Improving Interdisciplinary Care on the General Medical Ward

A thesis presented for the degree of

Doctor of Philosophy (PhD)

Samuel Pannick MA MB BS MRCP

NIHR Imperial Patient Safety Translational Research Centre

Department of Surgery and Cancer

Imperial College London

August 2016

Abstract

General medical wards deliver the majority of inpatient care. Despite technological and therapeutic advances, these wards expose 10% of patients to preventable adverse events, and disproportionately contribute to preventable hospital deaths. Improving ward team performance is often proposed as a mechanism to improve patient outcomes.

The overarching goal of this thesis is to identify effective strategies to improve interdisciplinary team care on the medical ward. Chapter 1 introduces key concepts in healthcare quality, and specific issues in the delivery and measurement of interdisciplinary ward care. The existing literature for ward improvement strategies is then described. A narrative review identifies common targets for ward interventions [chapter 2], and a systematic review evaluates interdisciplinary team care interventions, finding little evidence of significant impact on objective patient outcomes [chapter 3].

The development and evaluation of prospective clinical team surveillance (PCTS) is then reported. PCTS is a novel interdisciplinary team care intervention, engaging staff to identify barriers to care delivery, with facilitation and feedback. A programme theory and mixed methods evaluation are presented, using a stepped wedge, cluster controlled trial [chapter 4]. Mixed-effects models show a significant reduction in excess length of stay with high fidelity PCTS [chapter 5]. Surveys, focus groups and auto-ethnography identify PCTS' mechanisms of action, and its impact on incident reporting, safety and teamwork climates [chapter 6]. Implementation outcomes, facilitators and barriers are described in chapter 7.

Other perspectives on improvement are also explored. A model of organisational alignment

is developed [chapter 8], and an interview study with patients and carers elicits their priorities [chapter 9]. Finally, chapter 10 summarises the findings, highlighting opportunities to develop medical ward outcome sets and construct a model of interdisciplinary team effectiveness. These can be used to support improvements in interdisciplinary care, through changes in policy and practice.

Declaration of originality

I declare that the work in this thesis is my own. Any contributions by my colleagues are acknowledged appropriately.

Copyright notice

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial-No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the license terms of this work.

Acknowledgements

I am immensely grateful for the wisdom and kindness of those who have guided me over the last three years. I would like to thank my wonderfully supportive supervisors. Nick Sevdalis welcomed me and gave me an academic home. His patience and good humour were always appreciated. I was lucky to find my way to Thanos Athanasiou, a true polymath: I treasured his advice, all the more so for its unflinchingly honest delivery. Iain Beveridge never faltered in his backing. I was continually thankful for his support and encouragement, as I was for Susy Long's. It was a true privilege to work with Bob Wachter, who gently reined in my wilder claims about healthcare delivery. I am indebted to him for his mentorship.

My family kept me going - especially my father, who is a hard act to follow. I thank the Imperial College Healthcare Charity for its support, and staff at the West Middlesex and St Mary's Hospitals. Fran Husson and Margaret Turley, PSTRC patient and carer representatives, generously helped develop the interview guide. Elliot Berger bravely read drafts whilst himself a medical patient. Hutan Ashrafian, Rachel Davis, and Ben Byrne all provided invaluable support for the systematic review. Max Johnston, Louise Hull, Tayana Soukup and Stephanie Archer coordinated focus groups, with further support from Stephanie Archer for interviews. As researchers, their professionalism was beyond reproach. More valuable still was their unhelpful advice as the reprobates of room 503.

My greatest thanks go to Daisy and Noah

Most of all to my wonderful Hayley, who makes it all possible

And to Mum, who put up with a lot of hospitals

Publications, grants and awards arising from this thesis

Publications

- Pannick S, Wachter RM, Vincent C, Sevdalis N. Rethinking medical ward quality. BMJ 2016; 355:i5417. DOI: 10.1136/bmj.i5417
- Pannick S, Sevdalis N, Athanasiou T. Beyond clinical engagement: a pragmatic model for quality improvement interventions, aligning clinical and managerial priorities. BMJ Qual Saf 2016; 25:716-25. DOI: 10.1136/bmjqs-2015-004453
- Pannick S, Davis R, Ashrafian H, Byrne BE, Beveridge I, Athanasiou T, Wachter RM, Sevdalis N. Effects of Interdisciplinary Team Care Interventions on General Medical Wards: A Systematic Review. JAMA Intern Med 2015; 175(8):1288-98. DOI: 10.1001/jamainternmed.2015.2421.
- Pannick S, Beveridge I, Ashrafian H, Long SJ, Athanasiou T, Sevdalis N. A stepped wedge, cluster controlled trial of an intervention to improve safety and quality on medical wards: the HEADS-UP study protocol. BMJ Open 2015; 5:e007510. DOI: 10.1136/bmjopen-2014-007510
- Pannick S, Beveridge I, Wachter RM, Sevdalis N. Improving the quality and safety of care

on the medical ward: A review and synthesis of the evidence base. Eur J Intern Med 2014; 25(10):874-87. DOI: 10.1016/j.ejim.2014.10.013

Presentations

- Pannick S, Beveridge I, Athanasiou T, Long SJ, Sevdalis S. HEADS-UP: Prospective clinical team surveillance improves safety climate, incident reporting and patient outcomes. A cluster controlled stepped wedge trial. Plenary presentation (Best of Research and Innovations). Society of Hospital Medicine Annual Meeting (March 2016, San Diego, USA)
- Pannick S, Beveridge I, Long SJ, Athanasiou T, Sevdalis N. HEADS-UP: A novel intervention to improve staff engagement & safety climate with daily, team-based risk recognition. Royal College of Physicians Annual Conference (March 2015, Harrogate)

Invited talks

- Pannick, S. Diagnosing service problems. Management of the Labour Ward - joint meeting of the Royal College of Obstetricians and Gynaecologists and the British Maternal and Fetal Medicine Society (June 2016, London)
- Pannick, S. Diagnosing service problems. Management of the Labour Ward - joint meeting of the Royal College of Obstetricians and Gynaecologists and the British Maternal and Fetal Medicine Society (June 2015, London)

Grants

- HEADS-UP (Hospital Event Analysis Describing Significant Unanticipated Problems): Improving the quality of medical ward care with daily, team-based risk recognition (£43,993). Imperial College Healthcare Charity, 2014.

Awards

- HEADS-UP team, West Middlesex University Hospital NHS Trust. HEADS-UP: Best example of interprofessional learning in secondary care - highly commended. Excellence in Education and Training Awards, Health Education North West London, 2015.
- West Middlesex University Hospital NHS Trust. HEADS-UP. Patient safety - Finalist. HSJ Awards, 2015.

Peer review

- Peer reviewer for JAMA Internal Medicine, Implementation Science, BMJ Quality & Safety, Journal of Hospital Medicine, BMJ Simulation and Technology Enhanced Learning, and British Journal of Hospital Medicine.

Contents

Abstract iii

Declaration of originality v

Acknowledgements vii

Publications, grants and awards arising from this thesis ix

List of Figures xxiii

List of Tables xxv

Part I Introduction 1

Chapter 1. The medical ward microsystem 3

1.1 Introduction 3

1.2 Key concepts in patient safety and healthcare quality 5

1.2.1 Changing perceptions of patient safety 5

1.2.2 Systems approaches, blame and individual accountability 6

1.2.3 Definitions of quality in healthcare 7

1.2.4 Changes in healthcare safety over time 7

1.2.5 Quality improvement 8

1.2.6 Evaluating quality improvement interventions . . . 11
1.3 Medical ward design and care delivery . . . 13
1.3.1 Ward design . . . 13
1.3.2 The medical ward as a clinical microsystem . . . 14
1.4 Serious failings in medical ward care . . . 17
1.4.1 Problems in routine care . . . 18
1.4.2 Problems in emergency care . . . 19
1.5 Understanding ward-level performance . . . 21
1.5.1 Ward-level variation *vs* hospital-level variation . . . 22
1.5.2 Aggregate hospital-wide metrics . . . 23
1.5.3 Measuring quality in ward care . . . 26
1.6 Aims of the thesis . . . 29

Part II Evidence-based strategies to improve medical ward care 33

Chapter 2. Improving medical ward care: a narrative review 35

2.1 Introduction . . . 35
2.2 Aims . . . 36
2.3 Methods . . . 36
2.3.1 Search strategy . . . 36
2.3.2 Article selection . . . 37
2.4 Results . . . 38
2.4.1 Staffing and team composition . . . 55
2.4.2 Staff communication and collaboration . . . 58
2.4.3 Standardisation of the processes of care . . . 62
2.4.4 Early recognition and treatment of the deteriorating patient . . . 64
2.4.5 Ward safety climate . . . 65

2.4.6 Combined approaches . . . 66
2.5 Discussion . . . 66
2.5.1 Summary of findings . . . 67
2.5.2 Study limitations . . . 70
2.5.3 Implications for further work . . . 71
2.5.4 Conclusion . . . 71

Chapter 3. Interdisciplinary team care interventions: a systematic review 73

3.1 Introduction . . . 73
3.2 Aims . . . 74
3.3 Methods . . . 75
3.3.1 Search strategy & data sources . . . 75
3.3.2 Article selection: inclusion & exclusion criteria . . . 75
3.3.3 Screening & data extraction . . . 79
3.3.4 Assessment of study quality . . . 80
3.3.5 Statistical analysis . . . 80
3.4 Results . . . 81
3.4.1 Abstract screening & application of the inclusion criteria . . . 81
3.4.2 Study characteristics . . . 81
3.4.3 Outcome risk adjustment and analysis . . . 95
3.4.4 Complications of care and discharge location . . . 95
3.4.5 Length of stay . . . 96
3.4.6 Readmissions and use of outpatient care . . . 98
3.4.7 Mortality . . . 100
3.5 Discussion . . . 100
3.5.1 Summary of findings . . . 102
3.5.2 Study limitations . . . 103

3.5.3 Implications for further work . . . 104
3.5.4 Conclusions . . . 105

Part III Prospective clinical team surveillance for medical wards 107

Chapter 4. Prospective clinical team surveillance: a programme theory & study protocol 109

4.1 Introduction . . . 109
4.2 Aims . . . 110
4.3 Frontline staff: an underused source of information . . . 110
4.3.1 Frontline reflections on care delivery correlate with patient survival . 110
4.3.2 Frontline staff default to first-order problem solving, hindering organisational improvement . . . 111
4.3.3 Collating team experiences, rather than individual reports, may be beneficial . . . 112
4.4 Prospective clinical team surveillance: a programme theory . . . 113
4.4.1 Prospective clinical team surveillance: structured interdisciplinary team briefings, facilitated escalation, and feedback . . . 114
4.4.2 Embedded research model . . . 122
4.4.3 Proposed mechanisms of action . . . 124
4.4.4 Pilot implementation . . . 125
4.5 The HEADS-UP trial: study design and setting . . . 126
4.5.1 Setting . . . 128
4.5.2 Implementation . . . 128
4.5.3 Study population . . . 130
4.5.4 Study outcomes & hypothesis testing . . . 131
4.5.5 Sample size, power calculation and analysis . . . 132

4.5.6 Allocation and blinding . . . 135
4.5.7 Data management . . . 135
4.5.8 Ethical approval . . . 135
4.6 Discussion . . . 136

Chapter 5. PCTS: patient outcomes & processes of care 137

5.1 Introduction . . . 137
5.2 Aims . . . 138
5.3 Methods . . . 138
5.3.1 Study design & participating institutions . . . 138
5.3.2 Study outcomes and data collection . . . 139
5.3.3 Statistical analysis . . . 141
5.4 Results . . . 145
5.4.1 Intervention and control group characteristics . . . 145
5.4.2 HEADS-UP use during the study period . . . 145
5.4.3 Primary outcome: excess length of stay (eLOS) . . . 147
5.4.4 Secondary outcomes . . . 150
5.5 Discussion . . . 151
5.5.1 Summary of findings . . . 151
5.5.2 Limitations . . . 152
5.5.3 Implications for further work . . . 153
5.5.4 Conclusions . . . 153

Chapter 6. PCTS: mechanisms of action & qualitative impact 155

6.1 Introduction . . . 155
6.2 Aims . . . 156
6.3 Methods . . . 156
6.3.1 Data sources and analysis . . . 156

6.4 Ethics, consent and permissions . . . 160
6.5 Results . . . 160
6.5.1 Psychological safety, with a hard edge of accountability . . . 160
6.5.2 Rapid resolution and meaningful managerial follow-up . . . 164
6.5.3 Changing individuals: reversing normalised deviance, and catharsis . . . 167
6.5.4 Proactive teams and a revealed gap in practice . . . 169
6.5.5 Organisational change . . . 171
6.5.6 Safety & teamwork climates . . . 173
6.5.7 Submitted incident reports . . . 175
6.6 Discussion . . . 177
6.6.1 Summary of findings . . . 179
6.6.2 Limitations . . . 180
6.6.3 Implications for further work . . . 181
6.6.4 Conclusions . . . 181

Chapter 7. PCTS: an implementation evaluation 183

7.1 Introduction . . . 183
7.1.1 Characterising implementation . . . 183
7.1.2 Describing context . . . 184
7.2 Aims . . . 185
7.3 Methods . . . 185
7.3.1 Data sources . . . 185
7.3.2 Analytical approaches . . . 186
7.4 Results . . . 189
7.4.1 Local context . . . 189
7.4.2 Implementation fidelity . . . 200
7.4.3 Facilitators and barriers . . . 205

7.4.4 Other implementation outcomes . . . 212
7.5 Discussion . . . 214
7.5.1 Summary of findings . . . 215
7.5.2 Limitations . . . 217
7.5.3 Implications for future work . . . 218
7.5.4 Conclusions . . . 219

Part IV Future targets for quality improvement on medical wards 221

Chapter 8. The model of alignment 223

8.1 Introduction . . . 223
8.2 Aims . . . 224
8.3 Methods . . . 224
8.4 Results . . . 225
8.4.1 Engagement: a reciprocal commitment from staff and their organisation 225
8.4.2 Clinical engagement can be improved by co-design, local modification and strategic selection of QI interventions . . . 226
8.4.3 Beyond clinical engagement: the role of managers in QI . . . 228
8.4.4 Influencing middle managers to facilitate effective QI: status, incentives and resources . . . 230
8.4.5 A model of alignment for successful QI . . . 231
8.4.6 Applying the model of alignment to analyse HEADS-UP implementation at the primary study site . . . 236
8.5 Discussion . . . 237
8.5.1 Summary of findings . . . 237
8.5.2 The novel contributions of this model . . . 238
8.5.3 Limitations . . . 239

8.5.4 Implications for future work . . . 239
8.5.5 Conclusion . . . 240

Chapter 9. Patient-prioritised metrics 241

9.1 Introduction . . . 241
9.2 Aims . . . 243
9.3 Methods . . . 243
9.3.1 Establishing current practice in the display of performance metrics . . . 243
9.3.2 Semi-structured interviews . . . 244
9.3.3 Data analysis . . . 245
9.3.4 Sample size . . . 246
9.3.5 Ethical approval . . . 246
9.4 Results . . . 247
9.4.1 Current practice in the display of performance metrics . . . 247
9.4.2 Interview participant demographics . . . 253
9.4.3 Perceptions of quality in this setting . . . 253
9.4.4 Benefits of existing quality metrics . . . 262
9.4.5 Limitations of existing quality metrics . . . 264
9.5 Discussion . . . 271
9.5.1 Summary of findings . . . 271
9.5.2 Novel contributions of this work . . . 273
9.5.3 Study limitations . . . 274
9.5.4 Implications for future work . . . 276
9.5.5 Conclusion . . . 277

Part V Discussion 279

Chapter 10. Discussion 281

10.1 Thesis aims . . . 281
10.2 Summary of findings . . . 282
10.2.1 Aim 1: Categorise and evaluate existing improvement strategies for the medical ward . . . 282
10.2.2 Aim 2: Design, implement and evaluate a novel interdisciplinary intervention in this setting . . . 283
10.2.3 Aim 3: Identify novel improvement targets from organisational and patient perspectives . . . 285
10.3 Thesis limitations: a critique of methodological choices . . . 285
10.3.1 Literature reviews: narrative, systematic, or realist? . . . 286
10.3.2 PCTS evaluation: evaluability assessment, formative evaluation, or summative evaluation? . . . 287
10.3.3 Study outcomes: validity and accuracy . . . 290
10.4 Implications for research, clinical practice and health policy . . . 292
10.4.1 Avenues for further research in interdisciplinary ward care . . . 292
10.4.2 Implications for clinical practice and policy . . . 296
10.5 Personal reflections . . . 298
10.6 Concluding remarks . . . 298

Bibliography **301**

Appendix A. Systematic review: search terms **363**

A.1 Objective outcomes . . . 363
A.2 Program evaluations and trials . . . 364
A.3 Medical ward descriptors . . . 364

Appendix B. HEADS-UP: research and development authority approvals **367**

Appendix C. HEADS-UP: PRISMA diagram and supplementary results from generalised estimating equations models 371

Appendix D. HEADS-UP: focus group topic guide and survey instrument 375

Appendix E. HEADS-UP: a mobile application to improve automated data collection 383

Appendix F. Patient-prioritised quality metrics: interview topic guide, patient information sheets and ethical approval 391

List of Figures

1.1 Thesis structure 30

2.1 Common themes for medical ward quality & safety interventions 68

3.1 Information flow through the systematic review 82
3.2 Effect of interdisciplinary team care interventions on length of stay 97
3.3 Effect of interdisciplinary team care interventions on early readmissions . . . 99
3.4 Effect of interdisciplinary team care interventions on early mortality 101

4.1 Prospective clinical team surveillance driver diagram 116
4.2 Unintended events on medical wards 118
4.3 Early iteration of HEADS-UP pro forma 119
4.4 Late iteration of HEADS-UP pro forma 120
4.5 Interlinked intervention components 123
4.6 Stepped wedge cluster trial design 127

5.1 The relationship between engagement, briefing frequency, and the probability of excess length of stay 148

6.1 Empirical support for HEADS-UP study hypotheses 178

8.1 The model of alignment 234

9.1 Patient information board 248
9.2 Shift-by-shift staffing levels, comparing numbers required to numbers present 250
9.3 Hospital-acquired problems, displayed as the date of the last recorded event . 251
9.4 'Friends and Family Test' results displayed as a star rating 252
9.5 Thematic map for patient and carer interviews 254

10.1 A model of interdisciplinary team care effectiveness 295

C.1 Patient flow diagram for the HEADS-UP study 372

List of Tables

1.1 Organisational summary mortality statistics 24

2.1 Five strategies to improve medical ward care 39

3.1 Systematic review: PICO formulation of the research question 76
3.2 Systematic review: inclusion & exclusion criteria 77
3.3 Systematic review: characteristics of included studies 84

4.1 HEADS-UP hypotheses & their corresponding outcomes 133

5.1 Characteristics of the participating institutions 140
5.2 Coding of monthly HEADS-UP implementation fidelity 143
5.3 Ward & admission characteristics in the control and intervention groups . . . 146
5.4 Intention-to-treat and per protocol analyses for patient outcomes and processes of care (mixed-effects models) 149

6.1 Survey respondents and completion rates 174
6.2 Safety and teamwork climate scores 176

7.1 World Health Organisation implementation outcomes 188
7.2 Implementation strategies & programmatic fidelity 204

8.1 Managerial collaboration in selected QI interventions 233

C.1 Intention-to-treat and per protocol analyses for patient outcomes and processes of care (generalised estimating equations models) 373

Part I

Introduction

Chapter 1

Delivery and measurement of interdisciplinary team care in the medical ward microsystem

1.1 Introduction

The ward is the fundamental delivery unit of medical care for most hospitalised patients. It is a place of treatment, where patients receive their therapy; a temporary home, where they take their meals and receive their visitors; and - for many patients - a resting place, where they are comforted in their final days. Despite enormous advances in medical understanding, these basic premises of ward care have changed little.

The concept of how the ward environment might contribute to - or detract from - patients' health and healing has undergone several revolutions. In the 19th century, Florence Nightingale observed the advantages of natural light, ventilation and cleanliness in ward design (Hurst, 2008). Her prescription for a ward that could improve patients' outcomes was adopted globally. With few specific interventions that could actually help patients, the environment of care became disproportionately important. In the late 20th century, this primacy was largely reversed, with

increasingly effective therapies for a range of medical conditions. Clinical practice focused on the selection of evidence-based treatments for specific diseases, paying less attention to the systems of care delivering those treatments. Similarly, academic funding prioritised the identification of novel biomedical therapies, with less investment in health services research to evaluate how they were used in the real world (UK Clinical Research Collaboration, 2015).

With rapid, ongoing, therapeutic and technological advances, the sheer complexity of modern healthcare now itself poses a major challenge. It is beyond any *individual* healthcare provider to guarantee that her patients receive all the therapies from which they are likely to benefit, or to keep those patients entirely safe from iatrogenic harm. To remedy this, we are turning our attention once again - much as Nightingale did - to the local *systems* in which we work. Whereas she focused on the physical characteristics of the hospital environment, we are increasingly attuned to its processes, personalities, and communities - as well as its bricks and mortar.

This thesis focuses on the work of general medical wards. These hospital units care for adult patients whose medical treatment primarily requires drug therapy and limited procedural interventions, rather than surgery. Patients on medical wards are increasingly elderly and frail, with multiple chronic illnesses, and typically require community support from both formal and informal carers. Rapid physiological deterioration is common. Caring for these medical patients requires large interdisciplinary teams; highly individualised care planning; and difficult choices between pursuing curative treatment or palliation. This type of care represents the majority of inpatient work, accruing significant healthcare costs. Approximately £5 billion is spent on ward care each year in the National Health Service (NHS), a quarter of NHS inpatient expenditure (Monitor, 2014).

Healthcare quality and safety research has mostly taken place in the hospital setting; however, that research activity has largely focused on areas considered 'evidently high risk', i.e., surgery and anaesthesia. Dixon-Woods et al. (2009) highlight the 'relative neglect of more apparently mundane settings, such as medical wards'. In this chapter, I will set out why medical wards are worthy of study in their own right. I will describe:

(i) The evolving understanding of ward systems, and their impact on the delivery of care;

(ii) Evidence of serious failings in medical ward care; and

(iii) Characterisations of ward-level performance.

Any description of ward care must be seen in the context of the wider focus on healthcare safety and quality from the beginning of the 21st century. I will therefore review some fundamental concepts in the broader patient safety literature, before turning more specifically to medical wards.

1.2 Key concepts in patient safety and healthcare quality

1.2.1 Changing perceptions of patient safety: from isolated harms to a global public health problem

Patient safety - the absence of preventable harm to a patient during the process of health care - is now considered a serious global public health issue (World Health Organisation, 2014a). It took considerable time for the topic to capture such international attention, despite well-publicised cases where individual patients had died due to medical errors (Knox, 1995; Robins, 1995). Academic reports were similarly ineffective in generating public pressure, despite evidence that around 9% of patients were affected by adverse events[1] (Vries et al., 2008). Harms were typically understood as the result of individual clinicians' mistakes, with little appreciation of the extent of the problem, or its systemic causes (Li et al., 2015).

A seminal report by the Institute of Medicine, *'To err is human'* (Kohn et al., 2000), finally launched the modern patient safety movement. It represented a distinct change of course, with a clear message about the extent of harm - targeted, for the first time, directly to the lay public (Elwyn and Corrigan, 2005). Much attention focused on the report's estimate of up to 98,000

[1] An adverse event is an unintended injury or complication, resulting in prolonged hospital stay, disability or death, caused by healthcare rather than the underlying disease process (Vries et al., 2008).

American deaths per year from medical error, generating a momentum that led to notable increases in academic patient safety publications and research awards (Stelfox et al., 2006). The effect, in practice, was most apparent in sustained reductions in hospital-acquired infections, an important cause of harm (Pronovost et al., 2016). Progress in other areas, however, has been far more limited, with a growing understanding that most patient safety problems are more complex and pervasive than was first thought to be the case (National Patient Safety Foundation, 2015). Still, there is no longer any credible debate that patient safety is an important public health issue.

1.2.2 Systems approaches, blame and individual accountability

In addition to forcing patient safety onto the public and professional agendas, a key contribution of the Institute of Medicine's approach was its emphasis that 'errors' and 'safety' reflected systems properties, more than the competence or efforts of individual clinicians (Donaldson, 2008). Systems changes, rather than a focus on individuals, would therefore hold the key to improving safety and quality. A similar message emerged from the UK's report one year later, *'An organisation with a memory'* (Department of Health, 2000).

The need to understand systems failings as the foundational causes of most errors has now been widely accepted. However, this does not absolve individual clinicians of their responsibilities. A 'just culture' requires personal accountability for individual performance, in addition to an atmosphere of trust that encourages the recognition of systemic inadequacies (Reason, 1997). Tensions between the conflicting approaches of 'no blame' and 'accountability', and indeed between individual and collective accountability, must be balanced if patients and clinicians are to feel that they have been treated fairly (Wachter, 2013). That this nuanced debate is taking place at all is a major step forward from the automatic blame response that errors once provoked.

1.2.3 Definitions of quality in healthcare

Having established that patient safety was one component of high quality care, the Institute of Medicine then developed a more comprehensive description of the quality construct. In *'Crossing the Quality Chasm'* (Institute of Medicine, 2001), six key characteristics of quality were listed: safe, effective, patient-centred, timely, efficient and equitable.

More recent reports describe fewer quality goals. In the NHS, quality was defined as patient safety, patient experience and the effectiveness of care (Department of Health, 2008). In the USA, the Institute for Healthcare Improvement described goals for high value care in a 'Triple Aim': improving the patient experience of care, improving the health of populations, and reducing the per capita cost of care (Berwick et al., 2008). This has since been expanded to a 'Quadruple Aim', to include improving the experience of providing care, i.e., improving job satisfaction and reducing burnout for healthcare providers (Bodenheimer and Sinsky, 2014; Sikka et al., 2015).

How best to define healthcare quality clearly depends, to some extent, on the sociopolitical context in which that care is delivered. Nonetheless, these different descriptions of quality share common components, notably patient experience and an element reflecting the effectiveness of care in improving health. Patient safety may be listed separately as a component of quality, as it is in *'Crossing the Quality Chasm'* and in *'High Quality Care For All: NHS Next Stage Review'* (Department of Health, 2008). Alternatively, it might be incorporated into a broader category of patient experience (Stiefel and Nolan, 2012).

1.2.4 Changes in healthcare safety over time

There is limited evidence that the safety of healthcare has improved in recent years. Administrative data indicate that adverse drug events, catheter-associated urinary tract infections (CAUTIs), and pressure ulcers all decreased substantially in the USA between 2011 and 2014 (Agency for Healthcare Research and Quality, 2016). Record reviews show a 30% fall in preventable adverse event rates in the Netherlands between 2008 and 2011/2012 (Baines et al., 2015).

However, the large reductions in adverse event rates and complication rates were statistically insignificant. Previous work by the same group in the Netherlands had also demonstrated little progress in reducing healthcare-associated harms (Shojania and Thomas, 2013).

Paradoxically, this might reflect rising standards of care. Over time, as more complications of healthcare are brought within an expanded 'perimeter of safety' - viewed as unacceptable, potentially preventable patient safety issues - significant reductions in adverse event rates are increasingly difficult to achieve (Vincent and Amalberti, 2015). The clear exceptions to this rule have been healthcare-associated infections, particularly central line-associated bloodstream infections (CLABSIs). Once accepted as the inevitable sequelae of intravenous devices, CLABSIs are now the target of well-defined sociotechnical efforts to prevent them (Pronovost et al., 2006; Dixon-Woods et al., 2011; Marsteller et al., 2012). Successes in CLABSI reduction are the clearest possible demonstration of improved patient safety in recent years.

1.2.5 Quality improvement

Quality improvement (QI) interventions, broadly defined as purposeful efforts to secure positive change (Portela et al., 2015), can be broadly divided into two conceptually different approaches. The first approach targets some underpinning aspect of care - for example, team structures or routines, non-technical skills, or generic operational challenges faced by staff. The second aims to improve the management of a *specific* clinical condition (a diagnosis, or a complication of that diagnosis). Here, I will discuss the relative attention afforded to each approach, and its limitations.

1.2.5.1 Improving underpinning aspects of care

Healthcare delivery is not streamlined. Operational inefficiencies are commonplace: indeed, process failures are an integral part of frontline healthcare delivery (Tucker and Edmondson, 2003). Nurses spend 10% of their time simply working around 'routine' problems: disruptions and errors in the supply of information or materials are an expected part of their work (Tucker,

2004). These problems do not attract significant attention, either in national safety initiatives or within individual institutions (Tucker et al., 2008; Singer et al., 2009).

