surg*. 2015 Jun 19;261(2):243–50.
272. McDonald KM, Schultz EM, Chang C. Evaluating the state of Quality-Improvement Science through Evidence Synthesis: Insights from the Closing the Quality Gap Series. *Perm J*. 2013;17(4):52–61.
273. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews - A Product from the ESRC Methods Programme. 2006.
274. Light, RJ; Pillemer D. Summing up: the science of reviewing research. Cambridge, MA, USA: Harvard University Press; 1984.

275. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med*. 2013 Aug;46(1):81–95.
276. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004 Mar;94(3):361–6.
277. U.S. Department Of Health and Human Services - Food and Drug Administration. Changes or Modifications During the Conduct of a Clinical Investigation; Final Guidance for Industry and CDRH Staff [Internet]. 2001. Available from: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm082158.pdf>
278. NHS England. Standardise, educate, harmonise. Commissioning the conditions for safer surgery. Report of the NHS England Never Events Taskforce. Patient Safety Domain, 4-8 Maple Street, London; 2014.
279. Liabo K, Stewart R, Rees R, Oliver S. Patient and public perspectives shaping scientific and medical research: panels for data , discussions, and decisions. *Patient Intell* [Internet]. 2012;4:1–10. Available from: <https://www.dovepress.com/patient-and-public-perspectives-shaping-scientific-and-medical-research-peer-reviewed-article-PI>
280. Ziebland, S; Coulter, A; Calabrese, J; Locock L. Understanding and using health experiences - Improving patient care. Oxford OX2 6DP: Oxford University Press; 2013.
281. Nasser M. How can research publication be improved? *BMJ Blogs*. 2015.
282. Scott-Lichter D, and-The Editorial Policy Committee -, Council of Science Editors. Council of Science Editor's White paper on promoting integrity in scientific journal publications 2012 Update. [Internet]. 3rd ed. Wheat Ridge, CO, USA; 2012 [cited 2014 Mar 16]. Available from:

283. Staniszewska S, Adebajo A, Barber B, Beresford P, Brady L, Brett J, et al. Developing the evidence base of patient and public involvement in health and social care research: the case for measuring impact. *Int J Consum Stud*. 2011;35(6):628–32.
284. Maslin-Prothero S. Developing user involvement in research. *J Clin Nurs*. 2003 May;12(3):412–21.
285. NHS England. First steps towards quality improvement: a simple guide to improving services [Internet]. NHS Improvement. 2012 [cited 2014 Feb 12]. Available from: <http://qi.elft.nhs.uk/resource/first-steps-towards-quality-improvement-a-simple-guide-to-improving-services/>
286. Balik B. How Do You Involve Patients in Improvement? Audio and Video Resource [Internet]. Institute for Healthcare Improvement. 2015. Available from: <http://www.ihl.org/education/IHIOpenSchool/resources/Pages/Activities/Balik-IntegratingPatientsInDesign.aspx>
287. Staniszewska S, Brett J, Mockford C, Barber R. The GRIPP checklist: strengthening the quality of patient and public involvement reporting in research. *Int J Technol Assess Health Care*. 2011 Oct;27(4):391–9.
288. Jones, Emma L; Lees, N; Martin, G; Dixon-Woods M. How Well Is Quality Improvement Described in the Perioperative Care Literature? A systematic review. *Jt Comm J Qual Saf*. 2016;42(5):196–206.
289. Ritchie J, Lewis J. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Qualitative Research Practice: A Guide for Social Science Students and Researchers. London EC2A: SAGE publications Ltd; 2003.
290. Charmaz K. *Constructing grounded theory: a practical guide through qualitative analysis*. London: SAGE Publication Ltd; 2006.
291. Atkinson, R; Flint J. Snowball Sampling. In: Lewis-Beck M, Bryman A,

- Futing Liao T, editors. The SAGE Encyclopedia of Social Science Research Methods. CA: Thousand Oaks, SAGE publications; 2003.
292. Monitor. Facing the future: smaller acute providers [Internet]. Monitor, Wellington House, 133-155 Waterloo Road, London, SE1 8UG. Publication code: IRRES 05/14; 2014 [cited 2014 Mar 1]. Available from: <https://www.gov.uk/government/news/smaller-hospitals-have-a-future-in-the-nhs-but-need-to-change>
 293. Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, Titchen A. An exploration of the factors that influence the implementation of evidence into practice. *J Clin Nurs*. 2004 Nov;13(8):913–24.
 294. Bryman A. *Social Research Methods*. 3rd ed. Oxford: Oxford University Press; 2008.
 295. The Royal College of Surgeons -; The Surgical Specialities: 1 – General Surgery [Internet]. [cited 2015 May 14]. Available from: <https://www.rcseng.ac.uk/media/media-background-briefings-and-statistics/the-surgical-specialties-1-general-surgery>
 296. Horan B. NHS Hospital & Community Health Service (HCHS) monthly workforce statistics - Provisional Statistics. Workforce and facilities team, NHS Digital. 2016.
 297. Williams TE, Satiani B, Thomas A, Ellison EC. The impending shortage and the estimated cost of training the future surgical workforce. *Ann Surg*. 2009;250(4):590–7.
 298. De Vries EN, Hollmann MW, Smorenburg SM, Gouma DJ, Boermeester MA. Development and validation of the SURgical PATient Safety System (SURPASS) checklist. *Qual Saf Health Care*. 2009 Apr;18(2):121–6.
 299. Flanagan A. Order of Authorship. In: American Medical Association., editor. *American Medical Association Manual of Style: A Guide for Authors and Editors* [Internet]. 10th ed. Oxford OX2 6DP: Oxford University Press; 2007. p. 128–40. Available from: www.amamanualofstyle.com

300. Glaser BG, Strauss A. The discovery of grounded theory: strategies for qualitative research. New York: Aldine Publishing Co; 1967.
301. Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. *Health Policy (New York)*. 2008;85(2):148–61.
302. Pennathur PR, Thompson D, Abernathy JH, Martinez EA, Pronovost PJ, Kim GR, et al. Technologies in the wild (TiW): human factors implications for patient safety in the cardiovascular operating room. *Ergonomics*. 2013 Feb;56(2):205–19.
303. Burnett S, Benn J, Pinto A, Parand A, Iskander S, Vincent C. Organisational readiness: exploring the preconditions for success in organisation-wide patient safety improvement programmes. *Qual Saf Health Care*. 2010 Aug;19(4):313–7.
304. Marshall, Catherine; Rossman G. *Designing Qualitative Research*. 4th ed. London EC1Y 1SP: Thousands Oaks: Sage; 2006.
305. Strauss, Anselm L; Corbin J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. SAGE Publication limited; 1990.
306. Blumer H. *Symbolic interactionism: Perspective and method*. Engelwood Cliffs, New Jersey: Prentice – Hall Inc; 1969.
307. Hammersley M. The dilemma of qualitative method: Herbert Blumer and the Chicago tradition. London EC4P 4EE: Routledge; 1989.
308. Charmaz K. Grounded Theory: Objective and Constructivist Methods. In W. Luttrell (Ed.) *Qualitative Educational Research: Reading in reflective methodology and transformative practice*. (pp.183-207). London: Routeledge; 2010.
309. Mays N, Pope C. Rigour and qualitative research. *BMJ*. 1995;311(6998):182–4.
310. Long T, Johnson M. Rigour, reliability and validity in qualitative research.