Few large-scale improvement interventions have formally addressed these characteristics of ward care. One notable exception was the *'Releasing Time to Care: The Productive Ward™ (PW)'* programme. Introduced in the UK in 2007 by the NHS Institute for Innovation and Improvement, PW was a modular, self-directed toolkit. Its goal was to 'empower ward teams to identify areas for improvement, by giving staff the information, skills and time they need to regain control of their ward and the care they provide' (Kings College London, 2009). Based on the principles of the Toyota Production System, also known as 'Lean thinking' (Womack et al., 1991), PW claimed to reduce ward activities that did not add value, releasing staff time for work that actually met patient needs.

The idea of a generic, cross-cutting, ward-level intervention was distinctly different to the initiatives that had preceded it, as PW leaders described (Kings College London, 2009):

> 'The ward environment is somewhere that in service improvement sort of initiatives has actually been largely ignored, because they've gone down pathways, they've gone down specific improvements in terms of clinical issues... but never looking at the mundane processes that make up the [ward's] activities... that really hugely impact on... patient care and patient experience.' (Lizzie Cunningham, NHS Institute)

Despite a £50 million investment in its nationwide dissemination, significant voluntary local uptake (Kings College London, 2009), and international adoption (Broek et al., 2014; Hamilton et al., 2014; White et al., 2014), there is little evidence of PW's effectiveness. A systematic review of PW literature found numerous 'experiential' publications extolling its benefits, with few negative reports, and little other substantive data. This suggests publication bias (Wright and McSherry, 2013); only one report has been published independent of the NHS Institute for Innovation and Improvement, which developed the programme (Wright and McSherry, 2014).

PW's key claim - to release time for nurses to spend with their patients - remains largely unproven. A review of possible efficiency improvements associated with PW found that, where nursing time *had* been freed up, the majority of it was then reinvested in completing more PW modules (Kings College London, 2011). At the time of writing, efforts are underway to re-evaluate PW, ten years after its original introduction (National Institute for Health Research, 2016).

A recent systematic review of other 'Lean' interventions in healthcare similarly found no evidence of positive effects on health outcomes, patient satisfaction or actual financial costs. The effect on process outcomes was inconclusive (Moraros et al., 2016). However, the evidence was of poor quality, and it may be that Lean interventions focused on waste from the patient's perspective (rather than the provider's) are more likely to improve outcomes (Blackmore and Kaplan, 2016).

At best, we could say that the known effect of these interventions is unclear. Tucker et al. (2008) argued that improving work systems in general, rather than focusing more narrowly on specific clinical conditions, might increase both the safety and efficiency of ward care, and improve patient experience. This is no simple task: cross-cutting interventions (whether they focus on operational issues, transfers of information, teamwork or other non-technical skills) are not easily implemented [see 1.4.1.2].

1.2.5.2 Improving the management of specific clinical conditions

The majority of QI interventions adopt the second conceptual approach, addressing the management (prevention, diagnosis and treatment) of specific clinical conditions. This is deliberate: improvement teams are typically urged to choose goals that are specific, measurable, achievable, realistic and timely (Doran, 1981). For example, the Dutch national safety programme *'Prevent Harm, Work Safely'* included improvement modules on optimising care for acute coronary syndromes, the prevention of renal failure due to iodinated contrast agents, and the prevention and treatment of severe sepsis, amongst others (Baines et al., 2015).

Whether these discrete, single-topic QI efforts make a substantial contribution to overall levels of healthcare quality and safety is far from certain. As these modular improvement strategies accumulate, so do their unintended consequences, including an unmanageable proliferation of performance metrics [see 1.5.3.4], and competition for limited staff time and resources. There is also the perception that single-issue QI does little to ease the complex, collaborative care that staff must routinely deliver. This has been likened to 'killing mosquitoes one at a time, but not draining the swamp' (Dixon-Woods, 2014b).

The most definitive improvement in any single group of clinical conditions has been the reduction in CLABSIs [see 1.2.4]. However, focal successes in CLABSI reduction are not representative of the QI field in general, which has not achieved similarly broad reductions in other patient harms. Subsequent single-issue QI initiatives have been nowhere near as impressive. This may be because CLABSIs are not representative of the common challenges that wards face [see 1.5.3]. Moreover, sustained CLABSI improvement was brought about when synergistic regulatory and social pressures aligned with technical advances in infection prevention. Such alignment is extremely rare in QI, which is more often faced with competing priorities, contradictory policies, and interventions whose effects are contested. Newer QI programmes now face a very different cultural context to the early CLABSI initiatives, which had a clearer field (Dixon-Woods et al., 2013).

CLABSI reduction is, in some ways, a cautionary tale. Meaningful improvement on this single topic required a powerful combination of sustained pressures, acting at micro-, meso- and macro-organisational levels. With so many overt and subtle effects, intended or otherwise, the individual contribution of any single QI initiative becomes very hard to determine.

1.2.6 Evaluating quality improvement interventions

The QI literature would suggest that the majority of QI interventions have been effective. However, this may reflect considerable publication bias (the preferential selection of reports with positive results) and the choice of inadequate study designs (Shojania et al., 2004). Improvement

interventions can be assessed with a range of study types, reflecting different trade-offs between research rigour and pragmatic, dynamic interventions (Portela et al., 2015). Each methodology has its particular strengths and weaknesses, and complementary quantitative and qualitative evaluations may be needed to assess what works, how, and in what contexts (Dixon-Woods, 2014a).

The key issue for QI studies, whatever their design, is causation: did a particular intervention lead to improvement *beyond what could have been expected otherwise*? Every evaluation needs to answer this question. Randomised controlled trials (RCTs) and their variants are considered the best designs for these inferences about causality, even if they too have drawbacks (Portela et al., 2015). However, the predominance of uncontrolled before-after studies is a bigger concern. These studies often falsely attribute improvements in care to an intervention, when the same change would have been seen regardless (Effective Practice and Organisation of Care, 2014).

QI interventions are bedevilled by a phenomenon known as the 'rising tide'. This is a positive secular trend (widespread, incremental improvements in the quality and safety of care), occurring contemporaneously with the QI evaluation (Chen et al., 2016). Importantly, rising tides are the result of the same public and professional pressures that lead to the development of QI interventions: their coexistence is no coincidence. Several major QI studies reporting null results have identified and acknowledged the influence of these secular trends (Benning et al., 2011; Bion et al., 2013). The majority of the QI literature, however, has not been as rigorous. The extent to which QI interventions can be relied upon as a delivery vehicle for improved care standards - as opposed to regulation, professional incentives, or structural investment - remains unknown.

The prioritisation of patient safety and high quality care, the understanding of failings as predominantly systems deficiencies, and the contested value of quality measurement and improvement are important themes throughout this thesis. I will now return to the main thrust of this chapter: describing medical ward systems and their quality of care.

1.3 Medical ward design and care delivery

1.3.1 Ward design

Hospital ward design has evolved in three distinct stages. Nightingale wards, featuring a long nave in which all patients could be easily observed from a nurses' station, were the design of choice from the mid-19th century until the 1950s. Later 'bay' designs, separating wards into numerous rooms that each housed a small number of beds, allowed for more flexible use of the available accommodation. Bays could then be designated as 'male' or 'female' as required (NHS Estates, 2002; Hurst, 2008).

Importantly, this design conceded for the first time that other priorities (here, an appropriate organisational response to fluctuating demand) had to be balanced against the ease of nursing observation. This concession paved the way for the increased prioritisation of patient privacy and dignity, with a third generation of hospital wards expected to routinely provide single rooms (Department of Health, 2011).

There are clear trade-offs between the different ward designs. For example, improved patient privacy may come at the expense of reduced supervision by staff, negative effects on staff teamwork, and reduced social interaction for patients themselves (Maben et al., 2015). Even apparently simple changes to the ward environment (however desirable in theory) may have complex, even harmful, effects on the nature of ward work (Wachter, 2015). Systems leaders have focused on the assumed technical benefits of their proposed changes, with less attention to the socioadaptive adjustments that these changes demand. A more holistic paradigm for the ward environment - the microsystem - can be adapted from the literature on organisational effectiveness.

1.3.2 The medical ward as a clinical microsystem

The performance of any organisation's core activities relies on its microsystems, the small frontline units that deliver its services. High-performing institutions outside of healthcare recognise that optimising microsystem performance is fundamental to the wider organisation's success (Quinn, 1992). For example, they continually re-engineer the frontline interface between the customer and the institution's core competency, and develop systems to capture microsystem information. Here, I will describe the distinct characteristics of medical wards as clinical microsystems, and identify demographic changes that will increase both their influence and workload.

1.3.2.1 Unit-level delivery of interdisciplinary care & production of the outcomes of care

Health systems in turn are made up of multiple clinical microsystems. These are functional interdisciplinary units, delivering care to particular groups of patients, with shared clinical and business aims, linked processes, and a common information environment (Nelson et al., 2002). These clinical microsystems - rather than the wider organisations to which they belong - produce clinical and experiential outcomes, on which the quality of care is ultimately judged.

In this thesis, I describe the work of medical ward microsystems as 'interdisciplinary', following an established typology (Nancarrow et al., 2012) and consistent use of this term in previous research (O'Leary et al., 2012a). Interdisciplinary work sits on a spectrum of integration between 'multidisciplinary', where individual contributions are made separately, and 'transdisciplinary', where workload is so integrated that disciplinary boundaries are at least partly dissolved (Hibbert et al., 1994; Thylefors et al., 2005):

> 'Outcomes [of interdisciplinary work] can only be accomplished through the interactive effort and contribution of the disciplines involved; this implies a high level of communication, mutual planning, collective decisions and shared responsibilities.

These independent contributions have to be coordinated.' (Nancarrow et al., 2012)

Medical ward interdisciplinary microsystems are then distinguished from other microsystems by the nature, and variety, of their workload.

1.3.2.2 The importance of interpersonal networks, information flow & uncertainty

Medical wards are an archetypal clinical microsystem, with distinct expertise, team members, and budgets. In caring for a particularly heterogeneous group of patients, however, these microsystems are defined more by their interpersonal networks and information flow than they are by any specific clinical or business aims. This sets them apart from other microsystems, which typically have clearer clinical goals, arising from a more narrowly defined patient population.

Two examples highlight this difference more clearly. A haemodialysis unit provides a single treatment to patients with end-stage renal failure: its work can largely be described with reference to the people, skills and equipment required to provide and supervise this single therapy. Surgical wards, even with a broader patient group to serve, also have their own unifying reference point: the surgical operation. Although the indications for each patient's operation, and indeed the details of each procedure, will clearly vary, surgical patients must be prepared for the experience in a similar way, and will need access to similar pathways in the post-operative period to facilitate their recovery. Common post-operative complications can be anticipated and managed in a standardised fashion (Pucher et al., 2014). Surgical wards, therefore, are defined in terms of their relationship with the operating theatre. Their goals revolve around the appropriate selection of patients for surgery, the timely delivery of safe and effective surgery, and patients' subsequent rehabilitation.

The key characteristic of a general medical ward, by contrast, is the need to manage patients for whom there is no single best pathway of diagnosis or treatment. Patients often present with non-specific symptoms, which may or may not represent an exacerbation of one of their many pre-existing conditions. Empirical treatment can be started, and concluded, with no definitive diagnosis ever established. Perhaps more than any other inpatient clinical microsystem, general

medical wards trade in variety and uncertainty. As institutions seek to establish alternatives to hospital admission[2], only those patients requiring substantial nursing assistance, intense medical supervision, or both, will meet the high threshold for entry. The complexity of care delivered in the ward environment will continue to increase, as these demographic changes take effect.

Recognising this as a key challenge, alongside the management of multimorbidity, has two important implications. First, it suggests that interventions to help interdisciplinary teams address the complexity of their workload may be more beneficial than disease-specific QI efforts [see 1.2.5.1]. Second, other inpatient microsystems - that have traditionally cared for patients with more singular medical needs - may need to incorporate elements of medical ward practice, as their own patient demographics become more complex.

1.3.2.3 An influential inpatient microsystem

Medical team practice is increasingly influential in other ways. Outside of the operating theatre or intensive care unit, leadership of care is increasingly deferred to medical teams. In some institutions this is a deliberate and positive change, with hospitalists (specialist inpatient physicians) invited to manage or co-manage the patients admitted to surgical services (Auerbach et al., 2010). For others, it is a more pragmatic phenomenon, often reflecting the immediate availability of on-site medical expertise with limited access to an equivalently senior surgeon (Royal College of Physicians, 2013).

The effective expansion of medical ward practice has important resource implications, to which restrictive organisational admission policies, and the need to assist non-medical teams, both contribute. In the UK, there have been only limited efforts to rigorously reassess the workflow of inpatient medical teams, although a programme to assess the impact of consultant-led acute care is now in place (National Institute for Health Research, 2014). The need for this work is borne out by empirical evidence of poorly understood ward changes (Maben et al.,

[2]For example, urgent care centres, ambulatory care centres, and hospital at home schemes make possible increasingly intensive medical care delivery, without the patient requiring admission to a hospital bed.

2015), widespread reports of workplace dissatisfaction (Royal College of Physicians, 2013), and exemplar evaluations of service developments elsewhere (O'Leary et al., 2015b; Bilimoria et al., 2016; University of Pennsylvania, 2016).

It is unsurprising, given these pressures, that medical teams might be unable to consistently deliver high quality care. There is evidence that this is the case, with standards of ward care falling below an acceptable threshold, causing harm.

1.4 Serious failings in medical ward care

Medical ward research is complicated by the same factors that make care delivery difficult: long admissions, with frequent transfers of information (handovers); large, fluid and geographically dispersed interdisciplinary teams; and patient heterogeneity [box 1.1]. Consequently, we know more about optimising team practice in the operating theatre and intensive care unit (ICU), which are more stable environments, with better defined episodes of care.

Box 1.1 Specific challenges for medical ward practice & research

Long admissions with frequent information transfers

Amorphous, poorly defined episodes of care, commonly involving multiple care teams

Large, fluid & geographically dispersed interdisciplinary teams

Significant patient heterogeneity

Despite the widespread perception that ICU patients are at greater risk of a serious error, medical wards are equally hazardous - perhaps because there is a better understanding of how to prevent patient safety incidents in the ICU (Umscheid and Brennan, 2015). Medical wards and ICUs expose similar proportions of their patients to preventable adverse events (10.6% *vs* 10.9%), and at similar rates (1.0 *vs* 1.4 preventable adverse events/100 patient days) (Forster et al., 2011). A higher percentage of these medical ward adverse events are fatal (2.3% *vs* 1.4%). Given the

dispersion of the team, responsibility for these ward harms is then spread diffusely, deterring action. Suboptimal service delivery has been identified in both routine and emergency ward care.

1.4.1 Problems in routine care

1.4.1.1 Opportunities to provide needed care are frequently missed

Medical wards are low reliability microsystems, failing to deliver interventions of proven benefit. In routine ward care, opportunities to provide ideal interventions are taken up less than 75% of the time (Huang et al., 2008). Fewer than 50% of the missed opportunities are justified in the medical record. For treatment of ischaemic heart disease (a common reason for admission to medical wards), ideal intervention use and justified non-use may each be as low as 55% (Huang et al., 2008). Such poor performance may arise from teams' working styles. Ward teams can be inefficient, with providers working alone but assuming that their work will be coordinated with others' (Bharwani et al., 2012). In doing so, staff form working *groups*, rather than working *teams*. Any complacency that one's colleagues will anticipate and remedy care problems is unwarranted: 62% of care problems reach the patient, and 10% of those cause physical injury (Lubberding et al., 2011). With medical wards depending on high-performing interdisciplinary teams to manage uncertainty [see 1.3.2.2], the root causes of their reported problems relate to human behaviours (particularly behaviours around situational assessment), more so than in surgery or emergency medicine (Wagner et al., 2016).

1.4.1.2 Errors of omission are leading causes of preventable patient deaths

Missed opportunities to provide needed care are important, and not only as *prima facie* evidence of medical wards' systems failures. In this setting, errors of omission are much more likely than errors of commission[3] - and are far more likely to be serious errors at that (Hayward

[3] An error of omission is a medical error involving too little treatment (underuse). Errors of omission include diagnostic delays, subtherapeutic medication doses, and failures to provide indicated treatments. Errors of commission involve too much treatment, or hazardous treatment (overuse and misuse, respectively). Examples include excessive

et al., 2005). In a major retrospective case record review study, omission errors (poor clinical monitoring, diagnostic errors and inadequate drug or fluid management) were the principal problems associated with preventable inpatient deaths (Hogan et al., 2012). In the same study, general ward care was the phase of care in which problems contributing to death occurred most often. Suboptimal ward care led to 44% of preventable deaths; problems in postoperative and procedural care (including ICU care) accounted for only 8.3% (Hogan et al., 2012).

We see, then, that errors of omission on medical wards are common, serious, and leading causes of preventable deaths in hospital. A change in terminology - for example, using 'problem in care' rather than the more commonly used 'adverse event' - may be required to stop omission errors from being overlooked in record reviews (Hogan et al., 2012). Even then, errors of omission are judged more leniently than equally harmful errors of action - a phenomenon termed 'omission bias' (Spranca et al., 1991; Baron and Ritov, 2004) - and their causes are extremely heterogeneous, resisting simple solutions (Wong et al., 2015). This is a major challenge. Multiple, targeted, modular improvement interventions will be required to address them, each requiring substantial effort. The alternatives, cross-cutting interventions targeting non-technical skills such as teamwork or communication, are not easy substitutes. Team training and handover bundles have been effective, but require prolonged, multi-institutional, collaborative efforts (Neily et al., 2010; Starmer et al., 2014). The sustained commitment required for either of these options may be a daunting prospect for organisations struggling to balance their existing clinical and financial commitments.

1.4.2 Problems in emergency care

1.4.2.1 Five key failings in ward care for the acutely ill patient

An influential report by McQuillan et al. (1998) identified serious flaws in the delivery of emergency care on UK medical wards. Reviewing the ward treatment of 100 patients[4] immediately be-

medication doses, treatments that are contraindicated, or unnecessary interventions (Hayward et al., 2005).

[4] 51/100 of these patients were general medical patients.

fore their emergency admission to intensive care, McQuillan found at least 54 who had received suboptimal care, contributing to their morbidity or mortality. Five key failings in suboptimal ward care were identified:

- Lack of knowledge;
- Failure to appreciate clinical urgency;
- Failure to seek advice;
- Lack of supervision; and
- Failure of organisation.

McQuillan's five elements of suboptimal care were subsequently re-examined, in a review of the ward management of acutely ill patients. Massey et al. (2009) argued that the failings listed by McQuillan could be generalised beyond his case note review, and that they provided a useful frame on which to analyse future work. Modern training curricula would certainly recognise in McQuillan's report indirect descriptions of both technical skills (lack of knowledge) and non-technical skills (failures to appreciate clinical urgency and to seek advice). The problems he described, spanning institutional levels from the individual clinician to the wider organisation, also reflect the move to view errors through a system-wide lens [see section 1.2.2].

1.4.2.2 Competing priorities, sociocultural barriers and poor workforce planning perpetuate these failings

McQuillan's five failings remain relevant primarily because, to varying degrees, they all persist. Despite efforts to resolve them [see chapter 2], each failing is still recognisable in modern ward practice. There may be different reasons why each one has proven so refractory. For example, recurrent technical failures - such as the lack of clinical observations at the appropriate times - are common, despite automating technology. This might reflect the impact of other priorities competing for staff attention (Hands et al., 2013). Resource constraints (staffing ratios and

staff workload) might also determine clinical supervision and organisational failure (Massey et al., 2009). Importantly, these factors are impossible for individual UK hospitals to address independently, requiring regional coordination, but worforce planning has been neglected at a national level (Health Foundation, 2016).

Efforts to detect and treat clinical deterioration, on the other hand, have been limited more by the sociocultural barriers within teams (Shearer et al., 2012). These barriers complicate the escalation of care from one professional to the next. Indeed, nurses still have to consciously use specific referral 'strategies' to successfully obtain medical advice for their patients (Andrews and Waterman, 2005; Cox et al., 2006).

In summary, medical ward care failings are common and potentially serious, in both routine and emergency care. Their persistence may be due to psychological biases, complexity, workload, poor workforce planning, and sociocultural barriers. Improving the organisation of medical ward care has now become an international priority (American Hospital Association, 2013; Future Hospital Commission, 2013).

1.5 Understanding ward-level performance

We see from the previous sections that successful businesses in other industries focus on their own frontline microsystems; that medical wards are important, strained, microsystems in the inpatient setting; and that they can have significant failings. One possibility is that these failings simply reflect broader organisational struggles, i.e., that there are under-performing hospitals, rather than simply under-performing wards. In this section, I will describe evidence that important variations in the delivery of care are indeed found at the ward level, and that a focus on aggregated hospital-wide metrics is misguided. I set out criteria for performance measures that might better reflect ward-level quality of care.

1.5.1 Ward-level variation *vs* hospital-level variation

Even neighbouring wards within a single hospital can be very different. Holistic reviews of failing institutions describe islands of excellent practice coexisting with egregious deficiencies nearby (Bristol Royal Infirmary Inquiry, 2001; Francis, 2010). The differences within an institution may prove more important as targets for improvement than the contrasts between different hospitals.

Ward-level variation is evident in data describing both safety climate and adverse events. Safety climate is a superficial 'snapshot' of the safety culture within a work group or organisation (Flin et al., 2000; Flin et al., 2006), although the two terms are often used interchangeably [see box 1.2]. Safety climate scores vary more between the clinical microsystems of a single hospital than they do between hospitals (Pronovost and Sexton, 2005; Schwendimann et al., 2013), and individual wards contribute more to the variance in teamwork and safety climates than their wider departments (Deilkas and Hofoss, 2010).

Box 1.2 Definitions of safety culture & safety climate

Safety culture is 'the product of individual and group values, attitudes, perceptions, competencies and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's safety management' (Advisory Committee for the Safety of Nuclear Installations, 1993).

Safety climate can be regarded as the measurable surface features of the underlying safety culture (Flin et al., 2000).

Similarly, adverse event rates vary more *within* hospitals than they do *between* hospitals. Baines et al. (2015) calculated adverse event intraclass correlation coefficients (ICCs), to indicate the relative influence of each organisational level on the variance of the outcome. Department-level ICCs were consistently higher than those at the hospital level: departments had a greater influence over adverse event rates than the hospital to which they belonged. In addition, inter-hospital differences have reduced: hospital ICCs fell from 6.5% to 2.6% between 2004 and 2012.

Departmental level ICCs remained consistently higher (e.g., 9.7% in 2012), and were stable over the same time period.

These data would suggest that our improvement efforts should focus on frontline units: data capture and quality improvement interventions should both target this organisational level (Deilkas and Hofoss, 2010; Schwendimann et al., 2013). As we will now see, policymakers and healthcare regulators often take a different approach, using hospital-aggregated statistics to inform organisational performance management - despite concerns about those measures' validity and cost.

1.5.2 Aggregate hospital-wide metrics

The quality of inpatient care is most often judged on macrosystem data, aggregating together the processes or outcomes of numerous different units to produce an organisational summary. Outcomes are risk adjusted with routine administrative data to produce a single summary statistic, which claims to describe the comparative performance of the organisation. Various metrics have applied this methodology, including the Hospital Standardised Mortality Ratio (HSMR), and more recently the Summary Hospital-level Mortality Indicator (SHMI). In each case, a case-mix adjusted ratio of observed to expected deaths is produced; the number of expected deaths is derived from the mean of all the ratios in the sample, which changes from year to year. Although conceptually similar, the HSMR and SHMI have important differences [table 1.1].

1.5.2.1 Widespread use of summary hospital statistics for regulation and policy

Healthcare regulators and policymakers rely heavily on these hospital-wide measures, and the organisational ramifications of apparent underperformance can be severe. For example, in the wake of a report into secondary care failings in the UK (Francis, 2013), high HSMRs and SHMIs were used to select hospital trusts for intensive investigation (Keogh, 2013). Eleven of the fourteen trusts selected in this way were then placed into 'special measures', with executive-level leadership changes, suspension of financial privileges, and appointment of an external

Table 1.1: Organisational summary mortality statistics (Health and Social Care Information Centre, 2016b)

	HSMR	SHMI
Included deaths	In-hospital deaths	In-hospital deaths & deaths within 30 days of discharge
Selection of in-hospital deaths	Deaths from 56 diagnostic groups, accounting for around 80% of in-hospital deaths (depending on the services provided by the hospital)	All in-hospital deaths
Attribution of death	Death can be assigned to more than one hospital, depending on timing of patient transfer	Death is assigned to the last acute or secondary care provider
Adjustment for palliative care coding	Yes	No
Adjustment for social deprivation	Yes	No
Health care providers	All providers	Providers other than specialist hospitals

Abbreviations: HSMR - Hospital Standardised Mortality Ratio; SHMI - Summary Hospital-level Mortality Indicator.

director for improvement oversight (NHS England, 2016b).

As the author of the subsequent report acknowledged, the two aggregate mortality measures 'generated two completely different lists of outlier trusts [to review]' (Keogh, 2013). Nonetheless, the Care Quality Commission (CQC) - the regulator of health and social care services in the UK - continues to use HSMR and SHMI to inform its hospital inspections. Both measures are incorporated into its 'Intelligent Monitoring' tool, the accuracy of which has been questioned (Care Quality Commission, 2015; Griffiths et al., 2016a).

1.5.2.2 Concerns about validity and cost

Numerous international experts have raised concerns about the dubious value of these statistical constructs, which have never been validated either as safety indicators themselves or as screening tools for safety problems (Lilford et al., 2004; Shojania and Forster, 2008; Lilford and Pronovost, 2010; Scott et al., 2011; Gestel et al., 2012; Girling et al., 2012; Hogan et al., 2014; Hogan et al., 2015). Nor do these standardised ratios provide any actionable information: comprehensive chart reviews are then required to determine what problems exist (if any) and what needs to be done to resolve them. These reviews carry significant costs, which should be included when estimating the overall costs of detecting a problem with standardised outcome ratios (Shojania and Forster, 2008). There are also opportunity costs: the longstanding preoccupation with aggregated organisational statistics has precluded any meaningful investment in ward-level metrics.

We see from this that the key justification of hospital-wide statistics - that they meaningfully inform judgements about healthcare quality - is highly contentious. They are neither accurate nor cheap, nor do they facilitate quality improvement. Their continued use, instead of developing ward-level data capture, is likely to be a false economy.

1.5.3 Measuring quality in ward care

If aggregated hospital-wide statistics fail to describe quality of care, how best can wards judge their work? Operationalising quality - translating it into measurable variables - is difficult, even though high standards of efficient, patient-centred care are conceptually easy to value (Mosher and Kaboli, 2015). There are several key challenges. Performance measures need to be sensitive to changes in care; attributable to the team responsible for care; reflect the complex interdisciplinary nature of ward work; and ultimately be consistent with a burden of reporting that is manageable in practice. Each of these criteria is discussed in turn.

1.5.3.1 Sensitivity to changes in care

Many outcomes (e.g., 30-day readmissions) are determined primarily by the characteristics of the local patient population, and the resources of the local community, not by the quality of care provided during the inpatient admission (Joynt and Jha, 2012). Medical patients' outcomes, in particular, are dependent on their comorbidities and socioeconomic status. They differ in this respect from surgical patients, for whom an adverse outcome is more representative of the quality of inpatient care (Tsai et al., 2013).

Performance metrics should not penalise medical wards for factors beyond their control; nor should they fail to recognise tangible improvements in care. Participants in previous ward QI initiatives have already voiced concerns that typical performance metrics would not reflect the changes they were seeing at ward level (Kings College London, 2009). There may be parallel concerns that ward-level data are inadequately powered to detect changes in the quality of care. These latter concerns can largely be addressed through the use of statistical process control and interrupted time series techniques (Fretheim and Tomic, 2015). Repeated small samples can also be adequately powered for the purposes of QI interventions (Etchells et al., 2016).

1.5.3.2 Attribution to the responsible care team

Medical patients are typically cared for by numerous teams, who are in continuous transition throughout lengthy admissions. The overall standard of care is dependent not only on team members' actions during a given shift, but on the quality of handover between shifts. Unlike surgical patients, whose care in the operating theatre and in the immediate aftermath is definitive, medical patients have no single climactic episode of care or intervention around which there is professional 'ownership' and responsibility.