- Clin Eff Nurs [Internet]. 2000 [cited 2016 Apr 11];4(1):30–7. Available from:
<http://www.sciencedirect.com/science/article/pii/S1361900400901067>
311. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*. 2001;322(7294):1115–7.
 312. Soloman B, Felder R. Felder Soloman Index of Learning Styles Questionnaire [Internet]. Available from:
<http://www.engr.ncsu.edu/learningstyles/ilsweb.html>
 313. Medicines and Healthcare products Regulatory Agency. Medical Device Alert: All metal-on-metal (MoM) hip replacements Ref MDA/2012/036 [Internet]. 2012 [cited 2015 Aug 15]. Available from: <https://www.health-ni.gov.uk/articles/medical-device-alerts-mdas-2003-2012>
 314. Seale C, Silverman D. Ensuring rigour in qualitative research. *Eur J Public Health*. 1997;7(4):379–84.
 315. Data Protection Act 1998 - Act of parliament of the United Kingdom of Great Britain and Northern Ireland [Internet]. London, Great Britain: Crown Copyright, The stationary office; 1998 [cited 2014 May 17]. p. 1–86. Available from: <http://www.legislation.gov.uk/ukpga/1998/29/section/7>
 316. Social Science Applied to Healthcare Improvement Research. 010 – Standard Operating Proceedure (SOP) - Information Governance guidelines - ensuring that you're protecting confidential and sensitive information - SAPPHIRE - 2013-10-01 v1. 2015.
 317. UK Parliament. Statutory Instrument No. 1031: The Medicines for Human Use (Clinical Trials) Regulations [Internet]. London; 2004. Available from: http://www.legislation.gov.uk/uksi/2004/1031/pdfs/uksi_20041031_en.pdf
 318. Polgar, Stephen; Thomas S. Introduction to Research in the Health Sciences. 5th ed. Commissioning Editor: Steven Black, editor. London: Churchill Livingstone Elsevier; 2008.
 319. Bogner, A; Littig, B; Menz W. Interviewing Experts. Hampshire RG21

6XS: Palgrave Macmillan; 2009.

320. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract* [Internet]. 2000 Feb;17 Suppl 1:S11-6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10735262>
321. Larochelle M, Hyman N, Gruppi L, Osler T. Diminishing surgical site infections after colorectal surgery with surgical care improvement project: is it time to move on? *Dis Colon Rectum*. 2011 Apr;54(4):394–400.
322. Pronovost PJ, Jenckes M, To M, Dorman T, Lipsett PA, Berenholtz S, et al. Reducing failed extubations in the intensive care unit. *Jt Comm J Qual Improv*. 2002;28(11):595–604.
323. Bloos F, Müller S, Harz A, Gugel M, Geil D, Egerland K, et al. Effects of staff training on the care of mechanically ventilated patients: a prospective cohort study. *Br J Anaesth*. 2009 Aug;103(2):232–7.
324. Brattebø G, Hofoss D, Flaatten H, Muri a K, Gjerde S, Plsek PE. Effect of a scoring system and protocol for sedation on duration of patients' need for ventilator support in a surgical intensive care unit. *Qual Saf Health Care*. 2004 Jun;13(3):203–5.
325. Taber DJ, Pilch N a, McGillicuddy JW, Bratton CF, Chavin KD, Baliga PK. Improved patient safety and outcomes with a comprehensive interdisciplinary improvement initiative in kidney transplant recipients. *Am J Med Qual*. 2013;28(2):103–12.
326. Stevenson KS, Gibson SC, MacDonald D, Hole DJ, Rogers PN, Byrne DS, et al. Measurement of process as quality control in the management of acute surgical emergencies. *Br J Surg*. 2007 Mar;94(3):376–81.
327. Parks JK, Klein J, Frankel HL, Friese RS, Shafi S. Dissecting delays in trauma care using corporate lean six sigma methodology. *J Trauma*. 2008;65(5):1098–104.
328. Duncan F, Haigh C. Measuring and improving the quality of postoperative

- epidural analgesia for major abdominal surgery using statistical process control charts. *J Clin Nurs*. 2013 Oct;22(19–20):2748–57.
329. Baker R, Newland R. Continuous quality improvement of perfusion practice: the role of electronic data collection and statistical control charts. *Perfusion* [Internet]. 2008 Jan 1 [cited 2015 Jan 10];23(1):7–16. Available from: <http://prf.sagepub.com/cgi/doi/10.1177/0267659108093853>
 330. Vetter TR, Ali NMK, Boudreaux AM. A case-control study of an intraoperative corneal abrasion prevention program: holding the gains made with a continuous quality improvement effort. *Jt Comm J Qual Patient Saf*. 2012;38(11):490–6.
 331. Thomassen Ø, Brattebø G, Sjøfteland E, Lossius HM, Heltne J-K. The effect of a simple checklist on frequent pre-induction deficiencies. *Acta Anaesthesiol Scand*. 2010 Nov 24;54(10):1179–84.
 332. Berenholtz SM, Pronovost PJ, Lipsett PA, Hobson D, Earsing K, Farley JE, et al. Eliminating catheter-related bloodstream infections in the intensive care unit. *Crit Care Med*. 2004 Oct;32(10):2014–20.
 333. Roberts S, Kennedy D, MacLeod AM, Findlay H, Gollish J. A framework for the development and implementation of an advanced practice role for physiotherapists that improves access and quality of care for patients. *Healthc Q*. 2008 Jan;11(2):67–75.
 334. Gray J, Boland L, Machemer CH, Nelson M. Elimination of excess perioperative wait time and its impact on immediate clinical outcomes among patients undergoing surgery. *J Nurs Care Qual*. 2010;25(3):248–54.
 335. Cronin, J; Livhits, M; Mercado, C; Chen, F; Foster, N; Chandler, C; Gibbons, M; Ko, CY; Chen D. Quality improvement pilot program for vulnerable elderly surgical patients. *Am Surg*. 2011;77(10):1305–8.
 336. Ramsey, C; Ormsby, S; Marsh T. Performance-improvement strategies can reduce costs. *Heal Financ Manag*. 2001;Suppl 2-6.

337. Bertolaccini L, Rizzardi G, Filice MJ, Terzi A. “Six Sigma approach” - an objective strategy in digital assessment of postoperative air leaks: A prospective randomised study. *Eur J Cardio-thoracic Surg*. 2011;39(5):e128–32.
338. Meissner W, Ullrich K, Zwacka S. Benchmarking as a tool of continuous quality improvement in postoperative pain management. *Eur J Anaesthesiol*. 2006;23:142–8.
339. Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual “richness” for systematic reviews of complex interventions: case study (CLUSTER). *BMC Med Res Methodol* [Internet]. 2013;13(118). Available from: <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-118>
340. Burkitt KH, Mor MK, Jain R, Kruszewski MS, Mccray EE, Moreland ME, et al. Toyota Production System Quality Improvement Initiative Improves Perioperative Antibiotic Therapy. 2009;15(9):633–42.
341. US Congress. H.R.3548 - Improving Trauma Care Act of 2014. 2014;August. Available from: <https://www.congress.gov/bill/113th-congress/house-bill/3548>
342. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol*. 2002 Feb;31(1):115–23.
343. Abdul Wahid NN, Moppett SH, Moppett IK, Keith Moppett I. Quality of Quality Accounts: transparency of public reporting of Never Events in England. A semi-quantitative and qualitative review. 2016;109(5):190–9.
344. Catchpole KR, Dale TJ, Hirst PDG, Smith PJP, Giddings TAEB. A Multicenter Trial of Aviation-Style Training for Surgical Teams. *J Patient Saf*. 2010;6(3):180–6.
345. Smith CD, Spackman T, Brommer K, Stewart MW, Vizzini M, Frye J, et al.

- Re-engineering the operating room using variability methodology to improve health care value. *J Am Coll Surg*. 2013 Apr;216(4):559–70.
346. Song, J; Chung K. Observational Studies: Cohort and Case-Control Studies. *Plast Reconstr Surg*. 2010;126(6):2234–42.
 347. Dekkers OM, Egger M, Altman DG, Vandenbroucke JP. Research and Reporting Methods: Distinguishing Case Series From Cohort Studies. *Ann Intern Med*. 2012;156:37–40.
 348. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Heal Care*. 2003;12:458–64.
 349. Fretheim A, Tomic O. Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement. *BMJ Qual Saf*. 2015;24(12):748–52.
 350. Effective Practice and Organisation of Care (EPOC) Group. Interrupted time series (ITS) analyses: a resource for EPOC review authors [Internet]. Cochrane Collaboration. 2013 [cited 2016 Oct 15]. p. 1–13. Available from:
[https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/21 Interrupted time series analyses 2013 08 12_1.pdf](https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/21%20Interrupted%20time%20series%20analyses%202013%2008%2012_1.pdf)
 351. Tricco, A; Anthony, J; Ivers, N; Ashoor H; Khan, P; Blondal E; Ghassemi, M; MacDonald, H; Chen, M; Kark Ezer, L; Straus S. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services-systematic review and meta-analysis. *CMAJ*. 2014 Sep 15;1–11.
 352. Mauger B, Marbella A, Pines E, Chopra R, Black ER, Aronson N. Implementing quality improvement strategies to reduce healthcare-associated infections: A systematic review. *Am J Infect Control*. Elsevier Inc; 2014 Oct;42(10 Suppl):S274-83.
 353. Ford, Alan; Peat D. The Role of Language in Science. *Found Phys*. 1988;18(12):1233–42.