Performance measures must be appropriately assigned and then fed back to the responsible care team. This may require careful negotiation to establish boundaries of responsibility, especially in situations where patients are routinely transferred between wards for ongoing care. This is commonplace in the UK, as patients are stabilised first in an acute admissions unit (AAU) for around 48 hours, before moving to a downstream general or specialty ward. Where a performance measure may reflect the work of more than one team, those teams will require support to define their contribution to the overall outcome, and to collaborate on its improvement.

1.5.3.3 Reflecting complexity of care

As we have seen, medical wards are better described by the variety of their workload, and their extensive interdisciplinary networks, than by any single pathology or process [see 1.3.2.2]. How teams perform against these subtle challenges is difficult to measure. A common approach is to assess the quality of care for selected 'tracer' conditions, to represent the broader performance of the ward[5] (Kessner et al., 1973). Acute myocardial infarction, congestive heart failure and pneumonia have been used extensively as tracer conditions for hospital-based medical care (Wang et al., 2014; Venkatesan et al., 2016). Chronic obstructive pulmonary disease, stroke and venous thromboembolism have been included as tracer conditions in more recent US analyses

[5]Tracer conditions are common, well-defined conditions affected by medical care, with agreed appropriate care and a known epidemiology (Neuhauser, 2004).

(Centers for Medicare and Medicaid Services, 2016a).

Specific process and outcome measures for these tracer conditions are straightforward to define (and collect). However, a focus on specific elements of care for tracer conditions can inappropriately skew priorities and promote unnecessary treatment (Wachter et al., 2008). The prioritisation of tracer conditions - with a mandate to publicly report relevant process measures and outcomes - may be ineffective (DeVore et al., 2016; Joynt et al., 2016), and comes with opportunity costs. Disease-specific quality metrics do little to incentivise robust systems of care that protect complex medical patients from harm. Rather, their high profile selection implies that high quality care in this setting is best achieved by focusing on individual pathologies, one at a time.

Instead of measuring clinical processes and outcomes, it may be more effective to incentivise structures that facilitate implementation of best practice in a complex environment. NHS England recently issued an updated set of ten overarching clinical standards for acute care (NHS England, 2016a). These standards set out a different stall, requiring (amongst other things) shared decision making; early multi-disciplinary assessment; multi-professional handovers; repeated senior review; and attention to patient experience, safety and clinical effectiveness. This is a very different set of aspirations, better aligned with the needs of a complex patient population, yet still backed by a series of levers and incentives (NHS England, 2013). The two approaches need not be exclusive.

1.5.3.4 The burden of quality reporting

It is clear that performance measurement should be standardised. Inconsistent outcome measures and definitions are currently used by various provider organisations, specialty societies, countries and clinicians, dramatically slowing progress towards improvement (Porter et al., 2016). Greater consistency of outcome measurement is now a key goal for initiatives in both academic trials (Koroshetz, 2015) and routine practice (Porter et al., 2016).

However, these initiatives continue to focus on individual pathologies, with little consid-

eration of complex patients with multimorbidity. They are likely to add to the proliferation of performance data, which already far outstrip any organisation's capacity to use them. This is not a trivial concern: in addition to wasted staff time, reporting of quality metrics costs health systems billions of dollars each year (Casalino et al., 2016).

Leading experts in QI now openly call for the burden of quality measurement to be reduced - and quickly (Berwick, 2016; Wachter, 2016). Ward quality measures should be assessed not only for their sensitivity to change, attributability and reflection of complexity, but their contribution to the overall burden of reporting. Existing metrics and new metrics should both be reassessed with this in mind.

1.6 Aims of the thesis

The overall goal of this thesis is to explore the organisation of care on the general medical ward, and identify effective strategies for its improvement.

The specific thesis aims are:

1. To categorise and evaluate existing improvement strategies for the medical ward (Part II: chapters 2-3);

2. To design, implement and evaluate a novel interdisciplinary intervention to improve medical ward care (Part III: chapters 4-7); and

3. To identify new medical ward improvement targets from organisational and patient perspectives (Part IV: chapters 8-9).

A schematic outline of the thesis is shown in figure 1.1. In Part II, I review the existing evidence base for strategies to improve medical ward care. Chapter 2 is a narrative review, synthesising a disparate ward literature to define unifying themes of improvement attempts. Chapter 3 then explores one of these themes in more detail, in a systematic review evaluating the effects of interdisciplinary team care interventions on objective patient outcomes.

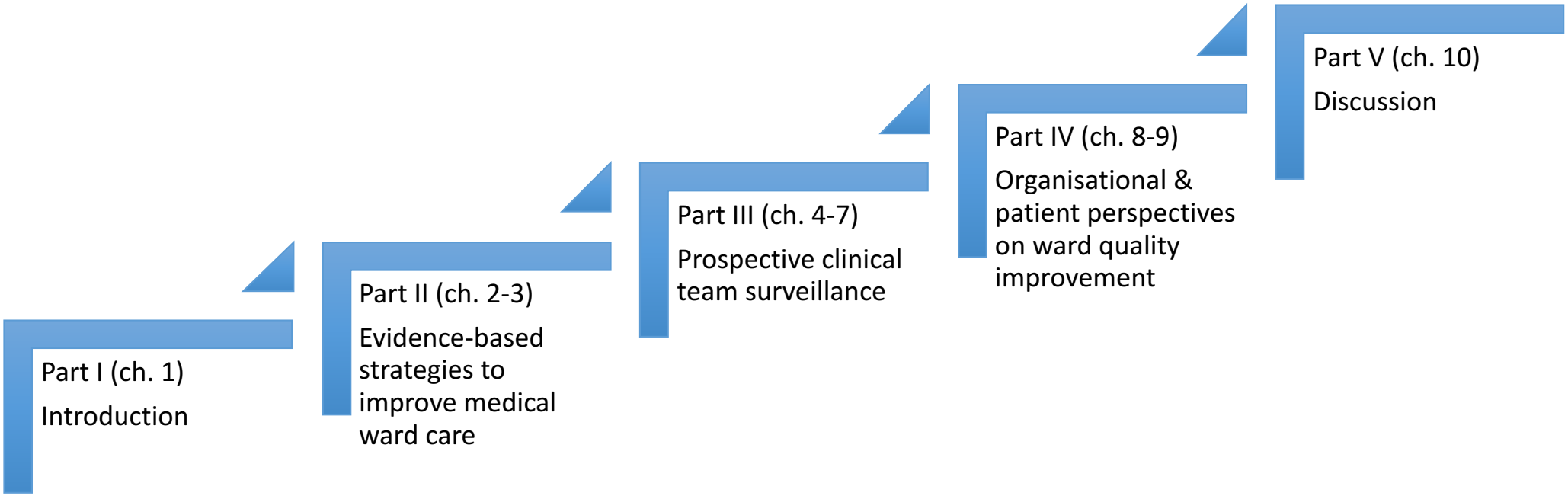


Figure 1.1: Thesis structure

The strengths and weaknesses of the studies identified in these reviews inform Part III. Here, I describe the development and mixed methods evaluation of a novel interdisciplinary team care intervention: prospective clinical team surveillance (PCTS). The design of the intervention (chapter 4), and its quantitative and qualitative evaluations (chapters 5 and 6, respectively) attempt to build on earlier work, and address some of its methodological shortcomings. An implementation evaluation in chapter 7 describes the implementation outcomes, facilitators and barriers for PCTS on medical wards, to better understand its adoption and spread in practice.

Part IV explores other perspectives on ward quality improvement. Chapter 8 uses the example of PCTS as empirical evidence to test a new model of organisational alignment for successful quality improvement. Chapter 9 then describes an interview study with patients and their carers, exploring their views on medical ward quality and its measurement. Part V (chapter 10) concludes the thesis, drawing together a discussion of its findings and limitations, as well as its implications for practice and future research. The main contributions of the thesis are then summarised, alongside my personal reflections on this work.

Part II

Evidence-based strategies to improve medical ward care

Chapter 2

Improving medical ward care: a narrative review

2.1 Introduction

In chapter 1, I described the predominance of single-issue QI initiatives in the ward environment, targeting the management of specific conditions. Less attention has been devoted to the cross-cutting factors that underpin ward care more generically. Relevant publications have aggregated into siloed literature streams, with no unifying overview of the evidence base for how best to organise, or improve, a general medicine service.

As a result, few clinicians are aware of the developing evidence for service delivery in this setting. Medical ward services evolve organically and idiosyncratically, rather than through rational design. Their organisation, particularly in the UK, has been determined by custom (the routines of more senior clinicians) as much as it has by resource constraints. There is a clear need for a synthesis of the disparate literature, both from an academic perspective and to inform frontline care delivery.

The work described in this chapter has been published as: Pannick S, Beveridge I, Wachter RM, Sevdalis N. Improving the quality and safety of care on the medical ward: A review and synthesis of the evidence base. Eur J Intern Med. 2014; 25(10):874-87. DOI: 10.1016/j.ejim.2014.10.013

Here, I conduct a narrative review of the existing literature on quality and safety in the medical ward setting. Kastner *et al* describe this methodology:

> '[A narrative review] produces an account of the evidence with commentary and interpretation. It can 'integrate' qualitative and quantitative evidence through narrative juxtaposition (discussing diverse forms of evidence side by side). It is less concerned with assessing evidence quality and more focused on gathering relevant information that provides both context and substance...' (Kastner et al., 2012)

This approach is helpful for the triangulation of different types of evidence, and is useful for theory-building where there is a diverse and potentially large evidence base (Dixon-Woods et al., 2005). With a focus on the breadth of evidence, rather than its depth, some authors might now describe this type of work as a 'scoping' review (Davis et al., 2009). However, the term 'scoping' is variably used and ambiguously defined (Levac et al., 2010), so I avoid it here.

2.2 Aims

The aims of this narrative review are to:

(i) Establish a high level overview of strategies to improve medical ward care;

(ii) Describe which strategies have been effective, and in what circumstances; and

(iii) Identify common methodological issues which would need to be addressed in future research.

2.3 Methods

2.3.1 Search strategy

The search strategy was an iterative one. I searched English-language publications in PubMed for articles pertaining to medical ward quality and safety, from inception until November 2013.

There were two broad strands of search terms. The first strand related to quality and safety outcomes, e.g., adverse events, complications, and objective and subjective healthcare outcomes. The second strand involved descriptions of the wards themselves, such as admission units, general medical wards, medicine inpatient service, or teaching wards.

Initial results were aggregated into themes, and as each major theme was defined, I identified seminal papers and systematic reviews, as well as the later studies that cited them. Where major reviews had been published on an overarching theme, the contributory papers relating to medical patients were analysed separately. I also searched the relevant literature curated by the Agency for Healthcare Research and Quality's patient safety network on its online database (Agency for Healthcare Research and Quality Patient Safety Network, 2014). Further themes and relevant studies were proposed by senior academic patient safety researchers in the UK and USA (Nick Sevdalis and Robert Wachter, respectively), initiating further hand searches.

This process continued until thematic saturation was reached - the point at which no further categories of improvement targets or interventions were identified. The strategy was intended to capture as broad as possible a representation of pertinent evidence for medical ward practice, without focusing unduly prematurely on any specialised topic.

2.3.2 Article selection

Articles were selected on the basis of their focus on medical ward care, and their relevance to quality and safety improvement. No papers were specifically excluded because of their choice of outcome measures. Clinical outcomes, process measures, and more subjective endpoints (e.g., patient- or staff-reported outcomes) were all included. Articles focusing solely on the episode of inpatient care, rather than transitions at admission or discharge, were highlighted. Studies describing computerised physician order entry systems, decision support systems or electronic communication devices were excluded. These comprise separate topics in their own right, and I felt that a focus on interventions not requiring significant capital investment would be of greater interest to most general medical audiences.

2.4 Results

Included studies described the type and incidence of problems on medical wards; observations of current practice; and attempts to improve outcomes for both staff and patients. The findings from the literature search are summarised in table 2.1. Those studies that inform a deeper understanding of how medical wards work, or describe an intervention, are explored further here.

Five major themes were identified within the evidence base:

1. Staffing levels and team composition;
2. Staff communication and collaboration;
3. Standardisation of the processes of care;
4. Early recognition and treatment of the deteriorating patient; and
5. Local safety climate.

The evidence synthesis across each of these five themes is presented in turn.

Table 2.1: Evidence for five strategies to improve medical ward care

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
Nurse staffing								
Kane, 2007	Meta-analysis	N/A	N/A	N/A	N/A	Patients per RN	6% reduction in medical patients' odds of death for every 1 RN full-time equivalent increase Dose-response relationship between nurse staffing & all-patient mortality No consistent relationship with several other patient outcomes presumed sensitive to nurse staffing	Strong correlation between nurse staffing & mortality, but further work needed to establish a theoretical model & demonstrate causality
Needleman, 2011	Retrospective observational study	Single tertiary academic medical centre, USA	N/A	N/A	N/A	Patient turnover: the sum of unit admissions, transfers, and discharges, divided by the start-of-shift census. Complete turnover equals 1; high turnover defined as more than one standard deviation above that unit's mean value	Mortality significantly associated with patient exposure to below-target or high-turnover shifts; hazard ratio 1.02 for below-target shifts.	No data on staff other than nurses. No data on patients with do-not-resuscitate orders.

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
						Staffing 'below target' if the total nursing hours worked in a shift was 8 hours less than the anticipated requirement		
Medical staffing & team composition								
Spellberg, 2012	Prospective controlled study	Single academic public teaching hospital, USA	Reduced trainee:consultant ratio from 5:1 to 2:1 or 3:1	1 year	Intervention teams in operation for 12 weeks of the study period	Maximum 6-7 new admissions per intern over 24hr call period. Average daily census 5 patients/intern. Maximum daily census 10 patients/intern	Better quality of educational experience & quality of life for staff, with more time devoted to teaching & patient care. No change to objective healthcare outcomes	Did not use a validated survey tool to assess staff experience, and response rates not given Compared overall mortality & length of stay, with no risk adjustment Did not recruit pre-specified number of patients to the experimental teams

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
McMahon, 2010	Prospective controlled study	Single community teaching hospital, USA	Reduced trainee:consultant ratio from 3:1 to 5:2	1 year	Intervention teams in operation throughout the study period	Maximum census 15 patients/team. Median 62 hours work/week. Average daily census 3.5 patients/internet in intervention group vs 6.6 in control group	Length of stay reduced by 0.5 days, with no change in readmission rate or quality of care	No formal cost analysis
			Geographic localisation				Significantly higher resident & intern satisfaction in the intervention teams	Did not use a validated survey tool to assess staff experience
			Multidisciplinary teams & more stable medical teams					39% trainees spent time in both intervention and control groups
Ong, 2007	Retrospective cohort analysis	Single tertiary care university hospital, USA	N/A	N/A	N/A	Team workload (number of admissions) on patient's admitting day	Increasing team workload on the day of admission increases length of stay, total costs & risk of inpatient mortality - particularly when more than 9 patients are admitted on the day of admission	Study not powered to detect differences in mortality

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
						Team workload during a patient's hospitalisation (average number of patients on the patient's team)	Increasing team workload during the entire stay reduced average length of stay & total costs	Only first hospitalisations in the study period were analysed
								Workload impact of co-management with ICU physicians not described
Ahmad, 2011	Prospective controlled study	Single university teaching hospital, UK	Twice-daily consultant ward round	12 months	All staff worked within new care system in intervention wards	Ward consultant inpatient census increased from 13 to 25 (inferred), with relief from most other clinical activities	Increased discharges & reduced average length of stay, with unchanged readmission rate & mortality	Control wards chosen for similar caseloads; no patient demographics to validate this choice. No comparison of patients' severity of illness
Bell, 2013	Retrospective observational study	91 acute hospital sites, UK	N/A	N/A	N/A	Not given	Reduced hospital adjusted case fatality rates, and lower 28-day readmission rates, in hospitals that have a continous admitting consultant presence for at least 4 hours per day	Survey responses informed analysis of consultant working patterns - likely response bias
								Full data only available for 58% of hospitals

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
								Confounding factors include nurse staffing, access to diagnostic services, workload, etc.
Fielding, 2013	Prospective controlled study	Single university hospital, UK	Consultant-delivered multidisciplinary teams, with dedicated pharmacist, senior nurse & shared medical assistant	Five months	Intervention teams in operation for 10 and 16 weeks of the study period	Median caseload of 9-14 patients per medical team	Shorter length of stay	Unclear whether intervention group improvements were due to consultant input or new team structure No comparison of patients' severity of illness
Kennedy, 2009	Observational & interview study	Three teaching hospitals, Canada	N/A	N/A	N/A	Not given	Trainees' decisions to seek clinical support are influenced by the clinical question, supervisor factors & trainee factors Trainees prioritise their own professional credibility when asking for help	Potential observer effect

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
Lucas, 2012	Cluster randomised non-inferiority trial	Single public teaching hospital, USA	2- vs 4-week inpatient consultant rotations	1 year	78% attending physicians participated in the intervention groups; individuals staffed a median of two 2-week rotations & one 4-week rotation	Not given	No difference in length of stay or unplanned patient revisits	Did not account for physician workload during each rotation
							Better self-rated measures of physician burnout & emotional exhaustion with 2-week rotations, but worse evaluations by trainees	No validated measure of physician burnout
Teamwork & communication								
Singh, 2012	Prospective quasi-experimental study	Single urban academic medical centre, USA	Geographic localisation	20 weeks	Intervention teams comprised 2/10 general medical teams, on 1/14 nursing units	Maximum 5 admissions/team/day	No change in 30-day readmissions or hospital costs	No power calculation

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
						Maximum patient census 12-16/team	Improved staff workflow & productivity, but trend towards higher patient length of stay	Faculty officer determined patient allocation
								Different workloads of intervention & control teams
O'Leary, 2009	Uncontrolled before-after study	Single tertiary care teaching hospital, USA	Geographic localisation	1 year	73% eligible patients localised to designated units in the study period, more so for teaching service patients	Mean number of patients on day of interview = 10 (doctor), 4 (nurse)	More frequent communication, with nurses & physicians better able to correctly identify one another	Uncontrolled before-after design
							Little impact on agreement on plan of care	Younger, less experienced nurses post-intervention may have focused more on communication
Curley, 1998	Non-randomised cluster trial	Single acute care county hospital (local tertiary referral centre), USA	Daily interdisciplinary rounds	6 months	Not given	15 patients/team	Lower length of stay & hospital charges	No data on clinical outcomes

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
								Limited descriptions of the rounds themselves (attendance rates, senior involvement, etc.)
Wild, 2004	Randomised controlled trial	Single community hospital, USA	Daily interdisciplinary rounds	2 months	Not given	Not given	No difference in length of stay or indirect measures of care quality	Small study, underpowered to detect length of stay changes
								Highly selected group of stable patients, many on well-defined clinical pathways
O'Mahony, 2007	Uncontrolled before-after study	Single university-affiliated urban community teaching hospital, USA	Thrice-weekly multidisciplinary rounds	1 year	Not given	Not given	Reduced length of stay, with improved performance on core quality measures	No description of intervention uptake
			Prompts and feedback on core quality measures				Improvements in self-reported trainee education & attitudes	Unvalidated survey tool for trainee self-reports
Vazirani, 2005; Cowan, 2006	Comparative quasi-experimental study	Single tertiary academic hospital, USA	Daily interdisciplinary rounds	2 years	91% charge nurse, 100% trainee & 29% consultant attendance at interdisciplinary rounds	3 RNs & 2.5 non-registered nurses per 15 beds	Improved physician reports of collaboration with nurses	Staff survey used unvalidated tool

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
			Hospitalist medical director			16 vs 22 patients per team (intervention vs control)	Reduced length of stay, increased hospital profit, and no adverse change in mortality or readmissions	Different workloads of intervention & control teams
			Advanced practice nurse contributing to inpatient & post-discharge care				No impact on patient ratings of hospital care or health-related quality of life	25% patients did not complete study
O'Leary, 2010; O'Leary, 2011	Cluster randomised trial	Tertiary care teaching hospital, USA	Daily interdisciplinary rounds, using structured communication tool	6 months	92% patients discussed each day; attendance exceeded 82% for each discipline	Not given	Improved nurse ratings of teamwork climate & collaboration	Intervention unit nurses were newer to the institution
		Physician teams already localised to their units					Reduced adjusted rate of adverse events	More patients under the care of a hospitalist in the intervention unit
							No change in length of stay or hospital costs	Adverse event screeners not blinded to patient group allocation
O'Leary, 2011	Cluster randomised trial	Tertiary care teaching hospital, USA	Daily interdisciplinary rounds, using structured communication tool	24 weeks	86% patients discussed each day; attendance exceeded 85% for each discipline	Not given	Higher nurse ratings of communication, collaboration, teamwork & safety climate	Case mix differences between control and intervention units

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
								Contamination between intervention and control groups: hospitalists worked on both units
Zwarenstein, 2009	Cochrane review	N/A	Practice-based interprofessional collaboration interventions	N/A	N/A	Not given	Healthcare processes & outcomes can be improved with these interventions. Small study sizes & heterogeneity limit inferences about their key elements or effectiveness	Overly-restrictive inclusion criteria
Early recognition & treatment of the deteriorating patient								
McNeill, 2013	Systematic review	N/A	Single & aggregate weighted scoring systems, emergency response teams, medical emergency teams & multidisciplinary outreach services	N/A	N/A	Not given	Weak evidence for the benefits of these interventions Comprehensive systems needed both to detect & reliably respond to patient deterioration. The most pertinent focus depends on the healthcare system	Predominance of uncontrolled before-after studies weakens the strength of the conclusions

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
							Aggregate scoring systems are more effective	
Bailey, 2013	Randomised controlled crossover study	Single academic medical centre, USA	Real-time automated alerts to nurse manager when patients met designated warning threshold	12 months	Not given	Not given	No impact on mortality or ICU transfers	Efferent limb of alert system was variable & not assessed - no indication that the alert was acted on
							Algorithm identified patients at higher risk of ICU transfer & death, but with low positive predictive value	Likely alert fatigue for clinicians, given the algorithm's low positive predictive value
Winters, 2013	Systematic review	N/A	Rapid response systems	N/A	N/A	Not given	Moderate evidence that these systems are associated with reduced rates of cardiorespiratory arrest & reduced mortality	Dependence on before-after studies hampers evidence base, as does likely publication bias
Standardisation of care - clinical protocols & checklists								
Ko, 2011	Systematic review	N/A	Checklists	N/A	Not given	Not given	Patients twice as likely to receive appropriate antibiotics within 8 hours	Interventions likely to have been implemented differently in each hospital
Robb, 2010	Uncontrolled before-after study	Single urban acute trust (incorporating 3 hospitals), UK	Care bundles	1 year	1200 care bundle sheets used in total, no further details given	Not given	Reduced hospital standardised mortality ratio for the targeted diagnoses	5/8 care bundles potentially relate to general medical ward care

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
								Confounding factors include introduction of an AAU & change in consultant working pattern
Rotter, 2010	Cochrane review	N/A	Clinical pathways	N/A	N/A	N/A	Reduced in-hospital complications & improved documentation No differences in readmissions / mortality Likely reduction in length of stay & hospital costs	4/6 relevant studies had moderate risk of bias
Standardisation of care - acute admissions units								
Scott, 2009	Systematic review	N/A	Acute admissions units	N/A	Not given	Not given	Reduced mortality & reduced length of stay, with no increase in readmissions	No prospective study adequately controlling for patient demographics or referral patterns from primary care Publication bias: only small fraction of acute admissions units published their outcomes

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
Rooney, 2008	Uncontrolled before-after study	Single teaching hospital, Ireland	Acute admissions unit	5 years	Not given	Median 17 admissions/unit/day	5.6% absolute reduction in all-cause hospital mortality	Support services prioritised AAU patients
						Physician / team workload not specified		No assessment of secular trends, nurse staffing, physician workload or multiple other confounders
Moloney, 2007	Uncontrolled before-after study	Single teaching hospital, Ireland	Acute admissions unit	2 years	Not given	Not given	Reduced hospital length of stay, with less variance between medical teams	Support services prioritised AAU patients
							No change in readmissions	No assessment of secular trends, nurse staffing, physician workload or multiple other confounders
Standardisation of care - behavioural nudges								
Caminiti, 2013	Cluster randomised controlled trial	Single university hospital, Italy	Data feedback (peer physicians' length of stay, patients ready for discharge, & local audit)	1 year (and long-term impact assessed again one year later)	Intervention in place for 11 months of the study period	Not given	Reduction in avoidable hospital days, with no change in 30-day readmission or mortality rates	Data collected only on one day each month

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
								Findings may not be applicable to other hospitals; 32% 'medically unwell' patients were in hospital for no apparent reason
de Jong, 2006	Retrospective cohort study	All hospitals in New York state, USA	N/A	N/A	N/A	Not given	Multihospital physicians conform to local standards for lengths of stay	Only two medical diagnoses (chronic obstructive pulmonary disease and congestive heart failure) included Confounding factors (workload, physician experience, etc.) not explored
Other multi-faceted interventions								
Lagoe, 2005	Uncontrolled before-after study (descriptive data)	Four acute care hospitals, USA	Hospitalists, feedback on physician performance, and incentives for community care homes	3 years	Not given	Not given	Average daily hospital census reduced by 12%. No effect on number of inpatient discharges	No statistical analysis No attempt to explore confounding factors or identify the most important component of the intervention

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
Blegen, 2010; Auerbach, 2012	Interrupted time series study	3 hospitals (academic university medical centre, non-teaching community hospital & integrated healthcare system), USA	Multidisciplinary team training sessions	2 years	Not given	1 RN per 4-5 patients	Higher staff ratings of safety culture	Multiple confounding factors over study period
		Predominantly hospitalist physicians	Multidisciplinary unit-based safety teams			Physician workload not given	Nurses rated safety culture more highly than other staff	Unclear why one intervention site responded differently to the other two
			Patient engagement with explicit daily goals				No change in readmissions or length of stay	Convenience sample of patients (with low response rate); may not be representative of population
							Improved patient perceptions of team function, & increased perception of safety gaps	Expensive intervention

(Table 2.1, continued)

Author, Year	Study design	Setting	Intervention(s)	Intervention duration	Intervention adoption	Workload measure(s)	Findings	Critique
Beckett, 2013	Statistical process control evaluation	Single hospital, Scotland	Geographic localisation Twice-daily consultant ward round Early identification & rescue of deteriorating patients Learning from adverse events End-of-life decision making	17 months	Not given	Not given	Reduced 30-day mortality, length of stay & cardiopulmonary resuscitation rate	Difficult to discern additional benefits of quality improvement programme above those of the ward-based team structure & increased consultant input

Abbreviations: AAU - acute admissions unit; ICU - intensive care unit; N/A - not applicable; RN - registered nurse.

2.4.1 Staffing and team composition

2.4.1.1 Nursing staff

2.4.1.1.1 Numbers There is limited evidence that better nurse staffing facilitates better care, with no prospective studies to assess the impact of improving nurse-to-patient ratios. The conclusion that nursing levels affect patient outcomes relies predominantly on a single centre's descriptive, longitudinal study (Needleman et al., 2011), and a meta-analysis of the dose-response relationship between nursing levels and patient deaths (Kane et al., 2007; Shekelle, 2013). However, not all patient outcomes presumed to be sensitive to nurse staffing show this relationship (Kane et al., 2007). Nonetheless, some states in the USA now mandate minimum nurse staffing levels based on this evidence. In the UK, the Francis Inquiry also recommended that formal tools be devised to establish staffing needs in individual units (Francis, 2013).

2.4.1.1.2 Skill mix The complexities of nursing provision do extend beyond a simple tally of nurse-to-patient ratios; relative contributions from 'registered' and 'unregistered' (or in the USA, 'licensed' and 'unlicensed') staff should be taken into account. A 'richer' skill mix, with higher proportions of care provided by registered nurses, is associated with improved outcomes for medical patients. These include (in observational studies) lower failure to rescue rates[1] (Needleman et al., 2002), reduced in-patient and 30-day mortality (Person et al., 2004; Estabrooks et al., 2005; Tourangeau et al., 2007), and improved patient satisfaction (Seago et al., 2006). Improving skill mix may be the least expensive of the options available to improve nurse staffing (Needleman et al., 2006), although a good learning climate on the ward may mitigate the impact of a poor skill mix (Chang and Mark, 2011).

2.4.1.1.3 Training Although a detailed discussion of nurse training lies outside the scope of this review, it is important to note that nurse education may also impact on patient outcomes.