354. McCulloch P. Developing appropriate methodology for the study of surgical techniques. *J R Soc Med*. 2009 Feb;102(2):51–5.
355. Harman, Joseph E; Gross AG. *The Craft of Scientific Communication*. Chicago: University of Chicago Press; 2010.
356. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials [Internet]. 2011. Available from: http://handbook.cochrane.org/chapter_8/8_5_the_cochrane_collaboration_s_tool_for_assessing_risk_of_bias.htm
357. Howick M, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. The Oxford 2011 levels of evidence [Internet]. OCEMB Levels of Evidence working group. 2011. Available from: <http://www.cebm.net/index.aspx?o=5653>
358. Davidoff F, Dixon-woods M, Leviton L, Michie S. Demystifying theory and its use in improvement. *BMJ Qual Saf*. 2015;(January):1–11.
359. Reeves ABC, Deeks JJ, Higgins JPT, Wells GA. *Cochrane Handbook for Systematic Reviews of Interventions: Chapter 13 : Including non-randomized studies*. version 5. Chichester, West Sussex: John Wiley & Sons Ltd; 2008. 1-34 p.
360. Marsteller JA, Holzmüller CG, Makary M, Sexton JB, Thompson DA, Lubomski LH, et al. Developing process-support tools for patient safety: Finding the balance between validity and feasibility. *Jt Comm J Qual Patient Saf*. 2008;34(10):604–7.
361. Cucchetti A, Mazzotti F, Pellegrini S, Cescon M, Maroni L, Ercolani G, et al. The use of the Hirsch index in benchmarking hepatic surgery research. *Am J Surg*. Elsevier Inc; 2013;206(4):560–6.
362. Garfield E. The history and meaning of the journal Impact Factor. *JAMA*. 2006;295(1):90–3.
363. Web of knowldege. Web of knowledge: Journal citation reports help page

- [Internet]. 2016 [cited 2016 Nov 24]. Available from: http://admin-apps.webofknowledge.com/JCR/help/h_impfact.htm
364. Miller K. Chapter 1: The challenge of organizational communication. In: Organizational communication: Approaches and processes. Stamford, CT: Cengage Learning; 2015.
 365. Cummings GG, Estabrooks C a., Midodzi WK, Wallin L, Hayduk L. Influence of Organizational Characteristics and Context on Research Utilization. *Nurs Res.* 2007;56(Supplement 1):S24–39.
 366. Yano EM. The role of organizational research in implementing evidence-based practice: QUERI Series. *Implement Sci.* 2008 Jan;3(29).
 367. NHS -England. NHS Standard Contract 2016/17 Particulars [Internet]. NHS England. Crown copyright; Available from: <https://www.england.nhs.uk/nhs-standard-contract/16-17/>
 368. Department of Health. Payment by Results Guidance for 2012-13. Payment by results guidance for 2012-13. 2012.
 369. Department of Health. The NHS as an Innovative organisation: A framework and guidance on the management of intellectual property in the NHS [Internet]. London: Crown copyright; 2002. Available from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4002660
 370. (ICMJE) IC of MJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [Internet]. 2015. p. 1–17. Available from: <http://www.icmje.org/recommendations/>
 371. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane database Syst Rev* [Internet]. 2012;Nov 14(11). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23152285>

372. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Assoc.* 2004;92(3):364–7.
373. Sadana R, Shook-Pui Lee M, Lee J. Health Research System Analysis (HRSA) Initiative: Methods for Collecting Benchmarks and Systems Analysis Toolkit - Tool 1: A brief overview of WHO Health Research System Analysis initiative. Geneva; 2006.
374. Caffrey L, Wolfe C, McKevitt C. Embedding research in health systems: Lessons from complexity theory. *Heal Res Policy Syst. Health Research Policy and Systems*; 2016;14(1):54.
375. Marshall M, Pagel C, French C, Utley M, Allwood D, Fulop N, et al. Moving improvement research closer to practice: the Researcher-in-Residence model. *BMJ Qual Saf.* 2014 Oct;23(10):801–5.
376. Chew S, Armstrong N, Martin G. Institutionalising knowledge brokering as a sustainable knowledge translation solution in healthcare: How can it work in practice? *Evid Policy.* 2013;9(3):335–51.
377. Adams M, Caffrey L, McKevitt C. Barriers and opportunities for enhancing patient recruitment and retention in clinical research: findings from an interview study in an NHS academic health science centre. *Heal Res policy Syst.* 2015;13(8).
378. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf.* 2016;25(8):626–32.
379. Lane J, Flagg J. Translating three states of knowledge - discovery, invention & innovation. *Implement Sci* [Internet]. 2010;5(1):9. Available from: <http://dx.doi.org/10.1186/1748-5908-5-9>
380. Gawande A. *The Checklist Manifesto: How to get things right.* London EC1R 0JH: Profile Books Ltd; 2010.
381. Dixon-Woods M. Perspectives on context: The problem of context in

- quality improvement. The Health Foundation. London; 2011.
382. Thomas W. Teaching and assessing surgical competence. *Ann R Coll Surg Engl.* 2006;88(5):429–32.
 383. Benner P. From novice to expert: Excellence and Power in Clinical Nursing Practice; Commemorative edition. Upper Saddle River, N.J: Prentice Hall; 2001.
 384. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. *Implement Sci. Implementation Science;* 2016;11(1):12.
 385. Squires JE, Graham ID, Hutchinson AM, Linklater S, Brehaut JC, Curran J, et al. Understanding context in knowledge translation: A concept analysis study protocol. *J Adv Nurs.* 2015;71(5):1146–55.
 386. Martin GP, Dixon-Woods M. Can we tell whether hospital care is safe? *Br J Hosp Med.* 2014;75(9):484–5.
 387. Guba EG, Lincoln YS. Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS, editors. *Handbook of qualitative research.* London: Thousand Oaks, SAGE Publication limited; 1994. p. 105–17.
 388. Kukull WA, Ganguli M. Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology.* 2012;78(23):1886–91.
 389. Zwarentein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ.* 2008;Nov 11(337).
 390. Rycroft-Malone J, Kitson A, Harvey G, McCormack B, Seers K, Titchen A, et al. Ingredients for change: revisiting a conceptual framework. *Qual Saf Health Care.* 2002 Jun;11(2):174–80.
 391. Schierhout G, Hains J, Si D, Kennedy C, Cox R, Kwedza R, et al. Evaluating the effectiveness of a multifaceted, multilevel continuous quality improvement program in primary health care: developing a realist theory of change. *Implement Sci.* 2013;Oct(8):119.

392. Power M, Tyrrell PJ, Rudd AG, Tully MP, Dalton D, Marshall M, et al. Did a quality improvement collaborative make stroke care better? A cluster randomized trial. *Implement Sci.* 2014;9(1):40.
393. Group. LS of E public policy. Maximizing the impacts of your research: a handbook for social scientists. Consultation draft 3 [Internet]. London, UK; 2011. Available from: <http://www2.lse.ac.uk/government/research/resgroups/LSEPublicPolicy/Home.aspx>
394. Leatherman S, Sutherland K. Designing national quality reforms: A framework for action. *Int J Qual Heal Care.* 2007;19(6):334–40.
395. Øvretveit J. Does improving quality save money ? A review of evidence of which improvements to quality reduce costs to health service providers. The Health Foundation. London; 2009.
396. The Central Office for Research Ethics Committees (COREC) and amended by the National Research Ethics Service (NRES) (a division of the National Patient Safety Agency). Defining Research: NRES guidance to help you decide if your project requires review by a Research Ethics Committee (rev. April 2013; June 2016) [Internet]. 2009 [cited 2016 Nov 4]. p. 1–4. Available from: <http://www.hra.nhs.uk/documents/2016/06/defining-research.pdf>
397. Martin G, Currie G, Weaver S, Finn R, McDonald R. Institutional Complexity and Individual Responses: Delineating the Boundaries of Partial Autonomy. *Organ Stud.* 2016;August 23:1–25.
398. Greenhalgh T. What is this knowledge that we seek to “exchange”? *Milbank Q.* 2010;88(4):492–9.
399. Van de Ven A, Johnson P. Knowledge for Theory and Practice. *Acad Manag Rev.* 2006;31(4):802–21.
400. Miller C. Medicine is not science: Guessing the future, predicting the past. *J Eval Clin Pract.* 2014;20(6):865–71.