[1] Failure to rescue was defined by Needleman et al. (2002) as death from one of five complications of care: pneumonia, shock or cardiac arrest, upper gastrointestinal bleeding, sepsis, or deep venous thrombosis.

In particular, units with higher proportions of Bachelor degree-qualified nurses have lower mortality (Estabrooks et al., 2005; Tourangeau et al., 2007), a lower risk of 30-day readmission (McHugh and Ma, 2013), and fewer complications of care (Blegen et al., 2013).

Interestingly, hospital-wide studies suggest that team composition – specifically, the aggregate level of education – contributes to an individual nurse's expertise, independent of their individual education and level of experience (McHugh and Lake, 2010). Medical wards may therefore require a critical mass of well-trained, expert nurses to guarantee high performance throughout the team. Taken together, these observational studies strongly suggest that adequate nurse staffing is necessary, but not sufficient, to improve patient outcomes. Skill mix, education and aggregate expertise are important moderating factors.

2.4.1.2 Physicians

2.4.1.2.1 Consultant involvement in care Physician staffing is similarly complex. There is no definitive agreement regarding the optimal degree of senior medical involvement in patient care, the substitution of less autonomous healthcare staff for doctors (Parekh and Roy, 2010), or the benefit of higher qualifications for clinicians managing medical patients (Norcini et al., 2002; Tourangeau et al., 2002; Tourangeau et al., 2006; Chen et al., 2006). Despite a firm consensus that consultant-delivered (as opposed to consultant-led) care should improve patient outcomes, there have been no major studies to compare the care delivered by fully accredited physicians versus postgraduate doctors in training (Academy of Medical Royal Colleges, 2012).

Attention has focused instead on the timing of consultant involvement, and the structure and workload of the teams they supervise. Consultants are likely to have an important role in coordinating team workload, both at the time of admission and throughout the inpatient stay. How the team handles this workload may have important effects on patient outcomes. Increases in team workload on a patient's day of admission are associated with increases in length of stay, total costs and inpatient mortality (Ong et al., 2007). Similarly, increases in monthly workload (monthly admissions) correlate with a higher risk of patient readmission (Averbukh

and Southern, 2012). A large retrospective analysis showed that UK hospitals providing at least four hours of daily consultant cover for new admissions had lower risk-adjusted mortality and readmission rates (Bell et al., 2013).

2.4.1.2.2 Timing and frequency of consultant input Increasing senior clinical input later in the admission may also be beneficial. In the UK, medical inpatients on general wards typically see their consultant physician only twice each week. Single-centre studies have shown that daily or even twice-daily consultant review is feasible, and may reduce length of stay without adversely affecting readmission rate or mortality (Ahmad et al., 2011; Fielding et al., 2013). In one study, consultants were accompanied by a pharmacist, senior ward nurse and medical assistant; it is impossible to separate the effect of consultants' frequent bedside presence from that of the novel (interdisciplinary) team structure (Fielding et al., 2013). Investment in auxiliary services may also be necessary if senior physicians are to improve patient outcomes via more frequent patient reviews (Carey et al., 2005).

Outside the UK, there is less direct evidence that medical patients benefit from greater senior clinical input. In the USA, increasing the ratio of consultant physicians to trainees (and simultaneously reducing the team's patient census) improved both the quality of junior doctors' work and their educational experience (McMahon et al., 2010; Coit et al., 2011; Spellberg et al., 2012). Objective healthcare outcomes and patient satisfaction were not affected (McMahon et al., 2010; Spellberg et al., 2012). Differences between the UK and the USA in what constitutes an 'acceptable' patient census per doctor may explain these discordant findings (Goddard et al., 2010; McMahon et al., 2010).

2.4.1.2.3 Consultant variation and stakeholder trade-offs Not all consultants are equal, however: those focusing on inpatient care (hospitalists) may reduce resource use and improve patient outcomes (Meltzer et al., 2002), even with relatively little of their time spent on direct patient care (O'Leary et al., 2006). What it is that contemporary consultant physicians actually do during their ward rounds has only recently been re-assessed (Stickrath et al., 2013), and little

is known about the effect of different ward round behaviours on clinical outcomes.

Maximising senior staff presence on the wards also requires some compromise. A fortnight of ward coverage may be tolerable for an individual consultant, but longer stints are associated with self-assessed burnout and emotional exhaustion (Lucas et al., 2012). Yet trainees rate short periods of consultant cover more harshly, illustrating the challenges of counterbalancing the demands of different staff members, patients and fiscal requirements.

Overall, it seems likely that increasing consultant-delivered care will improve hospital efficiency and patient outcomes most where trainees have traditionally been less closely supervised.

2.4.1.2.4 Junior doctors: working patterns & credibility Excessive working hours have long been considered a risk to the safe delivery of care by junior doctors. However, the anticipated benefits of restricting their duty hours have not materialised, with no clear improvements in patient outcomes (Fletcher et al., 2011). The unintended consequences of more frequent shift changes - such as the risks arising from inadequate handovers - may have replaced errors provoked by exhaustion [see 2.4.2.4]. Changes in team stability may also play a part. Juniors consider their professional credibility before calling for help (Kennedy et al., 2009), and it is unsurprising that the loss of consistent team relationships and supervision may contribute to team failings (Singh et al., 2007).

Having considered the evidence relating to individual staff groups on the ward, I will now describe the studies exploring their interdisciplinary practice.

2.4.2 Staff communication and collaboration

Ward teams function less well, and team members understand each other more poorly, than they might think. In general medicine, physicians dominate interactions with other health professionals, 'inflicting' information on their colleagues rather than exchanging it (Reeves et al., 2009). Despite this tendency, physicians still think highly of their ability to collaborate with nurses, a view not shared by the nurses themselves (O'Leary et al., 2010).

Worryingly, some junior doctors, while aware of communication problems, dismiss them as no threat to patient care (Weinberg et al., 2009). This may not simply be due to professional myopia. In a small Swiss study, nurses and junior doctors had discordant perceptions and unmet expectations about each other's role, and limited knowledge of each other's profession (Muller-Juge et al., 2013). Nonetheless, a Cochrane review of RCTs of interprofessional collaboration interventions only labelled them 'promising', rather than 'proven' (Zwarenstein et al., 2009). The review's inclusion criteria may have been overly restrictive (Shojania, 2013), and some of this literature will therefore be discussed in more detail.

2.4.2.1 Structured team practices

One important study highlighted the potential impact of a well-placed change in hospital practice. In a non-randomised cluster trial, bringing ward staff together in daily interdisciplinary rounds reduced length of stay and total hospital charges (Curley et al., 1998). Earlier smaller studies, with different patient demographics, had failed to demonstrate a comparable effect, suggesting that stable patients with straightforward diagnoses may have less to gain from an investment in multidisciplinary input. Trials also need to be adequately powered to detect relatively small changes in length of stay (Wild et al., 2004).

Subsequent research confirmed that regular interdisciplinary meetings (of varying frequency, and sometimes as part of multi-component interventions) had a role in reducing length of stay, without negatively affecting readmission rates or mortality (Cowan et al., 2006; O'Mahony et al., 2007). Standard practice has evolved since those studies, with more emphasis on the importance of interdisciplinary work, as well as baseline reductions in the average length of stay. It may now be more difficult to further reduce length of stay or hospital charges through interdisciplinary care arrangements (O'Leary et al., 2010).

Interdisciplinary interventions may still be worthwhile for other reasons, however. They can reduce the incidence of adverse events, and improve nurses' perceptions of teamwork and collaboration (O'Leary et al., 2010; O'Leary et al., 2011a). These perceptions are important as

they reflect a belief in the inherent value of an intervention; if well-received, trial interventions may be continued indefinitely (O'Leary et al., 2010). Importantly, not all collaborative interventions will be welcomed. Even simple changes to communication styles can be widely rejected, particularly if senior staff refuse to role-model them, they conflict with existing hierarchies, or if interprofessional work is not prioritised (Rice et al., 2010).

2.4.2.2 Geographic localisation

Forming a cohesive unit is a challenge for teams responsible for patients on multiple wards. Locating staff on the same ward (geographic localisation) has face validity, although its anticipated benefits have not all been borne out in objective studies. Improvements in communication, productivity and workflow may come at the unexpected cost of a greater length of stay, perhaps reflecting the perverse incentives of geographic localisation (O'Leary et al., 2009; Singh et al., 2012). For example, teams might retain patients in order to avoid new admissions to their assigned beds. Even when organisations attempt geographic localisation, bed pressures still result in patients being placed as 'outliers' or 'boarders' on other wards. Staff are well aware of problems arising from this inappropriate ward allocation (Goulding et al., 2012).

2.4.2.3 Team training

Practical changes that facilitate better teamwork and collaboration (e.g., making time for interdisciplinary meetings and improving staffing arrangements) might be more successful than teamwork training per se, which, despite its promise, can fail to improve patient outcomes (Vazirani et al., 2005; Auerbach et al., 2012). Very little research has evaluated team training in the general medical environment (O'Leary et al., 2012b). A Cochrane review of interprofessional education did not identify any studies with a significant focus on internal medicine (Reeves et al., 2013). Further, there is little published evidence that simulation training for medical teams – providing a setting in which they can rehearse their management of medical emergencies – improves patient outcomes. However, it does increase staff confidence and competence, with

improved team and non-technical skills (Schmidt et al., 2013; Shah et al., 2013).

2.4.2.4 Handover

Communication between shifts demands a separate focus. As inpatient care fragments, transfers of responsibility for ongoing patient care between different healthcare providers ('handovers', 'handoffs' or 'signouts') are increasingly frequent. Patients can be handed over 15 times in a five-day hospitalisation, and individual doctors participate in more than 300 handovers per month (Vidyarthi et al., 2006). Despite their ubiquity, the quality of each handover is typically poor, whether trainees or consultants are involved (Greenstein et al., 2013). Handover failures may arise, at least in part, from fundamental disagreements about what should be discussed (Helms et al., 2012). Perhaps as a result, receiving clinicians neither ask questions nor demonstrate the active listening behaviours that promote memory retention (Horwitz et al., 2009; Greenstein et al., 2013). With inadequate information to make clinical decisions, near misses and adverse events are inevitable (Arora et al., 2005; Vidyarthi et al., 2006; Horwitz et al., 2008).

Addressing these challenges requires a combination of structural and process changes, and clinical leadership (Vidyarthi et al., 2006). Standardisation with a structured handover form leads to fewer perceived errors by the covering doctor (Salerno et al., 2009). Electronic handover templates, especially those that auto-populate with accurate patient data, may be more efficient than handwritten versions, although their role in reducing adverse events is unclear (Petersen et al., 1998; Van Eaton et al., 2010; Graham et al., 2013). When designed well, the same handover tool may be used by the entire multidisciplinary team, which might be an important goal in itself in improving interdisciplinary communication (Collins et al., 2011; Vawdrey et al., 2013).

Consultant supervision of the handover process, with a structured teaching programme or regular feedback, improves the quality of the handover's content as well as junior doctors' self-confidence in conducting an effective handover (Chu et al., 2009; Bump et al., 2012). Peer evaluations of handover quality, by contrast, are more lenient and less reliable (Arora et al., 2013). Nonetheless, junior doctors identified by their peers as having exemplary handover skills agree

common strategies for effective handover: prioritisation of acutely ill patients, limiting discussion of straightforward patients and inactive issues, and avoiding duplication of information held in the medical record (Helms et al., 2012). Other recommendations include protected time for handover preparation, face-to-face discussion, and a designated interruption-free environment, although these may be difficult to implement in practice (Arora et al., 2009; Raduma-Tomas et al., 2012).

2.4.3 Standardisation of the processes of care

2.4.3.1 Acute admissions units (AAUs)

Medical teams are typically large, unstable, and geographically spread out — all factors thought to hamper team performance. AAUs[2] address some of these issues. AAUs are wards that concentrate and co-locate staff expertise in acute disease management for the initial period of a patient's admission. Popular in the UK but uncommon in the US, they are both effective and safe, reducing all-cause mortality and length of stay with no negative impact on readmissions (Moloney et al., 2007; Rooney et al., 2008). Overall, they have shown real benefits, despite a high risk of publication bias in their evidence base (Scott et al., 2009). Importantly, creating an AAU is no panacea: these units too may need specific help to improve process measures and patient outcomes (Beckett et al., 2013).

2.4.3.2 Checklists

As individuals, clinicians do not reliably provide interventions of proven value, with high rates of unexplained variation (Huang et al., 2008). Surgical checklists addressing this are associated with improved outcomes (Haynes et al., 2009), and there has been great interest in finding a medical equivalent. Medical checklists may find local acceptance if driven by senior physicians with clinical credibility and organisational support (Amin et al., 2012). There is currently little evidence for their impact on patient outcomes, however: a systematic review identified only a

[2] AAUs are also known as acute medical units (AMUs).

single formal description of checklists in an acute medical setting (Ko et al., 2011). Checklists continue to be devised organically, with no validated methodology for their development (Hales et al., 2008), and the same hierarchies that impede checklist adoption in the operating theatre could also hinder their uptake on medical wards (Amin et al., 2012).

2.4.3.3 Clinical protocols

Clinical protocols, promoting best practice when patients have a clear presenting symptom or acute diagnosis, can also aid standardisation of care. A Cochrane review concluded that these protocols are associated with reduced in-hospital complications and improved documentation, without negatively impacting on length of stay or hospital costs (Rotter et al., 2010). Most experts agree that poorly performing teams are likely to benefit from these interventions, although there are dissenters (Rotter et al., 2012; Vanhaecht et al., 2012).

Disease-specific guidelines typically do little to address multimorbidity, with its attendant risks of polypharmacy and adverse drug interactions (Boyd et al., 2005). Inappropriately slavish adherence to a protocol may actually diminish the quality of care provided, to the extent that the protocol can no longer be considered applicable (Lugtenberg et al., 2011). This is particularly relevant as the general medical inpatient population becomes older and medically more complex, often presenting with non-specific symptoms, for which there is no single best pathway of diagnosis or management.

Of the studies meeting the Cochrane review's inclusion criteria, only six described clinical pathways used substantially for medical ward patients; four of those had a moderate risk of bias (Rotter et al., 2010). However, a subsequent UK study from a single hospital demonstrated reduced mortality rates after implementing care bundles in targeted clinical areas (Robb et al., 2010). The question of how best to harness the benefits of clinical protocols for an increasingly complex, frail, elderly inpatient population has not yet been resolved.

2.4.3.4 Behavioural nudges

Physicians can also be 'nudged' towards standardisation. Showing physicians the untapped 'discharge potential' on their wards (patients who meet pre-defined criteria indicating their readiness for discharge), and overtly comparing physician-level average length of stay, reduced avoidable hospital days in one cluster randomised trial (Caminiti et al., 2013). Even with reservations about this study's generalisability, strategies promoting direct physician accountability and awareness of local practice may have wider benefits. Organisations might capitalise on physicians' tendency to adapt their own practice to the norms of each workplace (Lagoe et al., 2005; Jong et al., 2006).

2.4.4 Early recognition and treatment of the deteriorating patient

It has been difficult to prove the benefits of early warning scoring systems, tools that grade the patient's physiological disturbance and recommend appropriate action, with conflicting evidence and a paucity of high quality data (McNeill and Bryden, 2013). Lacking information specific to medical wards, hospital-wide data suggest that aggregate weighted scoring systems are most effective (McNeill and Bryden, 2013), but these remain entirely dependent on two factors.

Firstly, they should help clinicians recognise a deteriorating patient earlier than they would otherwise — the 'afferent limb' of the system. In day-to-day use in an Emergency Department, though, healthcare providers did not complete the necessary charts fully and even miscalculated the constituent scores (Wilson et al., 2013). It is unsurprising that the evidence supporting these tools is underwhelming, if they are routinely incomplete and numerically incorrect.

Secondly, clinicians should react appropriately to a documented deterioration. The 'efferent limb' of early warning scoring systems often involves deployment of a rapid response team, a group of clinicians with critical care skills, to quickly assemble sufficient expertise at the bedside of a sick patient. However, those teams' impact is limited by repeated failures to seek their help,

despite guidance for ward staff (McNeill and Bryden, 2013). Even an automated system alerting senior ward nurses to a patient's deterioration had no effect on clinical outcomes, perhaps because the activation of the nurses was no guarantee that appropriate action would be taken (Bailey et al., 2013).

A systematic review concluded that rapid response systems – comprising both early warning triggers and defined clinical responses – are associated with reduced rates of cardiorespiratory arrest outside the ICU and reduced mortality (Winters et al., 2013). Many of the reviewed studies were of poor quality, relying on before-and-after designs whose results may not be borne out when compared to more rigorously controlled research. Nonetheless, these rapid response systems have become an established feature of most healthcare organisations. Some authors contend that their 'success' is, at best, a mitigation of hospital failures earlier in the admission, and that a better approach would be to eliminate those failures at their source (Litvak and Pronovost, 2010).

2.4.5 Ward safety climate

High levels of safety culture and safety climate [see box 1.2] are associated with organisation-wide reductions in adverse events, and improving them is therefore a valid goal (Mardon et al., 2010). However, two systematic reviews found little evidence that interventions to improve safety climate improved patient outcomes: at the current time, such interventions are only 'potentially effective' (Morello et al., 2013; Weaver et al., 2013). Very little of the safety climate literature relates to adult medical wards. The available data are conflicting, in that improvements in teamwork and collaboration are not always paralleled by a change in safety climate (O'Leary et al., 2010; O'Leary et al., 2011b). Conversely, safety culture can be improved without definitively improving teamwork (Blegen et al., 2010; Auerbach et al., 2012).

Without a good theoretical understanding of how these factors interact, assessing different staff groups in turn is important. In operating theatres, for example, nurses are the least satisfied with collaboration and teamwork and are able to perceive greater improvements than other staff

(Makary et al., 2006). The most disenfranchised staff group may report the worst safety culture, although the identity of this group will vary between microsystems. Compared to the typical surgeon-dominated operating theatre, the medical ward is relatively 'owned' by its nurses, with physicians and pharmacists often seen as more transient visitors (Blegen et al., 2010). These subtleties complicate the evaluation of safety culture interventions.

2.4.6 Combined approaches

Some organisations have introduced wider-ranging programmes, successfully bringing together all five major themes identified in this evidence base (Lagoe et al., 2005; Beckett et al., 2013; McAlister et al., 2014). Their outcomes vary, from reduced length of stay in a large scale RCT to improved mortality in a single centre's longitudinal study (Beckett et al., 2013; McAlister et al., 2014). In each case, there was an overt organisational statement of the project's aims, as well as a commitment to the substantial organisational change that improvement would require. This was necessary, but not sufficient, for success: why an intervention does well in one setting and not another, even with similar organisational backing, remains unclear. This emphasises the importance of multi-centre trials (Blegen et al., 2010).

2.5 Discussion

This narrative review appears to be the first deliberate synthesis of the published literature on improvement strategies for general medical ward care. Many of the interventions described in this chapter had themselves been the focus of previous reviews. However, those reviews typically aggregated reports of their intervention's use across different settings. For example, the Cochrane review by Zwarenstein et al. (2009) included structured interprofessional interventions enacted in medical wards, surgery, nursing homes, and ambulatory care. Similarly, Ko et al. (2011) reviewed the use of safety checklists in all acute hospital settings, including intensive care, surgery, the emergency department and acute care departments.

The deliberate choice to summate the effects of complex interventions[3] in different settings has two main disadvantages. First, it ignores differences between microsystems, the organisational level at which important variation occurs [see 1.5.1]. This is particularly important when incorporating medical wards into broader summaries, given the specific challenges that they face [see box 1.1]. The efficacy of the intervention may depend very much on the characteristics of the microsystem, as will the details of the implementation model it requires.

Second, it obscures the relative paucity of high quality evidence pertaining specifically to medical wards, compared to more stable clinical environments such as the ICU and operating theatre. For example, Ko et al. (2011) found only one checklist study based on medical wards. Interventions tested in other settings may be extrapolated too broadly to the medical ward, with little specific evidence of similar effectiveness. Nonetheless, there was still sufficient evidence to elicit broad categories of improvement interventions in this microsystem, and to meet the other aims of this review, as summarised below.

2.5.1 Summary of findings

2.5.1.1 High level overview of strategies to improve medical ward care

Interventions to improve the quality and safety of medical ward care can be grouped into five common themes: staffing levels and team composition; interdisciplinary communication and collaboration; standardisation of care; early recognition and treatment of the deteriorating patient; and local safety climate. These categories have good face validity, addressing the common ward failings elicited by McQuillan et al. (1998) [figure 2.1; see also 1.4.2.1]. In doing so, they facilitate better care and promote improved outcomes.

[3] Complex interventions are commonly defined as 'interventions that comprise multiple interacting components, although additional dimensions of complexity include the difficulty of their implementation and the number of organisational levels they target' (Craig et al., 2008; Moore et al., 2015).

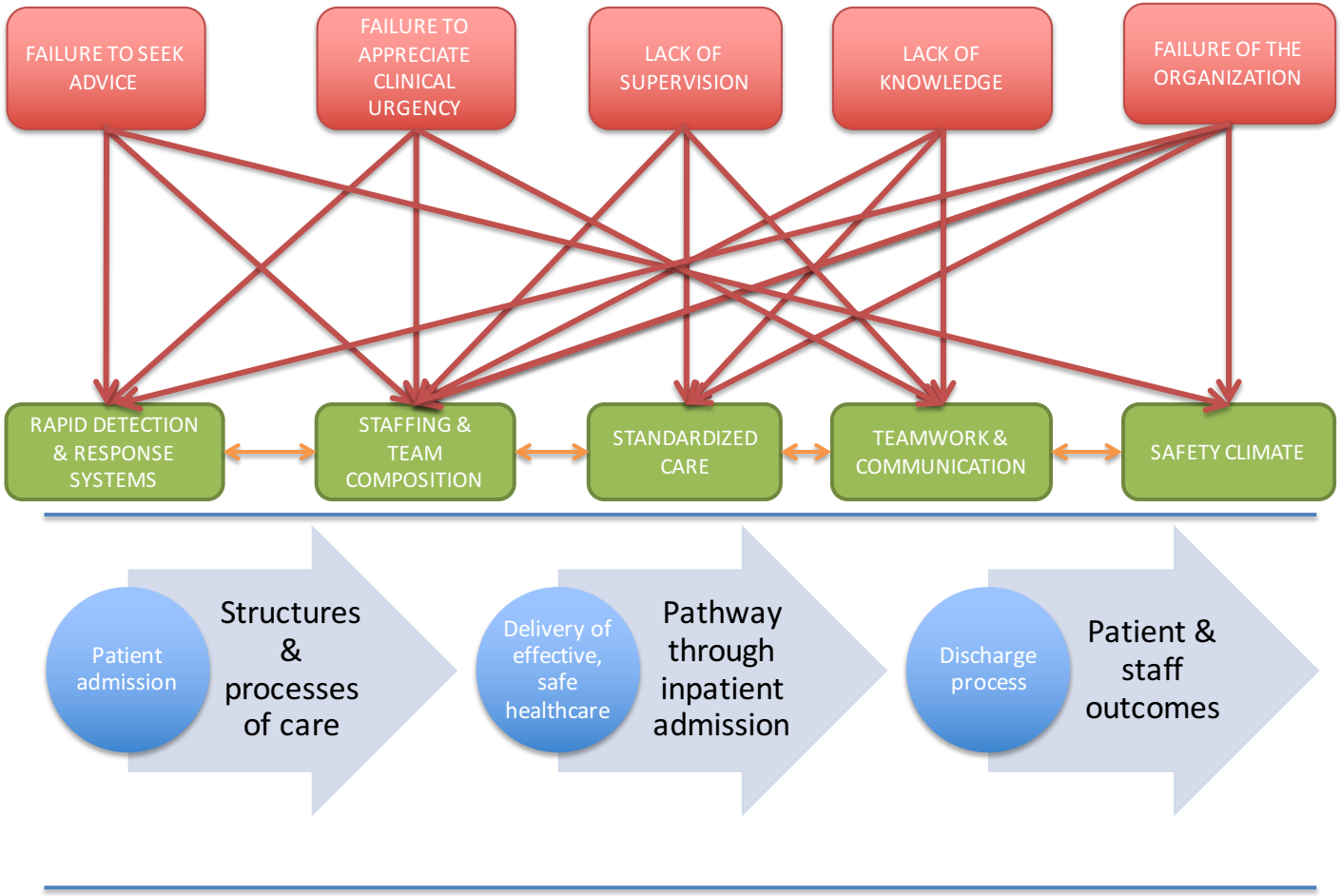


Figure 2.1: Common themes for medical ward quality & safety interventions. Intervention themes [green] address previously-identified failings [red], facilitating better care and promoting improved outcomes.

2.5.1.2 Which strategies are effective, and in what circumstances?

Many interventions in these common themes were reported to improve medical ward care to some extent. However, there was little definitive evidence of process improvements leading to better patient outcomes. Successful innovations in service delivery frequently combined multiple elements from these different themes, making it difficult to distinguish their individual effects. Overall, these improvement strategies are best described as promising, rather than proven.

Strategies to improve medical ward care are heavily context-dependent, and different ward stakeholders may have competing perspectives of what an effective intervention should achieve. Integrating interventions into existing workflow, along with meaningful and sustained senior support, may be necessary prerequisites for their success, but most reports did not focus on identifying important implementation factors. Comparison of the same type of intervention in different contexts is hampered by international variation in what constitutes an acceptable model of care.

2.5.1.3 Common methodological issues

Many of the studies discussed in this review shared the methodological failings that feature more broadly in QI literature [see 1.2.6]. Inadequate study designs, and failures to account for secular trends, were common. Specific additional issues included the failure to account for workload (an important confounding factor), the absence of an implementation evaluation, and a predominance of small, underpowered, single centre studies. Unit of analysis errors (interpreting results at the patient level when interventions were delivered at ward level, without adjusting for the clustering of results) were also common.

2.5.2 Study limitations

2.5.2.1 Search strategy

The principal limitation of this narrative review was its iterative search strategy. It is possible that relevant literature was not included as a result. However, given the wide-reaching primary aim for the review and the variety of interventions and study designs to be described, the search strategy could not be more exhaustive. In any case, the review reached thematic saturation, and later reports of novel interventions (published after the review period ended) have also been well categorised by the themes defined here. This suggests that the five major themes of medical ward improvement are sufficiently comprehensive.

2.5.2.2 Synthesis methodology

Much as the search strategy developed iteratively, so too did the synthesis of the included studies. The absence of a pre-defined analytical framework might be considered a limitation of this work. However, my approach adopted many of the recommended principles for narrative synthesis (Popay et al., 2006). Recommendations include clustering and tabulating the primary studies, to facilitate an inductive thematic analysis that 'translates' the dominant concepts reported in them. This process helps an author to explore the relationships within and between the themes. This is, by its nature, an iterative process; artificially constraining a theory-generating review to a pre-defined framework would have been inappropriate (Popay et al., 2006).

2.5.2.3 Strength of conclusions

This review was only able to draw tentative conclusions. This was a function both of the review methodology itself and the primary studies on which it relied. The methodological limitations of those primary studies were numerous [see 2.5.2.2], and many described medical ward data without focusing on medical wards *per se*. This precluded all but the broadest statements regarding the likely effectiveness of medical ward interventions. It is important to note that the

systematic reviews cited in this chapter also drew limited conclusions about the definitive benefits of these interventions, even with a narrower focus (Zwarenstein et al., 2009; Weaver et al., 2013).

2.5.3 Implications for further work

Calls for evidence-based changes to service delivery are increasing; this review is a timely summary of the evidence for care delivery interventions in the most familiar inpatient setting. A limited number of prominent improvement strategies were established within the broader literature; each of these can now be evaluated in more detail as required.

A number of other specific opportunities to contribute to the existing knowledge base were also identified. The literature currently lacks robust, mixed method assessments of medical ward interventions that can be disseminated at scale. Structured interventions should target caregivers who are disadvantaged by the existing workplace hierarchy. Evaluations should ideally include both clinical and staff-reported outcomes, representing multiple stakeholders. The development and evaluation of a novel interdisciplinary intervention, described in chapter 4, are informed by these findings.

2.5.4 Conclusion

The five strategies for medical ward improvement are supported by a limited evidence base. The next chapter focuses on one of these major improvement strategies: staff communication and collaboration. Interdisciplinary practice on the ward is increasingly structured, with assumed safety and quality benefits. The results of this narrative review suggest that a more formal evaluation of these benefits is warranted.