401. Poincaré H. *Science and Method*. London: Thomas Nelson and Sons.; 1914.
402. Engebretsen E, Heggen K, Wieringa S, Greenhalgh T. Uncertainty and objectivity in clinical decision making: a clinical case in emergency medicine. *Med Health Care Philos*. 2016;19(4):595–603.
403. Solberg, Mosser M (1997)-. Three faces of performance measurement: Improvement accountability and research. *J Qual Improv*. 1997;23(3).
404. Fabry G. Replication – The ugly duckling of science? *GMS Z Med Ausbild*. 2015;32(5):1–7.
405. Mosher H, Ogrinc G. Between the guidelines: SQUIRE 2.0 and advances in healthcare improvement practice and reporting. *BMJ Qual Saf*. 2016;25(8):559–61.
406. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med*. 2011;104(12):510–20.
407. De Salis I, Tomlin Z, Toerien M, Donovan J. Qualitative research to improve RCT recruitment: Issues arising in establishing research collaborations. *Contemp Clin Trials*. 2008;29(5):663–70.
408. Beard D, Rees J, Rombach I, Cooper C, Cook J, Merritt N, et al. The CSAW Study (Can Shoulder Arthroscopy Work?) - A Placebo-Controlled Surgical Intervention Trial Assessing the Clinical and Cost Effectiveness of Arthroscopic Subacromial Decompression for Shoulder Pain: Study Protocol for a Randomised Controlled Trial. *Trials [Internet]*. ???; 2015;16(1):210. Available from: <http://www.trialsjournal.com/content/16/1/210>
409. Lawton R, Taylor N, Clay-Williams R, Braithwaite J. Positive deviance: a different approach to achieving patient safety. *BMJ Qual Saf*. 2014;23(11):880–3.
410. Brewster L, Aveling E-L, Martin G, Tarrant C, Dixon-Woods M. What to

- expect when you're evaluating healthcare improvement: a concordat approach to managing collaboration and uncomfortable realities. *BMJ Qual Saf.* 2015;24(5):318–24.
411. Greenhalgh T, Annandale E, Ashcroft R, Barlow J, Black N, Bleakley A, et al. An open letter to The BMJ editors on qualitative research. *BMJ* [Internet]. 2016;352(February):i563. Available from: <http://www.bmj.com/lookup/doi/10.1136/bmj.i563>
 412. Hsu R, Douglas L. Leicester, Leicestershire and Rutland (LLR) Joint Primary and Secondary Care Mortality Review. 2014;(July):1–26. Available from: www.leicestershospitals.nhs.uk/aboutus/performance/learning-lessons-to-improve-care
 413. National Institute for Health and Clinical Excellence. Prevention and control of healthcare-associated infections: Quality improvement guide, NICE public health guidance 36. London SW1A 2BU; 2011.
 414. Øvretveit JC, Shekelle PG, Dy SM, McDonald KM, Hempel S, Pronovost P, et al. How does context affect interventions to improve patient safety? An assessment of evidence from studies of five patient safety practices and proposals for research. *BMJ Qual Saf.* 2011 Jul;20(7):604–10.
 415. Currie G, Lockett A, Finn R, Martin G, Waring J. Institutional Work to Maintain Professional Power: Recreating the Model of Medical Professionalism. *Organ Stud.* 2012;33(7):937–62.
 416. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials* [Internet]. 2009 Jan [cited 2014 Jan 25];10(9). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2654883&tool=pmcentrez&rendertype=abstract>
 417. Cartwright N. What are randomised controlled trials good for? *Philos Stud.* 2009;147(1):59–70.
 418. Harris A, McGregor J, Perencevich E, Furuno J, Zhu J, Peterson D, et al.