Chapter 3

Interdisciplinary team care interventions on medical wards: a systematic review

3.1 Introduction

In chapter 2, I described the ambiguous evidence supporting medical ward interdisciplinary collaborative interventions. These team-based interventions form one of the five major ward improvement strategies; they have been associated with numerous improvements in ward care, from reduced length of stay to lower adverse event rates. However, many positive reports relied on weak study designs, unable to properly distinguish chance variation from the effects of an intervention. Complex interdisciplinary interventions may also have unintended consequences, and their ability to bring about positive change may be lessening, as standard care improves.

In other settings, there is evidence that patient outcomes have been improved by interdisciplinary interventions. Interdisciplinary safety checklists in the operating theatre, for example,

The work described in this chapter has been published as: Pannick S, Davis R, Ashrafian H, Byrne BE, Beveridge I, Athanasiou T, Wachter RM, Sevdalis N. Effects of Interdisciplinary Team Care Interventions on General Medical Wards: A Systematic Review. JAMA Intern Med. 2015; 175(8):1288-98. DOI: 10.1001/jamainternmed.2015.2421

reduce surgical morbidity (Bergs et al., 2014; Gillespie et al., 2014; Haugen et al., 2015); and interdisciplinary initiatives in the ICU reduce CLABSIs (Pronovost et al., 2006; Bosk et al., 2009). Interdisciplinary collaborations are increasingly formalised in ward practice, with the expectation of similar improvements in care outcomes and efficiency. The benefits of these interventions on the medical ward have not been systematically assessed, and there is little agreement on how we might best judge their value [see 1.5.3].

In this chapter, I describe a systematic review to determine the impact of interdisciplinary ward team interventions - to my knowledge, the first review of its kind. The systematic search addresses the limitations of the iterative strategy used in the last chapter to categorise the broader evidence base [see 2.5.2]. The review also deliberately focuses on articles using higher quality study designs, given the methodological issues seen in much of the literature [see 2.5.1.3].

3.2 Aims

This study has two aims:

(i) To describe the range of objective patient outcomes used to evaluate general medical ward interdisciplinary team interventions; and

(ii) To summarise the performance of interdisciplinary team interventions against those outcomes.

I defined appropriate interventions as those that targeted interdisciplinary team care (ITC) on the medical ward [box 3.1].

Box 3.1 Interdisciplinary team care (ITC)

*The structured working practices dictating **which** different healthcare practitioners interact together to contribute to patient care, as well as **when** and **how** they do so.*

The definition incorporates staff who routinely attend to the patient (e.g., nurses and ward physicians), as well as those who intermittently provide specialist review and advice.

3.3 Methods

3.3.1 Search strategy & data sources

The research question was structured with the elements of the 'Population, Intervention, Comparison, Outcomes (PICO)' framework (Huang et al., 2006) [see table 3.1 for summary]. The question was framed as: 'For adult patients on general medical wards, what are the effects of ITC interventions on objective patient outcomes, compared to standard care or against other ITC interventions, as evaluated in primary studies with an appropriate design?'

The search strategy defined three components, each of which had to feature within an article's title or abstract, linked by the Boolean 'AND' operator. The three components were: an objective outcome of care, a term to indicate a trial or programme evaluation, and a descriptor to indicate the appropriate setting of a general medical ward. The search terms making up each component are listed in Appendix A.

The EMBASE, Ovid MEDLINE, and PsycINFO databases were then searched for studies meeting these criteria. Limits were applied to restrict matches to those published between 1998 and 2013, in English, applicable to adult humans, and to automate the removal of duplicate results. The reference lists of included studies were hand-searched for other relevant studies.

3.3.2 Article selection: inclusion & exclusion criteria

Specific inclusion and exclusion criteria were developed to refine the basic PICO structure [table 3.2]. RCTs, studies with an appropriate comparison group, and interrupted time series

Table 3.1: Summarised PICO formulation of the systematic review research question

PICO element	Descriptor summary
Population	Adult patients on general medical wards
Intervention	Any interdisciplinary team care intervention delivered during the inpatient admission
Comparison	Standard care, or another eligible intervention
Outcomes	Objective patient outcomes

were included. The criteria and terminology of the Cochrane Effective Practice and Organisation of Care (EPOC) review group (2014) were used to define study types.

Table 3.2: Systematic review inclusion & exclusion criteria

Inclusion criteria	Exclusion criteria
Population	
Adult patients (aged ≥18 years) on general medical wards	Children & adolescents Intensive care units, high dependency or coronary care (level 2) units Operating theatres or stroke units Whole-hospital initiatives not reporting medical patient outcomes or predominance of medical patients
Interventions	
Any intervention to alter ITC delivery, typically including ≥1 of the following components: • Interdisciplinary rounds • Geographic localisation of medical teams • Direct incorporation of allied health professional or subspecialty input into ward teams • Proactive or unsolicited patient management from a health professional not ordinarily part of the ward team	Pharmacotherapy (e.g., a trial of drug *vs* placebo) Medicine reconciliation Substitution of one type of healthcare provider for another (e.g., physician assistants instead of junior doctors) Direct comparison of specialist *vs* general medical care, without other changes to the system of care (e.g., nephrology *vs* internist care for haemodialysis patients, or hospitalist *vs* clinic-based faculty) Case managers / discharge facilitators Changes affecting a single staff group, with no other ITC components (e.g., unidisciplinary education) Patient-focused intervention, directly targeting health behaviours or understanding
Intervention delivered during hospitalisation	Intervention involving significant post-discharge follow-up or outpatient care
Comparators	
Usual care (as defined by the primary studies) Comparison of two eligible interventions	Comparison of an eligible intervention with an ineligible intervention

(Table 3.2, continued)

Inclusion criteria	Exclusion criteria
Outcomes	
Objective outcomes; e.g., mortality rate, length of stay, readmission rate, complications of care, or composites of these	Subjective / patient-reported outcomes Process measures (e.g., guideline compliance, or error reporting rather than objective error rates) Financial outcomes
Publication details	
Published between 1998 & 2013, in English	Published before 1998, after 2013, or in a language other than English
Appropriate study design, such as: • Randomised & non-randomised controlled trials • Prospective cohort studies with a valid comparison group • Interrupted time series	Inappropriate study design, such as: • Case report or case series • Letter to the editor or editorial • Meeting abstract • Case-control study • Retrospective cohort study • Uncontrolled before-after study • Secondary publication, e.g., systematic or non-systematic review

Abbreviations: ITC - interdisciplinary team care.

3.3.3 Screening & data extraction

All retrieved articles underwent title and abstract review. A second reviewer with systematic review expertise[1] independently screened a randomly-selected 15% of the articles. All articles thought to meet the review criteria were examined in full by both reviewers, and the final set of studies to be included was confirmed after discussion with an experienced third reviewer[2]. Quantitative data were extracted with a structured form, devised with an academic clinician with experience in meta-analysis[3]. If the required information was not available in the published report, its authors were contacted for further data or clarification.

Interventions were classified as low, medium, or high intensity. Implementation intensity (sometimes used interchangeably with 'implementation strength') is an emerging concept, defined as a measure of the activity required to support a programme's implementation (Hargreaves et al., 2016). Intensity may mediate the success or failure of the programme (Schellenberg et al., 2012). It can be described with reference to absolute levels of implementation (e.g., hours of contact per week) or relative scales, but is poorly defined: one study considered it the 'aggressiveness' of efforts to deliver the programme (Gold et al., 1993). The components of implementation intensity naturally vary from one programme to the next, but typically include elements of infrastructure (organisational arrangements for service delivery), type of services provided, resources, leadership and coordination, training, and information systems (Schellenberg et al., 2012). There is as yet no consensus or uniform approach in defining and measuring implementation intensity (Schellenberg et al., 2012). In this study, a global grade for implementation intensity took into account factors such as the frequency of the mandated use of the intervention; autonomy of the individuals on the interdisciplinary team to enact recommendations; resource provision; and follow-up.

[1] Rachel Davis, Senior Research Associate
[2] Nick Sevdalis, Professor of Implementation Science & Patient Safety
[3] Hutan Ashrafian, Clinical Lecturer

3.3.4 Assessment of study quality

All included studies were assessed for risk of bias using predefined criteria, based on the Cochrane EPOC group's guidance (Effective Practice and Organisation of Care, 2013). Studies were classed as having low, medium, or high risk of bias, using a scoring system employed previously for heterogeneous healthcare interventions (Davey et al., 2013).

Scores were based on adequate allocation sequence generation; allocation concealment; measurement of baseline outcomes and characteristics; data collection; blinding; protection against contamination; and freedom from selective outcome reporting or other biases. If all criteria were met, the risk of bias was graded as 'low'; if one to three criteria were scored as 'unclear' or 'not done', the risk was 'medium'; if four or more criteria were 'unclear' or 'not done', the risk was 'high'. A second reviewer[4] independently assessed each study, with disagreements resolved by consensus.

3.3.5 Statistical analysis

Interventions were categorised according to their focus on interdisciplinary team composition or practice. Interdisciplinary team composition interventions were subcategorised by the type of specialist input they described. The studies' outcome measures, and the adjustment methods used to account for their patient case-mix, were tabulated. For meaningful analysis, outcomes were grouped into early (occurring within 30 days of receiving the intervention) or late outcomes (31 days to 12 months).

Where appropriate, the associations between studies' intensity ratings and their outcomes were tested (χ^2 analysis; differences were considered significant at $p < 0.05$). The studies were too heterogeneous for summary effect statistics to be wholly reliable. Instead, the available data is presented as an exploratory analysis in weighted bubble charts, with pooled values and 95% confidence intervals (CIs) [figures 3.2, 3.3 and 3.4]. Bubble sizes reflect the sample size

[4] Ben Byrne, Clinical Research Fellow

of each study. The relative risk of outcomes, or weighted mean difference (for length of stay data), was calculated with DerSimonian and Laird random-effects modelling. Between-study and within-study variances contributed to study weighting. Analyses were conducted using Stata, version 12[5].

3.4 Results

3.4.1 Abstract screening & application of the inclusion criteria

A total of 6,934 potentially relevant citations were identified. After excluding abstracts from conferences, there was 100% agreement between the two reviewers on the initial inclusion of a 15% sample of abstracts (1,040 citations). After full text review, 30 studies met the inclusion criteria, 26 of which were identified from the initial database search and four from hand-searching reference lists. The flow of records through the different phases of the review is shown in figure 3.1.

3.4.2 Study characteristics

3.4.2.1 Trial designs & interventions

The 30 included studies comprised eight RCTs,[6] nine cluster RCTs,[7] 8 non-randomised cluster trials,[8] four controlled before-after studies,[9] and one interrupted time series (Auerbach et al., 2012). Study characteristics are summarised in table 3.3. Other than one example (Fine et al., 2003), studies compared an interdisciplinary team care intervention with usual care.

Interventions were grouped into two categories. The first group of interventions altered

[5] Hutan Ashrafian, Clinical Lecturer, assisted with the production of summary effect statistics.

[6] Soifer et al., 1998; Cole et al., 2002; Dey et al., 2005; Mannheimer et al., 2006; Pitkälä et al., 2006; Kircher et al., 2007; Lisby et al., 2010; Lesprit et al., 2013

[7] Solomon et al., 2001; Fine et al., 2003; Camins et al., 2009; Makowsky et al., 2009; O'Leary et al., 2010; Korbkitjaroen et al., 2011; O'Leary et al., 2011a; O'Leary et al., 2011b; Schillig et al., 2011

[8] Curley et al., 1998; Webster et al., 1999; Kucukarslan et al., 2003; Mudge et al., 2006; Manuel et al., 2010; Wald et al., 2011; Singh et al., 2012; Esparza et al., 2013

[9] Lai et al., 2009; Desan et al., 2011; Thanarajasingam et al., 2012; Saint et al., 2013

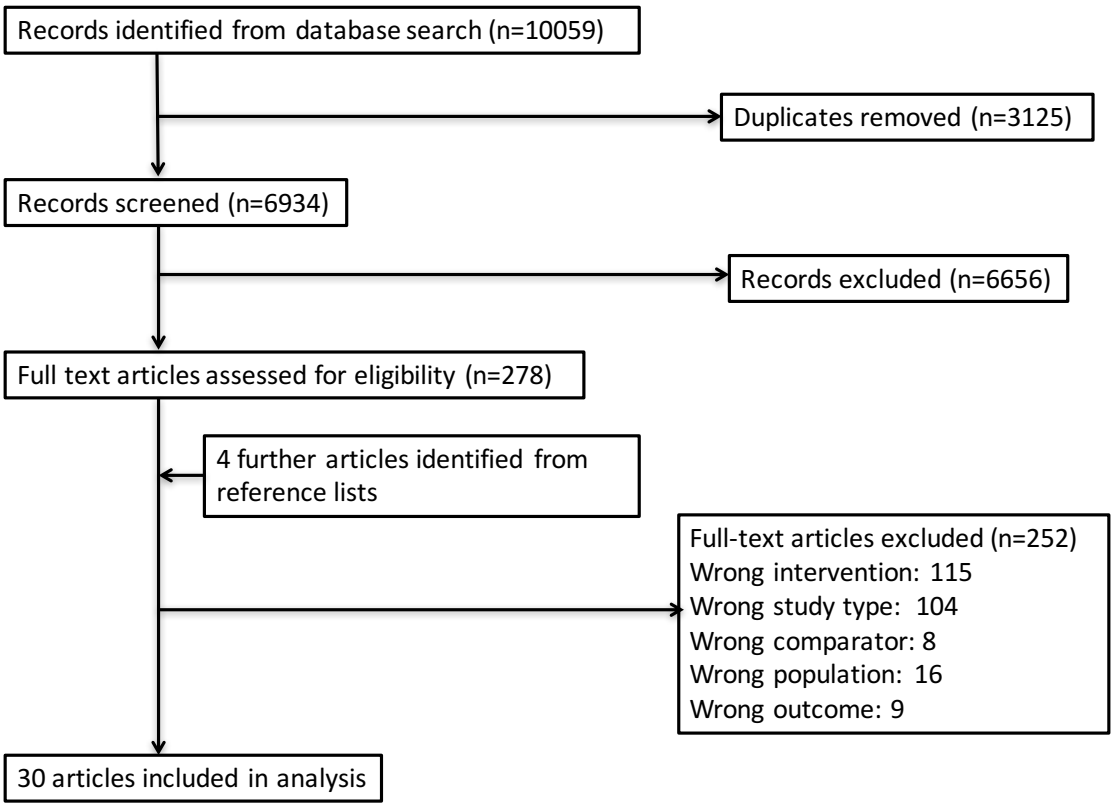


Figure 3.1: Information flow through the systematic review

the composition of the interdisciplinary team routinely attending to the patient, such as by requiring additional specialists or professionals to provide advice. The second group addressed team practice - the logistics of when, where, and how existing team members would work together.

The 30 studies included 66,548 patients with a mean age of 63 years, and a variety of primary diagnoses. Six studies targeted patients with specific diagnoses (delirium,[10] community-acquired pneumonia (Fine et al., 2003), acute stroke (Dey et al., 2005), and advanced liver disease (Lai et al., 2009)), although they were treated on general medical rather than specialist wards. One study evaluated a service for patients taking anticoagulant medication (Schillig et al., 2011).

Twenty-one studies were conducted in North America, eight in other developed countries, and one in an upper-middle-income country. Trials were conducted in a range of settings, including community and urban teaching hospitals, quaternary academic facilities, and Veterans Affairs hospitals. Five were multicentre studies.

Of the 20 studies that addressed interdisciplinary team composition, 15 investigated pro-active, unsolicited consultations from a specialist or specialist team.[11] Consulting professionals specialised in geriatrics (Webster et al., 1999; Cole et al., 2002; Pitkälä et al., 2006; Kircher et al., 2007), infectious diseases,[12] intravenous therapy (Soifer et al., 1998), stroke (Dey et al., 2005), pharmacotherapy (Mannheimer et al., 2006; Lisby et al., 2010; Schillig et al., 2011), or psychiatry (Desan et al., 2011). Four studies assessed the effect of embedding additional health care professionals in rounding teams, incorporating pharmacists (Kucukarslan et al., 2003; Makowsky et al., 2009), medical librarians (Esparza et al., 2013), or supervising medical subspecialists (Lai et al., 2009).

Ten studies investigated interdisciplinary team practice, often using interdisciplinary rounds (Curley et al., 1998; O'Leary et al., 2010; O'Leary et al., 2011a; O'Leary et al., 2011b) or a package

[10] Webster et al., 1999; Cole et al., 2002; Pitkälä et al., 2006

[11] Soifer et al., 1998; Webster et al., 1999; Solomon et al., 2001; Cole et al., 2002; Fine et al., 2003; Mannheimer et al., 2006; Pitkälä et al., 2006; Kircher et al., 2007; Camins et al., 2009; Lisby et al., 2010; Manuel et al., 2010; Desan et al., 2011; Korbkitjaroen et al., 2011; Schillig et al., 2011; Lesprit et al., 2013

[12] Solomon et al., 2001; Fine et al., 2003; Camins et al., 2009; Manuel et al., 2010; Korbkitjaroen et al., 2011; Lesprit et al., 2013

Table 3.3: Systematic review: characteristics of included studies

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Interdisciplinary team composition[a]						
Geriatrics						
Webster et al., 1999 (USA)[b]	Geriatric delirium care team (physician & nurse) proactively identify patients with delirium and guide management	High	LOS; in-hospital death; discharge disposition (not specified)	Non-randomised cluster trial	High	Reduced LOS; no difference in discharge disposition or mortality
Cole et al., 2002 (Canada)	Systematic delirium detection; Geriatician review and specialist nurse follow-up	High	LOS; discharge to the community; death within eight weeks of enrolment (all secondary outcomes)	RCT	Medium	No differences in LOS, discharge to the community, or survival
Pitkälä et al., 2006 (Finland)	Systematic delirium detection, followed by geriatrician review and specialist nurse follow-up; no definitive description of who delivered the intervention, or how	Low	Patients discharged to permanent institutional care or deceased at one year (composite primary outcome) Permanent institutional care at one year; acute hospitalisation days during delirium episode; total number of days spent in any institution and in permanent institutional care during the follow-up year; days spent in community care before permanent institutional care; one-year mortality (secondary outcomes)	RCT	Medium	No difference in primary outcome or any secondary outcome

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Kircher et al., 2007 (Germany)	Geriatrician, social worker and nurse consultation service; geriatrician summarised recommendations in a structured format; input from nurse and social worker as required	Medium	Living location; proportion of patients with ≥1 readmission; days of readmission at 12 months (all described as primary outcomes) 12-month survival; Timed-Up-And-Go-Test; hand grip (secondary outcomes)	RCT	High	No difference in any outcome
Hepatology						
Lai et al., 2009 (USA)	Multidisciplinary hepatology team; Direct supervision of a specialised housestaff team structure	High	LOS (primary outcome)	Controlled before-after study	High	Adjusted LOS reduced with intervention (when adjusted for CMI and MELD score); concurrent control group's LOS increased over the study period

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Infectious disease						
Solomon et al., 2001 (USA)	Multidisciplinary AUT (clinician-educator/ infectious disease physician/ clinical pharmacist); Systematic screening for unnecessary antibiotic prescriptions, followed by educational intervention to prescriber; 'Academic detailing' (one-on-one interactive educational outreach, delivered by a professional trained to discuss prescribing decisions in a manner likely to induce evidence-based practice change)	Medium	LOS; 30-day readmissions; in-hospital deaths; transfer to the ICU (all secondary outcomes)	Cluster randomised trial	High	No difference in any outcome (N.B.: intervention was not designed to affect these outcomes *per se*; they were used as balancing measures)
Fine et al., 2003 (USA)	Specialist nurse input to facilitate antibiotic management; Screening for patient stability sufficient to allow conversion from IV to oral antibiotic therapy, or discharge; Direct communication with primary team to discuss recommendations	Medium	LOS (primary outcome) 30-day mortality; medical complications; 30-day readmissions (secondary outcomes)	Cluster randomised trial	High	No change in LOS; fewer medical complications in intervention group; no differences in mortality or readmission

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Camins et al., 2009 (USA)	Multidisciplinary AUT (infectious disease physician & specialist pharmacist); Systematic screening for unnecessary antibiotic prescriptions, followed by academic detailing	Low	LOS; in-hospital mortality (secondary outcomes)	Cluster randomised trial	High	Shorter median LOS; no difference in mortality
Manuel et al., 2010 (Switzerland)	Infectious disease specialist; Systematic antibiotic prescription assessment with regard to need, choice of drug, route and dose; Feedback to treating physician	Low	LOS; in-hospital mortality (secondary outcomes)	Non-randomised cluster trial (with crossover)	High	No difference in either outcome
Korbkitjaroen et al., 2011 (Thailand)	Multidisciplinary infection control team (infectious disease physician & infection control nurse) identifying risk factors for hospital-acquired infection; Feedback to the treating team, including observations of adherence to infection control measures	High	LOS; in-hospital mortality; hospital-acquired infection rates (not specified)	Cluster randomised trial	High	Reduced hospital-acquired infection rate (driven by reductions in pneumonia and CAUTI) No change in LOS or mortality

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Lesprit et al., 2013 (France)	Infectious disease physician; Systematic antibiotic prescription assessment; Feedback to the treating physician	Low	LOS (primary outcome) ICU admission; in-hospital mortality; 60-day readmission for relapsing infection (secondary outcomes)	RCT	Medium	No change in LOS, in-hospital mortality or ICU admission. Fewer 60-day readmissions for relapsing infection
IV therapy						
Soifer et al., 1998 (USA)	IV therapy (nursing) team; Insertion and/or management of peripheral IV catheters	High	IV line-related bacteraemia; local complications from catheter insertion (not specified)	RCT	High	Reduced catheter-related bacteraemias and local complications
Medical librarian						
Esparza et al., 2013 (USA)	Clinical medical librarian embedded in medical rounds to answer specific questions posed by team members; Detailed answers supplied by email after rounds	Low	LOS; 30-day readmissions (not specified)	Non-randomised cluster trial	High	Higher readmission rates and LOS in intervention group
Pharmacotherapy						
Kucukarslan et al., 2003 (USA)	Clinical pharmacist; Participation in rounding and prospective evaluation of patient medications	High	Preventable adverse drug events (primary outcome) LOS; time to respond to therapy; readmission rate (secondary outcomes)	Non-randomised cluster trial	High	Fewer preventable adverse drug events/1000 patient-days No change in LOS, time to respond to therapy, or readmission rate

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Mannheimer et al., 2006 (Sweden)	Clinical pharmacologist & specialist nurse; Systematic screening for potential drug interactions contributing to patient symptoms; Feedback to the treating physician	Low	Readmission and/or death at 6 months (primary outcome) 6-month readmissions; 6-month deaths (secondary outcomes)	RCT	High	No difference in any outcome
Makowsky et al., 2009 (Canada)	Clinical pharmacist; Participation in rounding and provision of proactive clinical services	High	3-month readmissions; 6-month readmissions; LOS (secondary outcomes)	Cluster randomised trial (with crossover)	High	Reduced 3-month readmissions; no effect on 6-month readmissions Increased LOS in intervention group
Lisby et al., 2010 (Denmark)	Clinical pharmacologist & clinical pharmacist; Systematic medication review; Advisory feedback to treating physician	Low	LOS (primary outcome) 3-month readmissions; 3-month mortality; time to first readmission; number of readmissions; 3-month emergency department visits, 3-month outpatient care visits (secondary outcomes)	RCT	Medium	No difference in any outcome
Schillig et al., 2011 (USA)[c]	Pharmacist-directed anticoagulation management	High	Composite outcome of INR >5, major bleeding, or development of new thrombosis in-hospital or within 30 days of discharge (primary safety outcome)	Cluster randomised trial	High	No difference in composite outcome or its components

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Psychiatry						
Desan et al., 2011 (USA)	Psychiatrist; Proactive identification of patients with mental health issues; Early psychiatric consultation; Coordination with primary team	High	LOS (primary outcome)	Controlled before-after study	High	Reduced LOS; reduced proportion of cases with LOS >4 days
Stroke						
Dey et al., 2005 (UK)	Multidisciplinary stroke team (consultant & senior therapist); Advised primary team on acute stroke management; Coordinated early input from therapy groups	Low	6-week mortality; 12-week mortality (primary outcomes) Death or institutionalised care (secondary outcome)	RCT	Medium	No difference in any outcome
Interdisciplinary team practice[d]						
Curley et al., 1998 (USA)	IDRs including physician, nurse, pharmacist, nutritionist & social worker; Orders written during rounds	High	LOS; in-hospital mortality (not specified)	Non-randomised cluster trial	High	Reduced LOS; no difference in in-hospital mortality

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Mudge et al., 2006 (Australia)	Consistent multidisciplinary teams aligned with admitting medical units; Expanded senior clinical nurse role, with structured assessment of patient's premorbid function; Investment in allied health professional staffing to allow a consistent staff member from each discipline in each intervention unit; Early discharge planning	High	LOS; in-hospital mortality; 6-month mortality (primary outcomes) 6-month readmissions; discharge to residential care (secondary outcomes)	Non-randomised cluster trial	High	Reduced in-hospital mortality; No change in 6-month mortality, LOS, 6-month readmissions or discharge to residential care
O'Leary et al., 2010 (USA)	Structured IDRs & structured communication tool for newly admitted patients; Attended by nurses, resident physicians, pharmacist, social worker, & case manager	High	LOS (primary outcome)	Cluster randomised trial	High	No difference in outcome
O'Leary et al., 2011a (USA)	Structured IDRs & structured communication tool for newly admitted patients; Attended by nurses, resident physicians, pharmacist, social worker, & case manager	High	Adverse events (primary outcome)	Cluster randomised trial	High	Fewer patients experienced adverse events; reduced adverse event rates; reduced preventable adverse event rates

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
O'Leary et al., 2011b (USA)	Structured IDRs & structured communication tool for newly admitted patients; Attended by nurses, resident physicians, pharmacist, social worker, & case manager	High	LOS (secondary outcome)	Cluster randomised trial	High	No difference in outcome
Wald et al., 2011 (USA)[e]	IDRs, attended by nurses, physicians, social workers, case managers, physical/ occupational therapists, pharmacists, & volunteers; Standardised geriatric assessment; Focus on safety & early discharge planning; Educational curriculum for geriatric care	High	LOS; falls; 30-day readmissions; discharge location (secondary outcomes)	Non-randomised cluster trial	High	No difference in any outcome
Auerbach et al., 2012 (USA)[f]	Multidisciplinary teamwork training curriculum & unit safety teams	Medium	LOS; 30-day readmissions (secondary outcomes)	Interrupted time series	Medium	No difference in any outcome
Singh et al., 2012 (USA)	Localisation of each medical team's patients to a single nursing unit	Medium	LOS; 30-day readmissions (not specified) Inpatient deaths described as patient characteristic rather than outcome	Non-randomised cluster trial	High	No difference in any outcome

(Table 3.3, continued)

Author, Year (Country)	Intervention	Intensity	Objective patient outcomes (primary or secondary)	Trial design	Risk of bias	Results
Thanarajasingam et al., 2012 (USA)	Localisation of each medical team's patients to a single nursing unit Census cap of 14 patients/service	High	30-day readmissions; cardiac arrests; rapid response team events; ICU transfers; adverse events as defined by AHRQ patient safety indicators (not specified)	Controlled before-after study	High	Reduced 30-day readmissions, but not meeting the predefined $p < 0.01$ threshold for multiple comparisons; No change in rapid response team events, cardiac arrests, ICU transfers or complications of care
Saint et al., 2013 (USA)	IDRs, attended by doctors, nurses, pharmacist & clinical care coordinator; Modified attending rounds; Regular attending team meetings; Increased attending commitment to inpatient service provision; Educational curriculum for attendings and residents	High	LOS; 7-day readmissions; 30-day readmissions (not specified)	Interrupted time series with control cohort	Medium	No difference in any outcome

Abbreviations: AHRQ - Agency for Healthcare Research and Quality; AUT - antimicrobial utilisation team; CAUTI - catheter-associated urine tract infection; CMI - Case Mix Index; ICU - intensive care unit; IDRs - interdisciplinary rounds; IV - intravenous; LOS - length of stay; MELD - Model for End Stage Liver Disease; RCT - randomised controlled trial.