- The Use and interpretation of Quasi-Experimental Studies in Medical Informatics. *J Am Med Inf Assoc*. 2006;13(1):16–23.
419. Grix J. *Demystifying Postgraduate Research: From MA to PhD*. Birmingham: University of Birmingham Press; 2001.
 420. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question ? An introduction to the PICOT format for clinicians. *J Chiropractic Assoc*. 2012;56(3):167–71.
 421. Westfall JM, Mold J, Fagnan L. Practice-based research - “Blue Highways” on the NIH Roadmap. *JAMA*. 2007;297(4):403–6.
 422. Chopra V, Shojania KG. Recipes for checklists and bundles: one part active ingredient, two parts measurement. *BMJ Qual Saf*. 2013 Feb;22(2):93–6.
 423. O’Hara JK, Lawton RJ. At a crossroads? Key challenges and future opportunities for patient involvement in patient safety. *BMJ Qual Saf*. 2016;25(8):565–8.
 424. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE team; Process evaluation in randomised controlled trials of complex interventions. *BMJ*. 2006;332:413–6.
 425. Hildebrand MW, Host HH, Binder EF, Carpenter B, Freedland KE, Morrow-howell N, et al. Measuring Treatment Fidelity in a Rehabilitation Intervention Study. *Am J Phys Med Rehabil*. 2014;91(8):715–24.
 426. Bowen DJ, Kreuter M, Spring B, Linnan L, Weiner D, Bakken S, et al. How We Design Feasibility Studies. *Am J Prev Med*. 2010;36(5):452–7.
 427. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J*. 2015;350(19).
 428. Harrop E, Kelly J, Griffiths G, Casbard A, Nelson A. Why do patients decline surgical trials? Findings from a qualitative interview study embedded in the Cancer Research UK BOLERO trial (Bladder cancer:

- Open versus Laparoscopic or RObotic cystectomy). *Trials* [Internet]. *Trials*; 2016;17(35). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4719666&tool=pmcentrez&rendertype=abstract>
429. Gooberman-Hill R, Fox R. What can qualitative approaches bring to trauma outcome research? *Injury*. 2011 Apr;42(4):321–3.
 430. O’Cathain A, Hoddinott P, Lewin S, Thomas KJ, Young B, Adamson J, et al. Maximising the impact of qualitative research in feasibility studies for randomised controlled trials: guidance for researchers. *Pilot Feasibility Stud* [Internet]. *Pilot and Feasibility Studies*; 2015;1(32). Available from: <http://pilotfeasibilitystudies.biomedcentral.com/articles/10.1186/s40814-015-0026-y>
 431. Pearse R. Enhanced peri-operative care for high-risk patients (EPOCH) trial: a stepped wedge cluster randomised trial of a quality improvement intervention for patients undergoing emergency laparotomy [Internet]. Pragmatic Clinical Trials Unit Queen Mary University of London. 2014 [cited 2016 Oct 31]. p. 1–28. Available from: <http://www.epochtrial.org/docs/EPOCH Protocol revision V2 0 280414.pdf>
 432. Bloom BR, Michaud CM, Montagne JR La, Simonsen L. Chapter 4 Priorities for Global Research and Development of Interventions. In: Jamison D, Breman J, Measham A, editors. *Disease Control Priorities in Developing Countries*. 2nd ed. New York: Oxford University Press; 2006. p. 103–18.
 433. Strauss, Anselm L; Corbin J. *Basics of qualitative research: techniques and procedures for developing grounded theory*. 2nd ed. London: Thousand Oaks, Sage publications; 1998.
 434. Morse J. Designing funded qualitative research. In: Denzin N, Lincoln Y, editors. *Handbook of qualitative research*. 2nd ed. London EC1Y 1SP: Thousand Oaks, Sage publications; 1994. p. 220–35.
 435. McCulloch P. Patient safety and rocket science. *BMJ Qual Saf*.

2016;25(8):562–4.