[a]Interventions to alter the composition of the interdisciplinary team routinely attending the patient, through proactive and/or unsolicited involvement of medical specialists or allied health care professionals. [b]Results from phase 2 of the study relating to interdisciplinary team care intervention. [c]Length of stay excluded from analysis owing to changes over the course of the intervention (expansion of an advanced heart failure service). [d]Interventions to alter the practice of the established interdisciplinary team in terms of where, when, and how staff work together. [e]Authors argue that their intervention is not a typical acute care of the elderly unit but a hybrid general medical service (N.B.: no environmental changes or separate nursing training or protocols). [f]Results from phases 1 and 2 of the study, excluding the patient-targeted interventions of phase 3.

of interventions including interdisciplinary rounds (Mudge et al., 2006; Wald et al., 2011; Saint et al., 2013). Two studies assessed team localisation, with medical and nursing staff co-positioned in the same geographic area in the hospital (Singh et al., 2012; Thanarajasingam et al., 2012). One study described an interdisciplinary teamwork and communication programme (Auerbach et al., 2012).

3.4.2.2 Objective patient outcome measures

Of the 30 studies, 19 (63%) reported length of stay, readmission, or mortality as their primary outcome measure, or did not specify the primacy of their outcome measures. The most commonly reported objective patient outcome measures were length of stay (23 of 30 [77%]), complications of care (10 of 30 [33%]), in-hospital mortality (eight of 30 [27%]), and 30-day readmission rate (eight of 30 [27%]). Seven of nine studies (78%) that reported mortality after discharge used time points later than 30 days.[13] Seven of 30 studies (23%) reported requirements for institutional care either at discharge or subsequently.[14] One of 30 studies (3%) reported patients' use of community health care services following discharge (Lisby et al., 2010), and one of 30 studies (3%) reported objective assessments of functional status (Kircher et al., 2007). Studies reported a mean of 3.1 objective patient outcome measures.

Several studies did not seek improvements in these outcomes *per se*, targeting a different aspect of care. Interventions involving infectious disease specialists, for example, were often restrictive in nature, advocating narrower-spectrum antibiotic choices, or oral rather than intravenous therapy (Solomon et al., 2001; Fine et al., 2003; Camins et al., 2009; Lesprit et al., 2013). Those studies sought only to show that their intervention had no detrimental effect on patient outcomes, aiming instead to demonstrate improvements in processes of care.

3.4.2.3 Risk of bias and intervention intensity

No study had a low risk of bias; seven (23%) had a medium risk of bias,[15] and the remaining 23 (77%) a high risk of bias. Eight studies (27%) involved low-intensity interventions; the remainder were of medium or high intensity.

[13] Cole et al., 2002; Dey et al., 2005; Mannheimer et al., 2006; Mudge et al., 2006; Pitkälä et al., 2006; Kircher et al., 2007; Lisby et al., 2010

[14] Webster et al., 1999; Cole et al., 2002; Dey et al., 2005; Mudge et al., 2006; Pitkälä et al., 2006; Kircher et al., 2007; Wald et al., 2011

[15] Cole et al., 2002; Dey et al., 2005; Pitkälä et al., 2006; Lisby et al., 2010; Auerbach et al., 2012; Lesprit et al., 2013; Saint et al., 2013

3.4.3 Outcome risk adjustment and analysis

Of the 30 studies, 16 (53%) did not adjust their results for their patient case mix.[16] Eleven of 30 studies (37%) used patient characteristics as independent variables in a multivariate analysis to isolate the effect of their intervention.[17] Two of 30 studies (7%) adjusted results for predicted outcomes based on diagnosis-related groups (Webster et al., 1999; Thanarajasingam et al., 2012), and two of 30 studies (7%) used a disease-specific severity index (Fine et al., 2003; Lai et al., 2009). Only seven of 30 studies (23%) accounted for autocorrelated outcomes, in which patients treated by the same physician (or team) are more likely to have similar results than those who are not.

3.4.4 Complications of care and discharge location

Of 10 interdisciplinary team care interventions, five (50%) reduced complications of care.[18] Three of those five harnessed input from infectious diseases specialists or specialist nurses (Soifer et al., 1998; Fine et al., 2003; Korbkitjaroen et al., 2011). Both studies that reported preventable adverse events suggested improvement, either through embedding pharmacists in the clinical team (Kucukarslan et al., 2003) or formalising interdisciplinary rounds (O'Leary et al., 2011a).

Four of the five interventions that showed reduced complications of care were rated high intensity (Soifer et al., 1998; Kucukarslan et al., 2003; Korbkitjaroen et al., 2011; O'Leary et al., 2011a), but there was no apparent relationship between high-intensity categorisation and reduced complications (χ^2 = 0.476; p = 0.49). Zero of seven interventions successfully reduced the need for institutional care after discharge; four of those seven were rated high intensity (Webster et al., 1999; Cole et al., 2002; Mudge et al., 2006; Wald et al., 2011).

[16] Solomon et al., 2001; Cole et al., 2002; Kucukarslan et al., 2003; Dey et al., 2005; Mannheimer et al., 2006; Mudge et al., 2006; Pitkälä et al., 2006; Camins et al., 2009; Lisby et al., 2010; Manuel et al., 2010; Desan et al., 2011; Korbkitjaroen et al., 2011; Schillig et al., 2011; Wald et al., 2011; Esparza et al., 2013; Lesprit et al., 2013

[17] Curley et al., 1998; Soifer et al., 1998; Kircher et al., 2007; Lai et al., 2009; Makowsky et al., 2009; O'Leary et al., 2010; O'Leary et al., 2011a; O'Leary et al., 2011b; Auerbach et al., 2012; Singh et al., 2012; Saint et al., 2013

[18] Soifer et al., 1998; Fine et al., 2003; Kucukarslan et al., 2003; Korbkitjaroen et al., 2011; O'Leary et al., 2011a

3.4.5 Length of stay

Saint et al. (2013) documented significant secular reductions in length of stay. In this context, 16 of 23 interventions (70%) had no effect on the length of stay;[19] five of 23 (22%) reduced the length of stay;[20] and one (4%) may have increased the length of stay (Makowsky et al., 2009). A study that associated clinical librarian input with increased length of stay was too heavily confounded by patient group differences to support this conclusion, as acknowledged by its authors (Esparza et al., 2013).

Four of five effective interventions (80%) incorporated subspecialist input to the primary general medical ward team (Webster et al., 1999; Camins et al., 2009; Lai et al., 2009; Desan et al., 2011). Again, four of five effective interventions were of high intensity (Webster et al., 1999; Curley et al., 1998; Lai et al., 2009; Desan et al., 2011), without a significant overall correlation between intensity and effect on length of stay (χ^2 = 1.433; p = 0.49).

Of the studies that failed to demonstrate a statistically significant change in length of stay, nine of 16 (56%) reported increased length of stay with their intervention.[21]. Singh et al. (2012) explained how this result could be an unintended consequence of localisation interventions. However, the exploratory quantitative summary did not detect a significant change in length of stay overall. The weighted mean difference in length of stay for interdisciplinary team composition interventions was 0.087 (95% CI -0.083 to 0.257). For interdisciplinary team practice interventions, the weighted mean difference in length of stay was 0.001 (95% CI -0.035 - 0.037) [figure 3.2].

[19] Solomon et al., 2001; Cole et al., 2002; Fine et al., 2003; Mudge et al., 2006; Pitkälä et al., 2006; O'Leary et al., 2010; Korbkitjaroen et al., 2011; O'Leary et al., 2011b; Wald et al., 2011; Auerbach et al., 2012; Saint et al., 2013

[20] Curley et al., 1998; Webster et al., 1999; Camins et al., 2009; Lai et al., 2009; Desan et al., 2011

[21] Cole et al., 2002; Pitkälä et al., 2006; Lisby et al., 2010; Manuel et al., 2010; O'Leary et al., 2010; O'Leary et al., 2011b; Wald et al., 2011; Singh et al., 2012; Saint et al., 2013

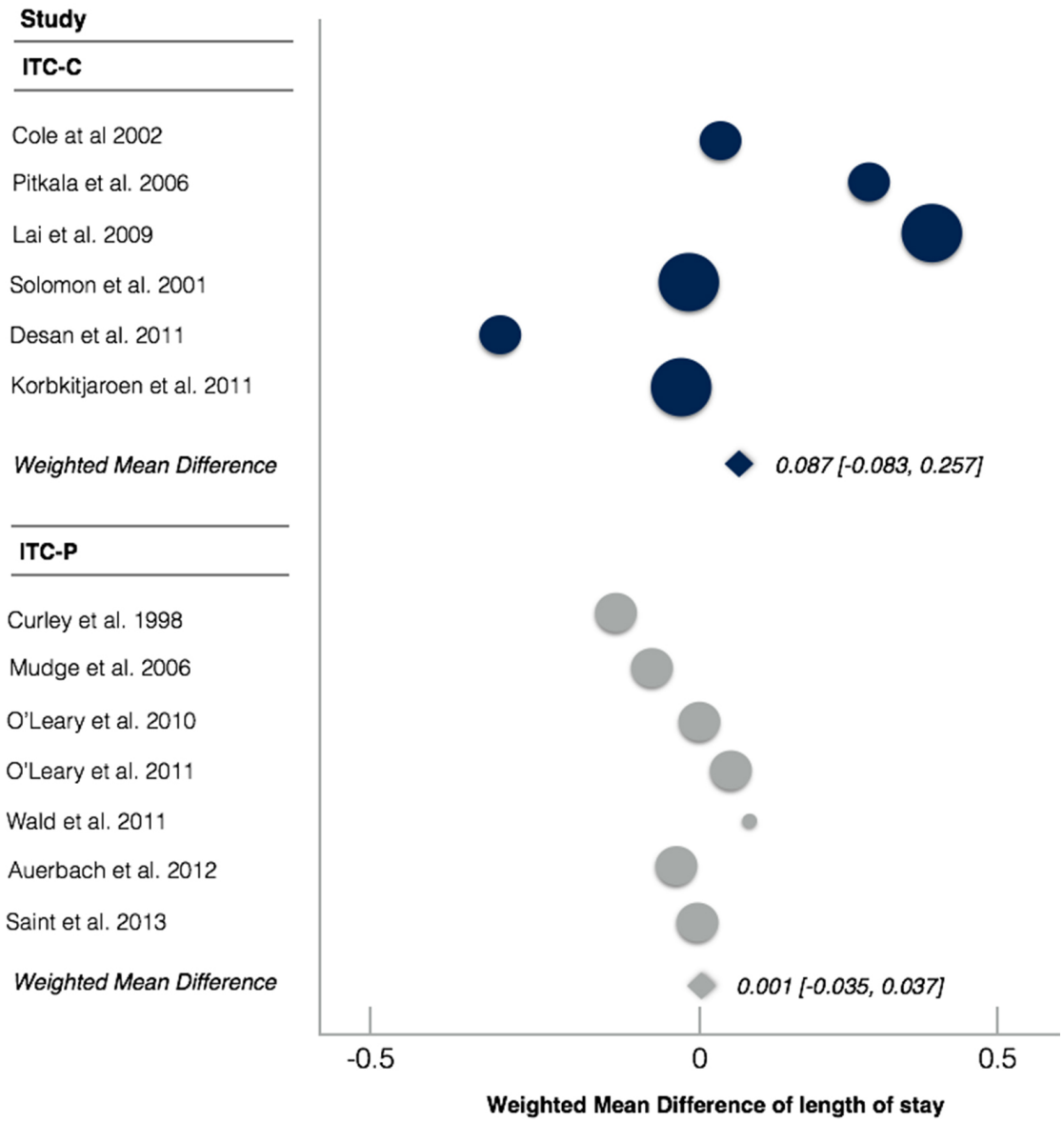


Figure 3.2: Effect of ITC interventions on length of stay

The figure's exploratory analysis indicates no significant effect of interdisciplinary team care interventions on length of stay. This was the case for each of the two subcategories: interventions altering interdisciplinary team composition (ITC-C; weighted mean difference 0.087; 95% CI -0.083 - 0.257) and interventions altering interdisciplinary team practice (ITC-P; weighted mean difference 0.001; 95% CI -0.035 - 0.037).

3.4.6 Readmissions and use of outpatient care

Twelve of 15 interventions (80%) did not reduce readmissions. One study (7%) reported readmissions at seven days (Saint et al., 2013); eight studies (53%) reported readmissions at 30 days;[22] five studies (33%) reported readmissions at points between 60 days and six months;[23] and one study (7%) reported readmissions at one year (Kircher et al., 2007). One study reported both three- and six-month readmissions (Makowsky et al., 2009).

One of three infectious disease interventions (33%) reduced 60-day readmissions related to relapsing infection (Lesprit et al., 2013); the other two similar interventions did not affect readmissions, albeit when measured at 30 days (Solomon et al., 2001; Fine et al., 2003). One of four pharmacy interventions reduced admissions (at three months), although the effect had dissipated by six months (Makowsky et al., 2009).

Zero of six interdisciplinary team practice interventions reduced readmissions.[24] One study suggested reduced readmissions after localising teams and introducing service census limits, but the results did not meet the predefined statistical threshold allowing for multiple comparisons (Thanarajasingam et al., 2012). Increased readmissions in the study by Esparza et al. (2013) were again attributed to confounding factors. Of the two interventions that reduced readmissions, one was of low intensity (Lesprit et al., 2013) and the other was of high intensity (Makowsky et al., 2009).

An exploratory analysis showed interdisciplinary team care interventions had an inconsistent effect on early readmissions. Team composition interventions tended to increase early readmissions, albeit with important confounding factors in the included studies: weighted risk ratio (RR) 1.341 (95% CI 1.120 - 1.607). Team practice interventions did not significantly reduce early readmissions: weighted RR 0.995 (95% CI 0.912 - 1.085) [figure 3.3].

[22] Solomon et al., 2001; Fine et al., 2003; Wald et al., 2011; Auerbach et al., 2012; Singh et al., 2012; Thanarajasingam et al., 2012; Esparza et al., 2013; Saint et al., 2013

[23] Mannheimer et al., 2006; Mudge et al., 2006; Makowsky et al., 2009; Lisby et al., 2010; Lesprit et al., 2013

[24] Mudge et al., 2006; Wald et al., 2011; Auerbach et al., 2012; Singh et al., 2012; Thanarajasingam et al., 2012; Saint et al., 2013

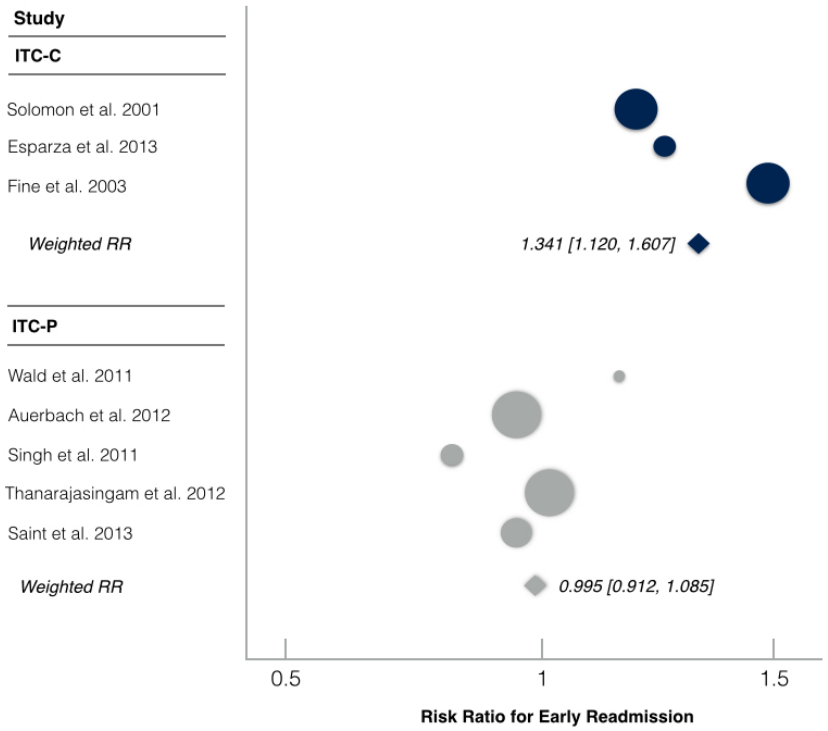


Figure 3.3: Effect of ITC interventions on early readmissions

The figure's exploratory analysis indicates no consistent effect of interdisciplinary team care interventions on early readmissions. Interdisciplinary team composition interventions (ITC-C) tended to increase early readmissions, albeit with important confounding factors in the included studies as described in the text (weighted risk ratio [RR] 1.341; 95% CI 1.120 - 1.607). Interdisciplinary team practice interventions (ITC-P) did not significantly reduce early readmissions (weighted RR 0.995; 95% CI 0.912 - 1.085).

Last, one study reported contact with primary health care (ambulatory visits to medical specialists, family physicians, and the emergency department) (Lisby et al., 2010) and one reported objective functional measures at 12 months (Kircher et al., 2007). The interdisciplinary team interventions had no effect on these outcomes.

3.4.7 Mortality

Only one of 15 (7%) studies that reported mortality rate showed a significant effect at any time point. Mudge et al. (2006) were the only group to describe a sustained increase in allied health care professional resourcing (physiotherapy, occupational therapy, social work, dietetics, and speech therapy) alongside an interdisciplinary structure to make use of those staff. Their intervention was rated high intensity. However, the observed reduction in in-hospital mortality did not persist at six months.

Team practice interventions tended to reduce early mortality, although only two studies contributed data: weighted RR 0.665 (95% CI 0.449 - 0.986). Interdisciplinary team composition interventions did not significantly reduce early mortality: weighted RR 0.925 (95% CI 0.816 - 1.049) [figure 3.4].

3.5 Discussion

This systematic review summarises the effects of ITC interventions on medical wards. For the first time (to my knowledge), a defined set of exclusion criteria were used to identify relevant studies with higher quality designs. It was important to stratify the evidence base in this way, describing these reports separately, given that so much of the ITC literature stems from low quality, uncontrolled studies. With an increasing emphasis on interdisciplinary care delivery, it was important to establish the likely true benefit of this type of intervention. Additional insight into the most relevant metrics for medical ward care was also required. I will now highlight the principal findings of the review.

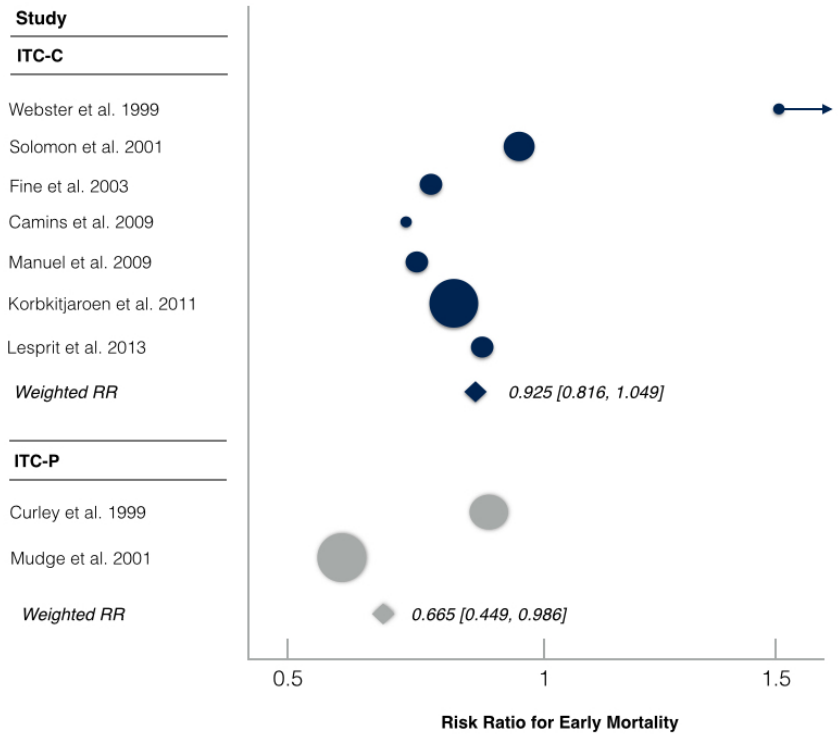


Figure 3.4: Effect of ITC interventions on early mortality

The figure's exploratory analysis indicates no significant reduction in early mortality with interdisciplinary team composition interventions (ITC-C; weighted risk ratio [RR] 0.925; 95% CI 0.816 - 1.049). The two interdisciplinary team practice interventions (ITC-P) tended to reduce early mortality (weighted RR 0.665; 95% CI 0.449 - 0.986). The arrow marker for the study by Webster et al indicates that the 95% CI for the RR (2.167) was out of the range of the plot.

3.5.1 Summary of findings

3.5.1.1 Characteristics of included studies

The majority of the 30 included studies were RCTs (randomised at the patient level), cluster RCTs (randomised at the level of the physician, team or ward), or non-randomised cluster trials. They described interventions changing interdisciplinary team composition (the make-up of the team routinely attending to the patient) and interdisciplinary team practice (the logistics of how the interdisciplinary team worked together). In contrast to many trials evaluating the efficacy of a drug or device, the studies included here had relatively few exclusion criteria.

Included studies took place in a variety of settings, from safety-net hospitals to large academic facilities, and across numerous countries with disparate health care systems. The findings of this review are therefore generally applicable to a broad range of medical patients internationally. However, the quality of the available evidence was relatively poor. Despite restricting the search strategy to higher quality study designs, 77% of the included studies were graded at high risk of bias, the remainder having a medium risk of bias. Only a minority (23%) accounted for the clustering of patient results by ward, an important statistical adjustment to reduce the risk of type 1 (false positive) errors.

3.5.1.2 The objective outcomes used to evaluate medical ward ITC interventions

ITC evaluations most commonly chose length of stay, complications of care, in-hospital mortality and 30-day readmission rate as their primary outcome measures. Length of stay was the most popular choice of outcome, identified in 77% of studies. Most other outcome metrics similarly described the immediate inpatient period (e.g., in-hospital mortality and complications of care). However several studies anticipated an effect on post-discharge and even late post-discharge outcomes, frequently reporting results at time points beyond 30 days.

Outcomes identified here were used both as the primary target for interventions, and as balancing measures to evaluate the unintended consequences of interventions with other primary

aims. Each study reported, on average, 3.1 objective patient outcomes. Only 47% of the studies adjusted their results for their patients' characteristics, with a variety of risk adjustment techniques.

3.5.1.3 The performance of ITC interventions against commonly chosen outcomes

Most ITC interventions do not affect length of stay, readmissions, or mortality rates. There is some evidence that improvements in interdisciplinary collaboration may reduce preventable complications of care and adverse events. One study described a significant contemporaneous secular reduction in length of stay, which these interventions did not reduce further.

Interdisciplinary interventions confined to the inpatient setting are unlikely to reduce readmissions, the need for institutional care after discharge, or mortality rates. However, simultaneous investments in staffing and team structures may reduce mortality rates in the short term. Where interventions did report improvement, it appeared that high-intensity delivery was a necessary pre-requisite, but was not in itself sufficient, for the desired impact.

3.5.2 Study limitations

The heterogeneity of studies included in this review limited the quantitative analysis, which could only be described as exploratory. Many of those studies had notable methodological limitations, increasing their risk of bias. The review also included controlled before-after studies with only one intervention or control site, which are particularly prone to confounding (Effective Practice and Organisation of Care, 2014). However, this review reflected the best published evidence currently available. Including a sufficiently broad range of studies is helpful, facilitating a deeper understanding of context-specific improvement strategies than relying on RCTs alone (Zwarenstein et al., 2009; Shojania, 2013).

The search strategy may not have been fully sensitive to all relevant articles: only 15% of retrieved articles were reviewed by more than one author. This is an acceptable degree of oversight for the field, particularly as all retrieved articles underwent title and abstract review,

rather than title review alone. Moreover, the two reviewers were in complete agreement as to the inclusion of articles for full text review. This suggests that the inclusion and exclusion criteria were sufficiently well defined to be applied consistently and reproducibly. There was no assessment of publication bias or selective reporting, and not all the authors contacted for further data replied. Given the tendency in QI literature to over-report positive results, this is unlikely to have substantially affected the review's overall findings.

3.5.3 Implications for further work

Study investigators have preferred outcome measures that can be extracted from administrative data: mortality rates, length of stay, readmission, and so on. Whilst easy to obtain, these have proved insensitive to changes in care. This is a common failing of administrative data. It is important to note it here, as ITC interventions were assumed to be effective when judged against these metrics. Administrators may have relied too heavily on changes in ITC to achieve reductions in length of stay, for example, neglecting more necessary changes in infrastructure or staffing. At the same time, the selection of inappropriate metrics may have the unintended consequence of eroding team function, setting unattainable targets that can demotivate team members (Mosher and Kaboli, 2015).

More work is needed to define valid measures of medical ward performance, and a core data set that all medical ward studies should report (Koroshetz, 2015). The patient perspective on this issue is explored in chapter 9. It is likely that additional efforts will be required to collect nuanced data, that better reflect ward quality. Preventable complications of care, or unnecessary length of stay, are possible examples of relevant - if resource-intensive - metrics. A modification of length of stay (excess length of stay) is used as the primary outcome in the evaluation of a novel interdisciplinary intervention in chapter 4.

Evidence from other clinical settings also suggests that collaborative, interdisciplinary interventions may not affect patient outcomes (Nielsen et al., 2007; Zwarenstein et al., 2009; Ven, 2012; Weaver et al., 2013; Reames et al., 2015). These negative results underscore the importance

of appropriate study design: findings from higher quality studies in this field differ from the conclusions of a broader evidence base. The selection criteria of systematic reviews play a key role: more permissive criteria may generate very different conclusions (Bhamidipati et al., 2016).

Investigators and reviewers should place greater emphasis on the choice of an appropriate evaluation methodology. Health service research could make greater use of the stepped wedge trial design, which has several pragmatic and statistical advantages over other approaches (Hemming et al., 2015a). However, ITC interventions are - by their nature - complex interventions, targeting a number of organisational levels with multiple interacting components, and are potentially difficult to implement (Craig et al., 2008). The Medical Research Council recommends that any assessment of their outcomes should be accompanied by a process evaluation (Moore et al., 2015). This clarifies the causal assumptions behind the intervention and how it was delivered in practice, to examine whether its effects varied with implementation quality. The process evaluation should extend to an analysis of the implementation context, which may shape the theory of how the intervention works, influence its outcomes, and be influenced by them in turn (Moore et al., 2015).

The next chapters apply these lessons to the evaluation of a novel, complex interdisciplinary intervention. Chapter 4 describes the design of a stepped wedge trial; chapter 5 the application of quantitative process data to identify varying intervention effectiveness with degrees of implementation; chapter 6 a qualitative analysis of the intervention's mechanisms of action; and chapter 7 its implementation context [see 7.1.2 for more theoretical detail on context].

3.5.4 Conclusions

ITC is a cornerstone of modern medical practice, incorporated into standard care. However, interventions to improve ITC have not proven effective in improving objective outcomes. Possible contributory factors include ceiling effects (high quality interdisciplinary collaboration at baseline) (O'Leary et al., 2016), ineffective interventions, and insensitive outcome metrics. In part III, I will describe the design and evaluation of a collaborative interdisciplinary intervention,

informed by these findings, and hoping to overcome some of their associated methodological challenges.

Part III

Prospective clinical team surveillance for medical wards

Chapter 4

Prospective clinical team surveillance: a programme theory & study protocol

4.1 Introduction

I showed in part II [chapters 2 and 3] that strategies to improve medical ward care have not been underpinned by universally high quality evidence. In particular, ward level interdisciplinary interventions may not improve patient outcomes, particularly when those interventions are rigorously evaluated. There are notable exceptions. Some multifaceted improvement strategies, properly assessed with appropriate study designs, have improved outcomes (Beckett et al., 2013; McAlister et al., 2014). However, these wide-ranging interventions may require substantial organisational change, which is a barrier to their wider dissemination. A simpler intervention with wider applicability is required, together with a sufficiently comprehensive evaluation to support its adoption outside of the trial setting.

The work described in this chapter has been published as: Pannick S, Beveridge I, Ashrafian H, Long SJ, Athanasiou T, Sevdalis N. A stepped wedge, cluster controlled trial of an intervention to improve safety and quality on medical wards: the HEADS-UP study protocol. BMJ Open. 2015; 5:e007510. DOI: 10.1136/bmjopen-2014-007510

Earlier, I noted that a limited number of interventions have sought to improve foundational aspects of ward care, rather than the management of specific clinical conditions [see 1.2.5.1]. This is an attractive approach, if largely unproven. Here, I develop a novel intervention to augment one underpinning facet of ward work: the use of frontline staff knowledge to improve care quality. I begin by considering why frontline knowledge of service provision is so important, and some of the barriers to the effective escalation and use of staff concerns. I then develop a programme theory of how a ward-level intervention could better translate this frontline knowledge into improved processes and patient outcomes. Finally, I describe the piloting of the intervention, and a study protocol to more definitively assess its impact.

4.2 Aims

The principal aims of this chapter are:

(i) To develop an interdisciplinary intervention to improve the use of frontline staff knowledge, with a programme theory for its effects;

(ii) To pilot the intervention, and establish its feasibility; and

(iii) To describe a study protocol for the structured implementation and evaluation of the intervention in two institutions.

4.3 Frontline ward staff: an important, underused, source of information

4.3.1 Frontline reflections on care delivery correlate with patient survival

Frontline staff are the clinical and non-clinical employees who are directly involved in the delivery of ward care and its operational support. They are a relatively untapped source of information about their own organisation and the effectiveness of its procedures and processes.

Few tools systematically capture staff knowledge to successfully improve clinical and operational processes (Stavropoulou et al., 2015). Centralised incident reporting systems, for example, have major failings, and physicians as a group do not engage with them (Farley et al., 2008; Shojania, 2008). Other strategies are required to detect and address problems (Olsen et al., 2007).

Efforts to capitalise on the frontline experience of daily care delivery will be worthwhile. Staff perceptions correlate with 'hard' outcomes, including patient survival and measures of 'harm-free' care (Tvedt et al., 2014; Lawton et al., 2015). Eliciting these perceptions might itself facilitate positive change: organisations that deliberately seek out discomfiting insights, consciously listening to their staff, develop more holistic improvement strategies (Dixon-Woods et al., 2014a).

4.3.2 Frontline staff default to first-order problem solving, hindering organisational improvement

From an organisational viewpoint, barriers to the delivery of high quality care often remain hidden. Staff do not typically volunteer information about lapses in care quality, nor the problems they encounter that hinder the delivery of high quality care. Operational issues are commonplace, but the vast majority of them are immediately addressed at the frontline, the staff member finding a way to compensate for the problem and complete the task that had been blocked or interrupted. Tucker and Edmondson (2003) describe this as first-order problem solving, i.e., generating short-term remedies that are successful at the time but do nothing to prevent the problem's recurrence. First-order problem solving is encouraged by an organisational emphasis on individual vigilance (encouraging each professional to independently resolve problems as they arise), few slack resources for quality improvement, and the progressive separation of managers from frontline work (Tucker and Edmondson, 2003). In contrast, second-order problem solving involves additional behaviours to address the underlying causes of the problem, and facilitate organisational improvement. This may involve communicating with the department responsible, escalating the concern, sharing ideas for its resolution, and implementing

and verifying process changes. Second-order problem solving behaviours are much rarer.

As a result, 7%-14% of each nursing shift is lost to system failures that could have been addressed (Tucker and Edmondson, 2003; Tucker, 2004; Tucker et al., 2014). To date, there has been little focus on the concerns that trouble frontline staff: local managers and national safety initiatives have turned their attention elsewhere (Tucker et al., 2008; Singer et al., 2009). In short, staff are discouraged from raising important operational concerns, both by counterproductive organisational values and the deliberate prioritisation of other issues.

4.3.3 Collating team experiences, rather than individual reports, may be beneficial

Conceptualising frontline experiences as team knowledge, rather than individuals' insights, might be a more helpful approach. This reframing could guide novel efforts to capture data, with the development of team-wide interventions. Importantly, it might also give those data sufficient authority to motivate change at higher organisational levels, where 'validating' any one individual's perceptions of care is seen as a major challenge (Martin et al., 2015).

Teams gathered together already identify numerous quality and safety issues, in the course of existing ward rounds and interdisciplinary meetings (Andrews et al., 1997; Lamba et al., 2014). Improvement interventions could take advantage of these existing fora. Alternatively, innovations can deliberately engage whole teams in new schemes to identify recurrent problems (Sujan et al., 2011; Lear et al., 2013). These initiatives differ from 'team huddles' that formalise shift-by-shift risk monitoring, which protect the patients at hand, without necessarily recording problems or preventing their recurrence (Brady et al., 2013; Goldenhar et al., 2013).

Observational reports have described how embedded observers can work with interdisciplinary teams to facilitate rapid identification and review of potential adverse events. This has been termed 'prospective clinical surveillance' (Andrews et al., 1997; Forster et al., 2006; Forster et al., 2008; Forster et al., 2011; Wong et al., 2015). Prospective clinical surveillance can be understood as a way to engage frontline staff in formally recording their routine experiences

of care delivery and its consequences, with a structure to support data capture. It may prompt teams to go beyond a simple formal recognition of problems, galvanising them to ameliorate defects in near-real time (Thomas, 2015). However, this strategy has never been evaluated in a controlled trial, nor linked to improved patient outcomes (Forster et al., 2011). The requirement for multiple external observers is also likely to impede wider implementation: a more frugal strategy is required.

Frontline teams may be able to accurately detect and report hazards without an observer. Structured team self-report captures the majority of major errors documented by a trained observer, at least in the operating room (Mason et al., 2013). Without a structure, teams recognise a much smaller proportion of errors, and even those that are recognised are not captured in the medical record or in incident reports.

4.4 Prospective clinical team surveillance: a programme theory

I have described how staff perceptions of care challenges correlate with patient outcomes, yet there are systemic incentives for staff to use only first-order problem solving, and organisations are poorly oriented to frontline clinicians' concerns. Observational reports suggested that a team-wide focus on problem identification and mitigation might change this, and could generate data to be taken up for organisational improvement. An intervention to promote this effective team approach was then developed, with a theoretical grounding.

QI interventions, particularly complex interventions [see 3.5.3], should be informed by theory - in both their development and evaluation. This theory need not be abstract: a programme theory is a deliberately practical, accessible framework, specific to each QI intervention. It describes the components of the intervention, its expected outcomes, and how those outcomes will be assessed, often in a logic model or driver diagram (Davidoff et al., 2015). A driver diagram begins with a clearly defined goal, identifies the high-level factors believed to have a direct impact on it, and subdivides those factors into underpinning goals (i.e., 'drivers') that help define the

actions to be taken (NHS Institute for Innovation and Improvement, 2013). The driver diagram for the novel complex intervention in this study is shown later in the chapter [figure 4.1]. The programme theory should also make clear the rationale and assumptions about the intervention's mechanisms (i.e., its 'theory of change'), in a narrative that links its inputs to outcomes. The full programme theory incorporates the description of the intervention's components, and the narrative about structures, behaviours and processes that are needed to 'achieve the aims and actions of the intervention' (Davidoff et al., 2015). Importantly, the programme theory also begins the process of evaluation, by defining what is to be tested, and suggesting how best that testing might be carried out (Parry et al., 2013). A comprehensive programme theory therefore promotes clarity about the intervention, its mechanisms of action, its key outcomes, and the measurement tools and analytical approaches that will be used to evaluate it (Davidoff et al., 2015).

In this section of the chapter, I outline the programme theory for an interdisciplinary team intervention, navigating first its components and then its proposed mechanisms of action.

4.4.1 Prospective clinical team surveillance: structured interdisciplinary team briefings, facilitated escalation, and feedback

As described above, structured team self-report is feasible in a clinical microsystem, and sufficiently accurate to capture a majority of the problems documented by an external observer (Mason et al., 2013). To foster second-order problem solving and organisational improvement, a system for effective team self-report should:

(i) promote the unassisted identification of safety concerns by frontline staff;

(ii) motivate them to resolve unit-level issues within their control;

(iii) record frontline successes and challenges to build a richer understanding of safety and resilience; and

(iv) engage leadership in attending to frontline concerns.

Such a system might achieve these goals by supplementing autonomous, structured frontline briefings for self-report, with facilitated escalation and feedback of the issues they identified [box 4.1]. I described this system as prospective clinical team surveillance (PCTS), emphasising its reliance on frontline teams - rather than an external observer - for primary data collection. With no requirement for an external observer in each area, improvement resource was assigned instead to a single facilitator who then coordinated escalation and feedback. As the driver diagram shows, improvements in interdisciplinary team care effectiveness would then be brought about by parallel improvements in ward teams' function and incremental support service improvements responding to their concerns [figure 4.1].

Box 4.1 Prospective clinical team surveillance

Prospective clinical team surveillance (PCTS) is the identification of recent clinical and administrative challenges affecting frontline care delivery, using a structured briefing for team self-report.

Issues raised at the team briefing may be amenable to immediate intervention, or may require a more detailed assessment with input from other departments. A facilitator summarises the briefing results for dissemination within the organisation, and coordinates follow-up and feedback.

Each of the three components of the intervention - interdisciplinary briefings, facilitated escalation, and feedback - is discussed in turn.

4.4.1.1 The HEADS-UP briefing

Designed to be used on adult medical wards, the Hospital Event Analysis Describing Significant Unanticipated Problems (HEADS-UP) briefing focused on the clinical and administrative problems most commonly identified in these areas, as described in a large, observational European study (Lubberding et al., 2011) [figure 4.2]. The briefing was developed with frontline clinical staff between October 2013 and November 2013 [see 4.4.4], in which time additional pertinent prompts were identified. These included frequent, or potentially serious, lapses in care that

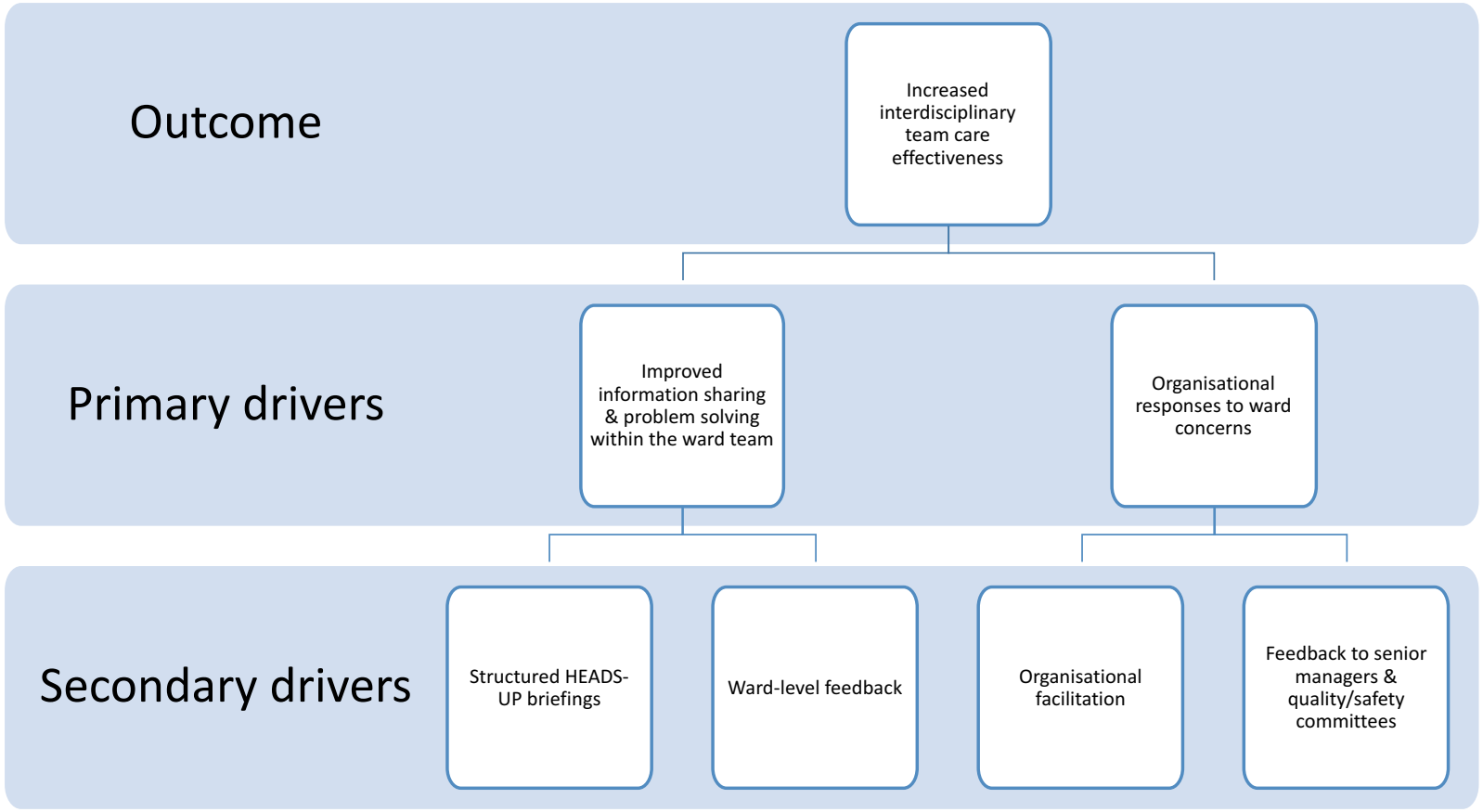


Figure 4.1: Prospective clinical team surveillance driver diagram

Increasingly effective interdisciplinary team care might be achieved through improvements in information sharing and problem solving within the ward team, allied with organisational responses to ward concerns. These will be driven by structured team (HEADS-UP) briefings, with ward-level feedback, organisational feedback and facilitation.

staff had noted on their wards. Briefings could be led by any member of the interdisciplinary team, regardless of seniority or role. Indeed, rotating leadership was encouraged, to embed HEADS-UP in routine practice.

As a result of early feedback from clinicians, a checklist-style pro forma provided the briefing structure, reducing the time taken for its completion. A visual format similar to the World Health Organisation's surgical safety checklist (World Health Organisation, 2014b) was adopted. The earliest versions of the pro forma [figure 4.3] underwent numerous iterations before arriving at the final design [figure 4.4][1]. The briefings ended with prompts for the team to resolve, or escalate, the concerns that they have discussed. These prompts were developed with feedback from both clinical staff and governance teams. Ward teams could adapt their briefings, whilst retaining the same overall format. Similar flexibility has been recommended in other QI work (Bosk et al., 2009; World Health Organisation, 2014b).

4.4.1.2 Facilitated escalation of HEADS-UP-identified concerns

Facilitation advanced the issues raised in the HEADS-UP briefings, helping to translate HEADS-UP-identified concerns into tangible organisational changes. In a recent update to the Promoting Action on Research implementation in Health Services (PARiHS) framework, facilitation was recognised as a key component of successful improvement efforts. It is the construct that 'activates implementation through assessing and responding to characteristics of the innovation and the recipients (both as individuals and in teams), within their contextual setting' (Harvey and Kitson, 2016). Facilitation is not merely an active element of implementation, but is itself a complex intervention, aligning and integrating the innovation, its recipients and its context (Harvey and Kitson, 2016).

Facilitation requires both a role (the facilitator) and a set of strategies and actions (the facilitation process); it is necessarily opportunistic and malleable, taking advantage of existing organisational levers and finding co-enthusiasts. Here, it involved (for example) working with

[1]The final design was produced with Matthew Harrison, Senior Designer at the HELIX Centre, a joint venture between Imperial College London and the Royal College of Art.

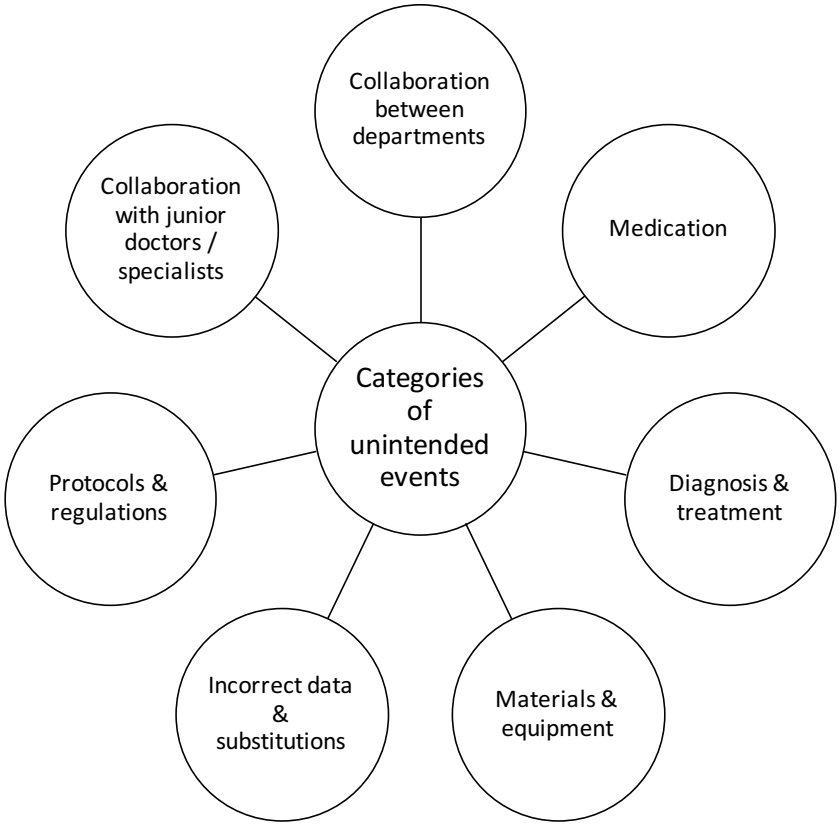
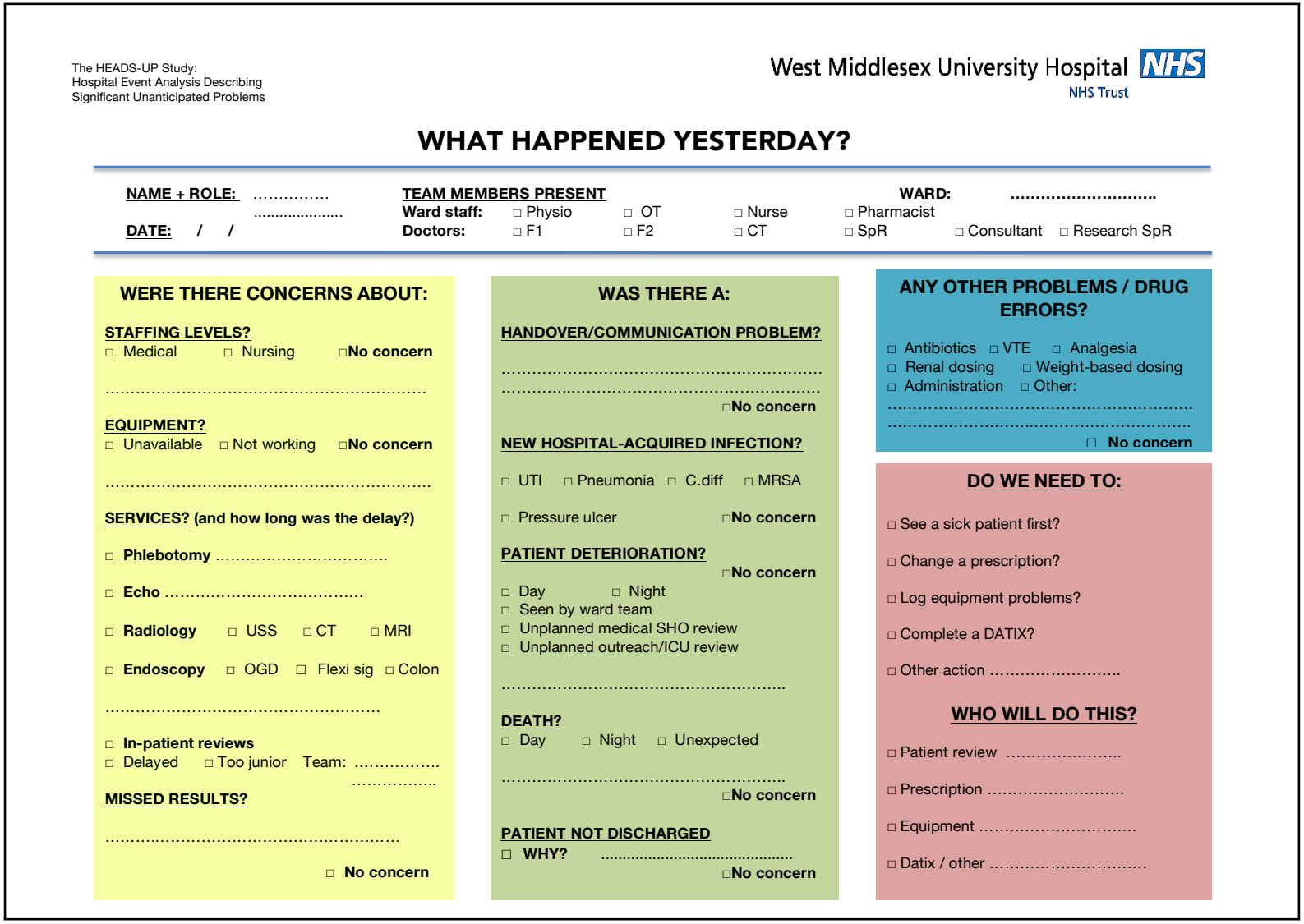


Figure 4.2: Categories of unintended events on medical wards, impacting negatively on patient care (after Lubberding et al., 2011)


The HEADS-UP Study:
Hospital Event Analysis Describing
Significant Unanticipated Problems

West Middlesex University Hospital NHS
NHS Trust

WHAT HAPPENED YESTERDAY?

NAME + ROLE: ……
……

DATE: / /

TEAM MEMBERS PRESENT
Ward staff: □ Physio □ OT □ Nurse □ Pharmacist
Doctors: □ F1 □ F2 □ CT □ SpR □ Consultant □ Research SpR

WARD: ……

WERE THERE CONCERNS ABOUT:

STAFFING LEVELS?
□ Medical □ Nursing □**No concern**

……

EQUIPMENT?
□ Unavailable □ Not working □**No concern**

……

SERVICES? (and how long was the delay?)

□ **Phlebotomy** ……

□ **Echo** ……

□ **Radiology** □ USS □ CT □ MRI

□ **Endoscopy** □ OGD □ Flexi sig □ Colon

……

□ **In-patient reviews**
□ Delayed □ Too junior Team: ……
……

MISSED RESULTS?

……

□ **No concern**

WAS THERE A:

HANDOVER/COMMUNICATION PROBLEM?

……
……
□**No concern**

NEW HOSPITAL-ACQUIRED INFECTION?

□ UTI □ Pneumonia □ C.diff □ MRSA

□ Pressure ulcer □**No concern**

PATIENT DETERIORATION?
□**No concern**
□ Day □ Night
□ Seen by ward team
□ Unplanned medical SHO review
□ Unplanned outreach/ICU review

……

DEATH?
□ Day □ Night □ Unexpected

……
□**No concern**

PATIENT NOT DISCHARGED
□ **WHY?** ……
□**No concern**

ANY OTHER PROBLEMS / DRUG ERRORS?

□ Antibiotics □ VTE □ Analgesia
□ Renal dosing □ Weight-based dosing
□ Administration □ Other:
……
……
□ **No concern**

DO WE NEED TO:

□ See a sick patient first?

□ Change a prescription?

□ Log equipment problems?

□ Complete a DATIX?

□ Other action ……

WHO WILL DO THIS?

□ Patient review ……

□ Prescription ……

□ Equipment ……

□ Datix / other ……

Figure 4.3: Early iteration of HEADS-UP pro forma

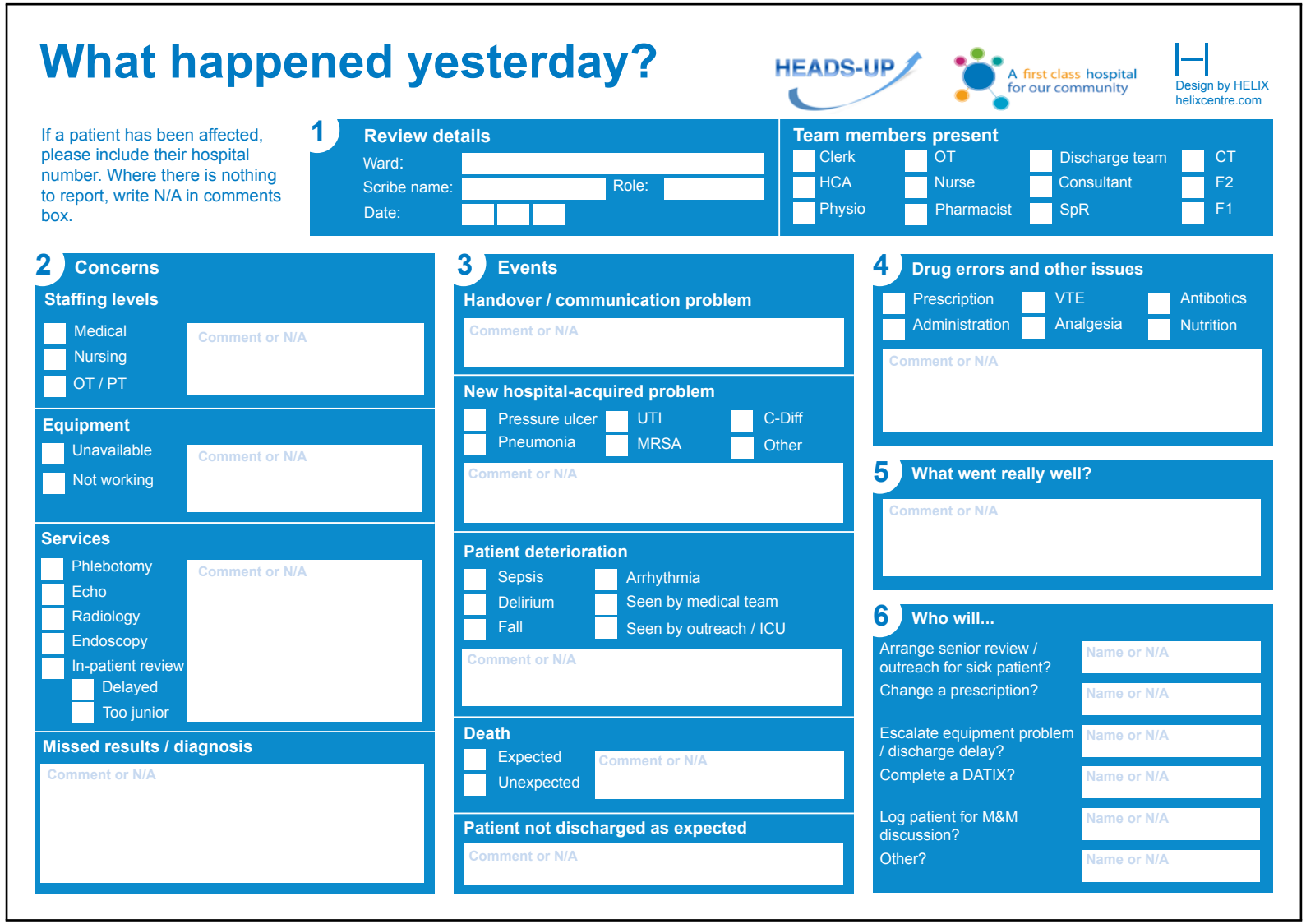


Figure 4.4: Late iteration of HEADS-UP pro forma

frontline teams to identify and document their areas of concern more effectively; championing those concerns in meetings with service leaders and safety committees; and following up on subsequent agreed actions when other priorities threatened resolution. Other studies that have used a facilitator in a similar way have described the facilitator as an *'animateur'*, bringing on board people over whom he has no direct managerial authority (Checkland et al., 2012; Moore and Buchanan, 2013). Detailed advice on operationalising facilitation for future improvement efforts is now available (Kitson and Harvey, 2015).

4.4.1.3 Feedback

Feedback was delivered to participating ward teams, to their managers, and to senior executives. The feedback summarised and disseminated the information collected in the daily briefings, highlighting HEADS-UP performance in each area, common concerns and challenges, potential opportunities for improvement, and recurrent or unresolved problems that might require additional support. HEADS-UP data were provided on request to service leads, to support their business planning. System changes arising from HEADS-UP were publicised, e.g., in existing departmental meetings, and via email and posters. Wherever possible, feedback delivery was timely, focused on solution-finding, signposted to relevant resources, and adopted a non-judgmental approach. These have been cited as characteristics of effective feedback delivery (Reynolds et al., 2016). Where teams provided more detail in their HEADS-UP briefings, feedback could be more specific, with clinical teaching sessions on topics identified as problematic. Although feedback modes (email, posters, clinical presentations, etc.) remained the same, style and content naturally evolved throughout the study period. Facilitation and feedback were targeted to participating clinicians; mid-level and senior hospital leadership; unit- and division-level quality meetings; and safety governance committees. Both facilitation and feedback were provided by an embedded clinical researcher [see 4.4.2].

Cycles of effective briefings, feedback and facilitation were intended to bolster meaningful implementation of HEADS-UP briefings at ward level. In turn, this would progressively

strengthen the programme's impact on incremental ward-level changes, and garner greater organisational support. Although described separately for the purposes of this programme theory, the intervention's components can be understood as interlinked [figure 4.5].

4.4.2 Embedded research model

The intervention was delivered by a clinician with joint academic and clinical commitments at the site contributing the majority of HEADS-UP wards[2]. The academic component to the role made up a 0.75 full time equivalent. A key characteristic of this intervention, therefore, was 'embedded research'. McGinity and Salokangas proposed a definition for this type of endeavour:

> 'Embedded researchers are defined as those who work inside host organisations as members of staff, while also maintaining an affiliation with an academic institution. Their task is seen as collaborating with teams within the organisation to identify, design and conduct research studies and share findings which respond to the needs of the organisation, and accord with the organisation's unique context and culture.' (McGinity and Salokangas, 2014)

Other important features of embedded research include the relationships between staff and the researcher, who is seen as part of the team, and the role of the researcher in contributing to the research capacity of the host organisation (Vindrola-Padros et al., 2016). Characterising the intervention as embedded research is important, because it illuminates various elements of context [see 7.1.2]. First, the appointment of a clinical research fellow represented a significant new investment in QI staffing for the medical department contributing most wards (albeit one partially paid for with clinical shifts). Senior managers were involved in the negotiation to establish the post, and therefore retained an interest in seeing it produce something of value. The research fellow's ongoing clinical work maintained links with the frontline teams imple-

[2] The clinician was a clinical research fellow at medical registrar grade (i.e., a sub-consultant, middle-grade doctor), with specialty experience in gastroenterology and internal medicine. The clinician was in a full-time clinical role at this site for one year prior to the start of the study.

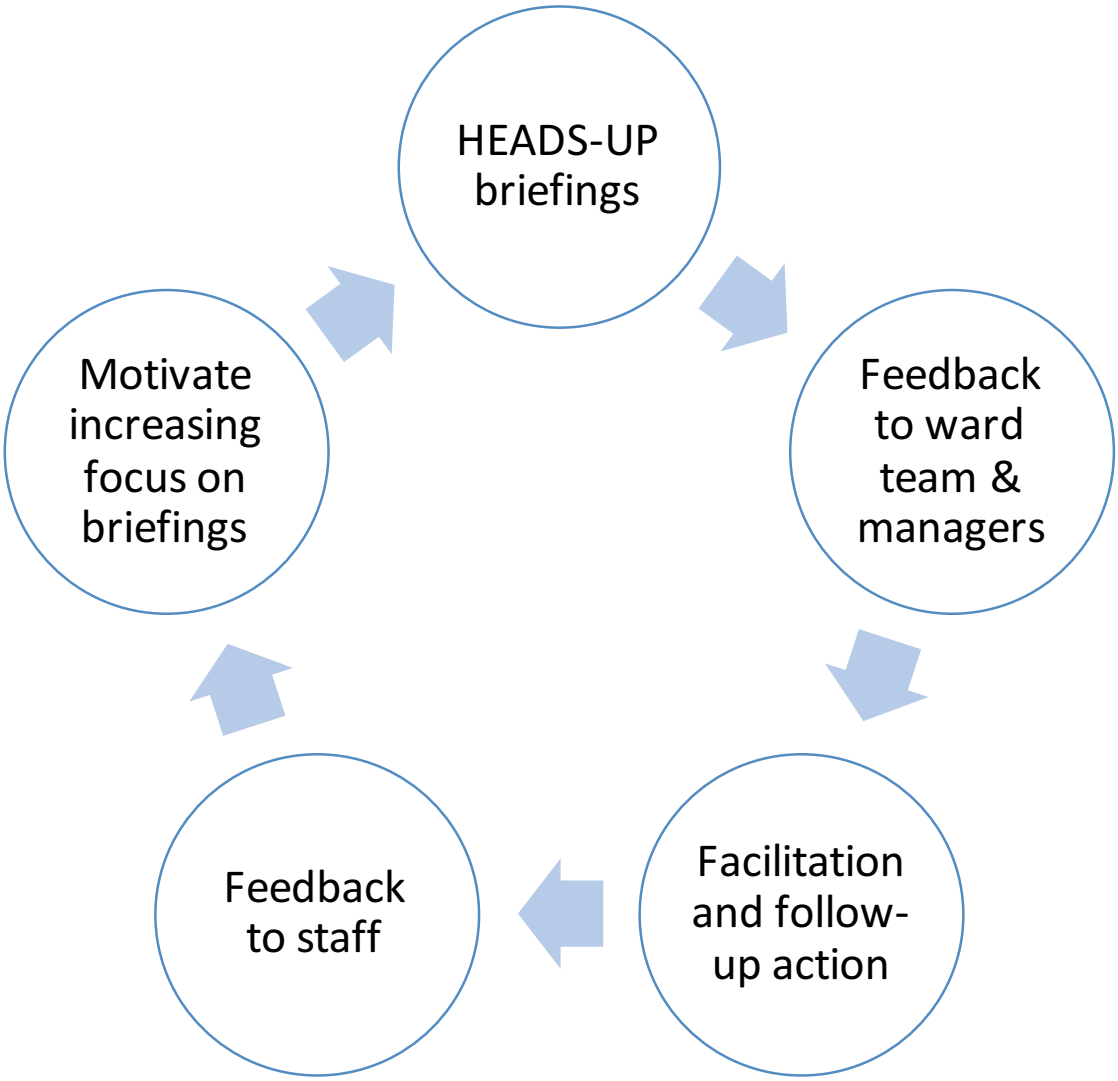


Figure 4.5: Interlinked intervention components

A virtuous cycle links the intervention components described in the programme theory. Changes arising from the programme would motivate increasing engagement with the frontline briefings, in turn increasing their ability to bring about change.

---

menting the intervention. This promoted a sense of local ownership of the intervention, rather than a feeling that it was externally imposed. Second, the post embodied a formal link with the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre. Through this link it enhanced the research capacity of the hospital. The financial investment in, and research engagement with, the programme stimulated local interest in seeing it meaningfully implemented.

4.4.3 Proposed mechanisms of action

I hypothesised that PCTS would improve the sharing and use of data within frontline teams, and between those teams and the wider organisation. Information may not be shared within healthcare teams for a variety of clinical and non-clinical reasons (Kennedy et al., 2009; Blackman et al., 2013). Introduction of PCTS would empower junior clinicians to voice their concerns, improving their teams' situational awareness - itself an important factor in mitigating risks (Brady et al., 2013). This would lead to improvement in units' safety and teamwork climates. With earlier team recognition of deteriorating patients, the process of escalation of care would be eased. Information generated by ward teams would inform their own practice, and prompt downstream service reorganisation. The combination of ward and support service improvements would ultimately improve clinical outcomes, with a dose-response relationship. Lastly, an explicit focus on team-wide recognition of adverse events would improve engagement with existing incident-reporting systems, with an increase in formally reported incidents.

Three specific mechanisms of action were proposed to underpin these anticipated changes: an increase in second-order problem solving, the generation of psychological safety, and increased managerial attention to operational failures.

4.4.3.1 Increased second-order problem solving and psychological safety

In order to increase second-order problem solving by frontline staff, managers should increase their availability during shifts, increasing the likelihood of their being informed of problems

(Tucker and Edmondson, 2003). They should also create a psychologically safe work environment, in which team members can express their concerns and admit their own errors (Tucker and Edmondson, 2003; Tucker, 2007). This may be more important for frontline system improvement than the efforts of highly motivated individuals (Tucker, 2007).

PCTS was intended to make second-order problem solving the default mode for problem resolution, rather than the exception. Although not directly impacting on the time managers are available, it would effectively increase the chance of their being informed of any problems, by providing a structured mechanism for that information transfer to occur. Structured briefings explicitly required that problems should be recognised as common and pervasive, not deemed rare indicators of personal failings. This would establish the psychological safety required for improvement.

4.4.3.2 Increased managerial attention to operational failures

PCTS also reinforced the need for managers to address common operational problems. Ordinarily, frontline workers 'do not control the organisational processes responsible for the majority of failures they encounter, and have a difficult task convincing managers that they warrant resolution' (Tucker, 2004). Systematic collation of these problems would support frontline staff in their requests for managerial help. Successfully resolving operational issues is likely to improve both safety and efficiency, in contrast to the perceived trade-offs between them (Tucker et al., 2008). With local evidence of this, managers would then be more willing to help resolve future issues.

4.4.4 Pilot implementation

HEADS-UP briefings were piloted on one gastroenterology ward at a district general hospital, over a 6 week period (October-November 2013). The briefings were initially led by a clinical research fellow [see 4.4.2], to explain the programme's intentions, demonstrate an 'ideal' briefing, and prompt feedback about how it might be improved. Leadership of the briefings was then

delegated to ward staff, at first with direct observation (with feedback to the briefing leader) and then unsupervised.

When led by a clinician, the HEADS-UP briefing typically took between five and eight minutes to complete. Clinical teams' use of the tool was deliberately unsupervised for the latter half of the pilot period, to gauge whether it was suitably concise and relevant to be used in practice without the researcher being present. The HEADS-UP briefing was completed unsupervised on 80% of working days. Taken together, these data suggested that a formal HEADS-UP trial was feasible.

4.5 The HEADS-UP trial: study design and setting

HEADS-UP (the operationalisation of PCTS, with HEADS-UP briefings, facilitation and feedback) was evaluated in a prospective stepped wedge, cluster controlled trial. The primary aim of the study was to assess the impact of HEADS-UP on clinically relevant patient outcomes. The secondary aim was to explore how changes in patient outcomes, if any, were mediated by changes in workplace climate and ward processes.

Stepped wedge study designs involve the sequential introduction of the intervention to each of the clusters (in this case, wards) over time. Clusters progressively moved from the control group to the intervention group [figure 4.6]. The staged implementation in a stepped wedge design is particularly helpful when simultaneous roll-out of an intervention to all clusters is impractical, e.g., because of logistical constraints, as was the case here. This trial design is typically used to evaluate interventions whose effects are predicted to be more beneficial than harmful (Mdege et al., 2011). Stepped wedge designs are increasingly used to test interventions in acute care (Fuller et al., 2012; Kitson et al., 2013; Poldervaart et al., 2013; Pearse, 2014).

The order in which clusters received the intervention here was guided by logistical restrictions, and to some extent the *de facto* recognition of clinicians enthusiastic to introduce HEADS-UP to their wards. Implementation guidelines for complex interventions, such as sur-

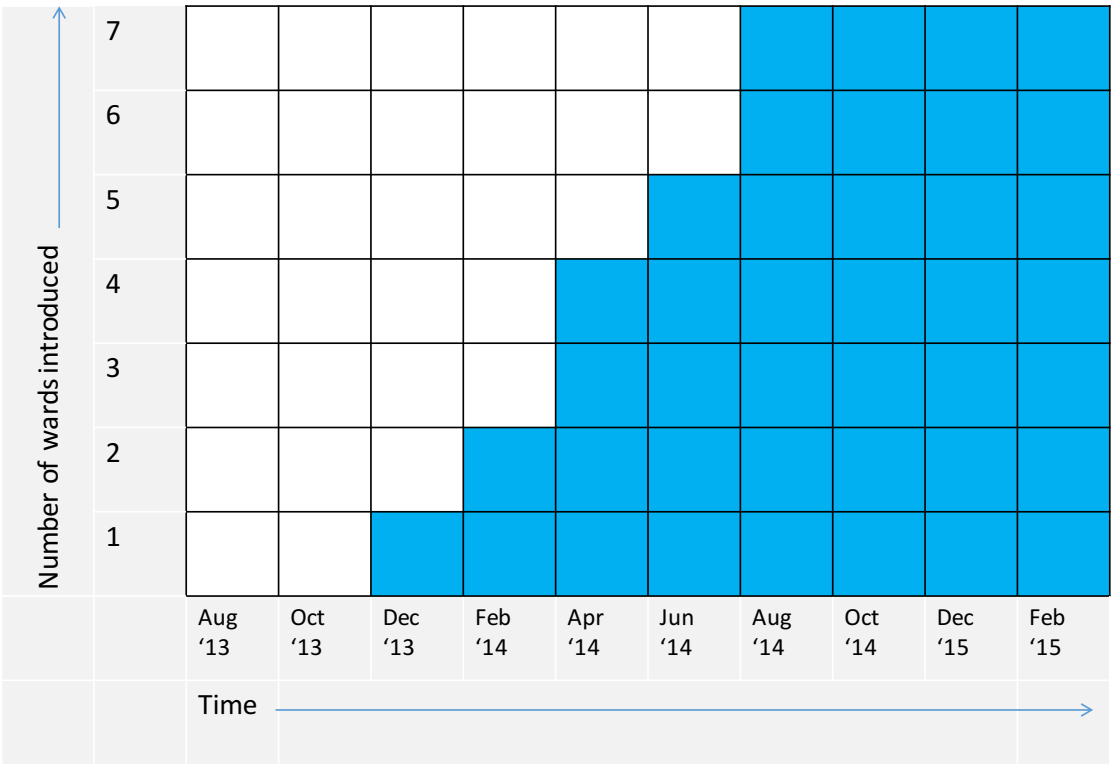


Figure 4.6: Stepped wedge cluster design in the HEADS-UP trial

---

New ward clusters introduced HEADS-UP briefings at two-month intervals. After discussion with senior staff about the optimum timing of implementation in their units, two wards joined the intervention group in April 2014, and another two in August 2014.

---

gical and obstetric safety checklists, advocate this approach (World Health Organisation, 2014b; World Health Organisation, 2016). Although cluster randomisation (randomising the order in which the clusters receive the intervention) might have been preferable, it was deemed important to recognise the impact of willing early adopters, who could lead their colleagues in rigorously implementing the intervention (Ham et al., 2002; Wilson, 2006; Frankel et al., 2008). In the event, unanticipated staffing changes, such as early retirements and maternity leave, limited the participation of these early adopters, and implementation order was determined primarily by logistical restraints. HEADS-UP briefings were introduced to clusters of wards at two-monthly intervals [see figure 4.6].

4.5.1 Setting

The study was conducted at two sites in London, UK. The first was a university-associated district general hospital. Clusters from this site were generated from the acute admissions and downstream medical (gastroenterology, respiratory and geriatric) wards. Each cluster comprised clinical areas that were physically linked, served by the same medical team, or both. This helped to limit contamination between groups. The second site was an academic hospital, where HEADS-UP was implemented on a geriatric ward [see table 5.1].

4.5.2 Implementation

Support for the study was sought from senior clinicians and executives with responsibility for clinical quality, safety and risk management at each site. The intervention was introduced to clinicians and managers with regard to their workflow, opportunities for feedback and learning, and monitoring of implementation fidelity.

4.5.2.1 Introduction to ward teams

Presentations to each clinical team publicised the intervention prior to its introduction. Interdisciplinary ward teams were asked to use HEADS-UP briefings on a daily basis during the

normal working week (Monday to Friday). Staffing and service provision were significantly reduced out of hours, and it was not thought practical to incorporate a HEADS-UP briefing at night or at weekends within those constraints. The initial HEADS-UP briefings in each unit was supervised by the research fellow, but rapidly delegated to the clinical team. The presence of a researcher was documented when appropriate.

4.5.2.2 Workflow change

To maximise clinical engagement with this intervention, teams' existing working patterns were disrupted as little as possible: they used HEADS-UP wherever it fitted most naturally into their existing schedule. This was ordinarily early in the day. No protected time was available for HEADS-UP training, but the ideal format for each briefing was discussed in departmental rounds, team meetings and with participating clinicians.

4.5.2.3 Feedback and learning

The information shared during the daily HEADS-UP briefings prompted a degree of reflective practice and immediate learning. In addition, a regular summary of the HEADS-UP events recorded from their clinical area was given to each team. The format of this feedback again depended on the team's existing schedule; where possible, it was incorporated into existing departmental educational or governance meetings, to place it in the appropriate context and minimise any additional time commitments. This feedback emphasised the ongoing impact of the information gathered during the HEADS-UP briefings, highlighting any subsequent quality improvement work or changes to support services.

HEADS-UP summaries were also shared with the governance committees responsible for the issues raised. Clinicians and managers already accountable for quality and safety in these clinical areas were expected to use the information appropriately to guide resource allocation, and make changes to routine processes or procedures as they saw fit. No specific guidance was issued as to how the HEADS-UP data should be used. However, significant safety concerns and

adverse events raised through HEADS-UP briefings were emphasised to the responsible clinical team or governance body, in order that they could take appropriate action. The need for this proviso was emphasised by peer reviewers of the published study protocol (Pannick et al., 2015).

4.5.2.4 Fidelity assessment

The fidelity of HEADS-UP implementation was evaluated primarily through the completed briefing pro formas, which recorded the team members present, number and type of concerns raised, and the decisions taken as a result. A narrative diary, describing the qualitative impact of HEADS-UP, as well as the downstream QI work and service changes arising from it, complemented the quantitative outcomes. The narrative account also recorded the observation of a number of HEADS-UP briefings, and the extent to which they held true to the perceived ideal in terms of participants, timeliness, focus and intent.

4.5.3 Study population

The study focused on adult medical patients admitted to study wards between 2013 and 2015. To isolate the effect of the intervention, patient-level exclusion criteria included:

- Time spent on the specified ward comprising less than 50% of the total inpatient stay;
- Discharge to a new skilled care facility or other hospital (i.e., not the patient's address at the time of admission. Discharge to a new facility typically incurs substantial delays, outside of the ward team's control.);
- Multiple intra-hospital ward transfers. A single transfer from the initial admissions unit to a downstream medical ward was permitted. One further transfer to an escalation area to facilitate discharge (whereby the patient spends less than 24 hours in the escalation area immediately prior to their discharge home) was also permitted;
- Admission to the high dependency unit or intensive care unit;

- Elective admission or direct admission from another hospital; and
- Surgeon-directed care for more than 24 hours during the inpatient stay.

4.5.4 Study outcomes & hypothesis testing

The primary study outcome was excess length of stay (eLOS): a surplus stay of 24 hours or more, compared to local peer institutions' Healthcare Resource Group (HRG)-predicted length of stay[3]. The local average HRG-predicted length of stay was generated with data from four other nearby hospitals, broadly subject to the same community service restrictions and healthcare economy demands. Secondary outcomes were:

- 30-day readmission;
- a composite of in-hospital death or death/readmission within 30 days of discharge;
- complications of care (hospital-acquired infections and pressure ulcers);
- escalation of care (a composite of emergency calls, referrals to the ICU outreach service, and transfers to the ICU [see 5.3.2]);
- staff engagement with the existing web-based incident reporting system (number, and type, of incidents reported); and
- safety and teamwork climates (using the relevant subsections of the well-validated Safety Attitudes Questionnaire (SAQ) (Sexton et al., 2006)).

The selected measures were similar to those used in other studies evaluating the quality of ward care (Timmermans et al., 2014), and have been assessed in a UK setting where appropriate (Sexton et al., 2006). Table 4.1 lists how the study's outcomes tested each of its hypotheses, and the rationale for their selection. The qualitative impact of the intervention, its implementation,

[3] HRGs are standard groupings of clinically similar patient events, which use common levels of healthcare resource (Health and Social Care Information Centre, 2016a).

and the multilevel barriers and facilitators to its introduction were also assessed through focus groups and auto-ethnography [described in detail in chapters 6 and 7]. Together with the patient- and ward-level outcomes described here, they completed a mixed methods analysis of PCTS.

4.5.4.1 Confounding factors

Staff workload and patient casemix were likely to be the two predominant confounding factors for this non-randomised study. Ward admissions were documented from hospital administration systems, and staff perceptions of workload assessed periodically with the validated NASA-Task Load Index (NASA-TLX) (Hart, 2006). Patient comorbidities were identified from administrative data with an updated version of the Charlson comorbidity index (Charlson et al., 1987; Health and Social Care Information Centre, 2014).

4.5.5 Sample size, power calculation and analysis

Local data from six of the seven anticipated clusters were evaluated. Between 2011 and 2013, a median of 522 patients matched the inclusion/exclusion criteria each month (range 316–722). Mean cluster eLOS rates ranged from 5.6% to 51.5%.

Cluster controlled trials require an estimate of the intraclass correlation coefficient (ICC); this is complex, and observed ICCs rarely match their predicted values. ICC estimation should take into account the results of previous studies, as well as the anticipated numbers of individuals and clusters in the index study. Greater numbers of individuals in a study reduce the width of the ICC confidence interval, mitigating the effect of a relatively small number of clusters (Eldridge, 2011).

In addition, clinical outcome measures tend to have lower ICCs than process measures (Eldridge, 2011). The observed ICC for falls in a multifactorial intervention on elderly care wards was only 0.007 (Cumming et al., 2008). However, ICCs for length of stay and appropriateness of stay in trials of inpatient care pathways were an order of magnitude higher (Kul et al., 2014). For this study, a more conservative ICC estimate of 0.06 was therefore adopted.

Table 4.1: HEADS-UP hypotheses & their corresponding outcomes

Study hypothesis	Outcome component(s)	Rationale for outcome selection
Incremental improvements to ward care & support services would improve clinical outcomes	Excess length of stay	Length of stay reflects efficient resource use, and possibly quality of care (Svendsen et al., 2009) eLOS increases study power compared to length of stay, without requiring an excessive number of wards or data collection period, and may be more likely to reflect interdisciplinary team care [chapter 3]
	Mortality	Correlates with quality of care (Jha et al., 2007) and may relate to performance in non-technical skill domains (Curry et al., 2011)
	Readmission	Need to confirm that any improvements in hospital efficiency do not come at the expense of increased medical readmissions, 37% of which are avoidable (Donze et al., 2013)
Improved situational awareness would mitigate patient risks	Complications of care	These outcomes are appropriate patient safety indicators (Savitz et al., 2005), reliably reported because of stringent mandated targets (NHS England, 2014a; NHS England, 2014d)
Earlier team recognition of the deteriorating patient would facilitate processes underpinning escalation of care	Escalation of care	Multidisciplinary interventions, increasing team situational awareness (Pronovost et al., 2003), may address staff reluctance to appropriately escalate concerns (Kennedy et al., 2009; McNeill and Bryden, 2013)
Team-wide recognition of adverse events would improve engagement with existing incident reporting systems	Number & characteristics of incident reports	More reports overall, with a lower contribution from reports of slips & falls, associated with more positive safety culture and risk management ratings (Hutchinson et al., 2009)
Empowerment of junior clinicians would improve perceptions of safety & teamwork	Safety & teamwork climate	Improved safety climate is associated with organisation-wide reduction in adverse events (Mardon et al., 2010) SAQ scores have substantial validity evidence, with data for inpatient internal medicine settings and in the UK. Higher scores have been associated with improved patient outcomes (Sexton et al., 2006; Havyer et al., 2014)

Abbreviations: eLOS - excess length of stay; SAQ - Safety Attitudes Questionnaire.

In any case, the power under a stepped wedge cluster randomised trial is relatively insensitive to ICC underestimation, compared to a parallel cluster design (Hemming et al., 2015b). Similarly, stepped wedge trial power is relatively insensitive to variations in the coefficient of variation (Hussey and Hughes, 2007). In the best-case prediction, the coefficient of variation approached the level at which adjusting for variable cluster size has a negligible impact on sample size, even in parallel cluster trials (Eldridge et al., 2006). No adjustment for coefficient of variation was therefore required.

Given the variation in baseline outcomes in the local data set, I then estimated the study's power to detect a one standard deviation reduction in eLOS on the wards with the highest and lowest baseline outcome rates. With 7840 patients in the trial (560 patients/month), and two-sided $p<0.05$, the study would achieve 100% power to detect a 14% absolute risk reduction. At worst, in the ward with the lowest baseline outcome rate, it would achieve 75% power to detect a 2.3% absolute risk reduction. The study's power to detect a 2–14% absolute risk reduction therefore lay between 75% and 100%. With this complex trial design, statistical uncertainty in power calculations is not considered unusual. The protocol of a large stepped wedge cluster randomised trial similarly produced a range within which its power might lie (Pearse, 2014). Experts also advocate more widespread recognition of the inherent limitations of statistical power thresholds (Campbell, 2013).

The stepped wedge design compares outcomes in each cluster before and after the introduction of the intervention. Overall differences in outcomes between pre-intervention and post-intervention periods were calculated. Primary analysis was on an intention-to-treat basis, with a separate pre-specified, per-protocol analysis of those units implementing the intervention with high fidelity. Analyses used the patient-level data described above, clustered within 'units', using random effects to model the correlation between individuals within the same cluster. Generalised linear mixed-effects models, and generalised estimating equations, formed the basis of the analysis (Hussey and Hughes, 2007). The choice between the two mathematical approaches is discussed in more detail in the next chapter [see 5.3.3.2]. Underlying temporal trends were

accounted for. No interim analyses were conducted, nor was a formal health economics analysis.

4.5.6 Allocation and blinding

Individual patients were not recruited separately to the study, so there was no anticipation of significant bias due to lack of allocation concealment (Hemming and Taljaard, 2016). Staff could not be blinded to their ward's assignment. Clinical outcome data and escalation of care data were extracted by local administrative staff blinded to the study, as part of their ordinary duties. Hospital peer group data were generated, under existing contracts, by *CHKS* (Alcester, UK) and *Dr Foster* (London, UK), also blinded to intervention group. Implementation data were recorded from the daily proformas completed by ward teams, before the clinical outcomes in each area were known.

4.5.7 Data management

Data were extracted directly from hospital administrative systems, with a monthly assessment by administrative staff to confirm their reliability. Where appropriate, anonymised data were held securely on password-protected hospital intranet systems. Given the nature of the intervention, the time scale of the study, and the extraction of outcomes from existing administrative systems, no data monitoring committee was required. The summarised trial data set was held by the clinical research fellow, and disseminated to local clinicians and managers as required after the study, with no contractual limitations.

4.5.8 Ethical approval

Permission for the study was sought from the relevant Research & Development authority at each participating institution. Both authorities approved the study as a service development initiative not requiring formal ethical evaluation [see appendix B]. The Office for Regulatory Compliance at the Imperial College Academic Health Science Centre initially advised that registration with a clinical trial database would not be required, given the nature of the

intervention. This decision was reviewed again, and the study was ultimately registered – prior to completion of data collection – in the ISRCTN registry, identifier ISRCTN34806867 (http://www.isrctn.com/ISRCTN34806867).

4.6 Discussion

This stepped wedge, cluster controlled trial assessed the impact of an intervention to improve quality and safety on medical wards. The intervention built on existing observational work about team-wide identification and management of failures in care delivery. A programme theory described how the intervention was anticipated to work, making explicit its assumptions, and generating a framework for evaluation.

The design of the HEADS-UP briefings, as well as their proposed implementation, intentionally minimised the disparity between the trial setting and daily clinical practice. This was intended to maximise the study results' applicability to other healthcare settings. However, the relatively prolonged data collection period made the study vulnerable to unanticipated changes in the local healthcare system. There was also a tension between effective programme promotion, essential for adequate staff engagement, and contamination of the control groups. In addition, downstream interventions arising from the programme could have impacted multiple wards, regardless of their participation in HEADS-UP briefings. The narrative record helped to highlight where this may have been the case, and contextualised the impact of the programme.

In summary, PCTS offered a novel, rapid, whole-team analysis of clinical and administrative challenges, including adverse events, at ward level. Facilitated organisational feedback aimed to improve the use of ward-level data in service delivery. This prospective trial aimed to identify whether PCTS was a useful addition to existing safety systems, with broader lessons for the implementation of safety and quality interventions in the complex medical ward environment. In the following chapters, I describe the results of the programme's mixed methods evaluation, beginning with its effects on patient outcomes and processes of care.

Chapter 5

The effect of PCTS on patient outcomes and processes of care

5.1 Introduction

In the last chapter, I proposed a programme theory for PCTS, which was then operationalised and piloted as the HEADS-UP intervention. HEADS-UP consisted of a daily, structured interdisciplinary briefing for medical ward teams, with facilitation and feedback to escalate and address ward concerns at higher organisational levels. The intervention underwent a mixed methods evaluation, including quantitative and qualitative analyses, auto-ethnography and semi-structured focus groups.

The multifaceted evaluation followed the Medical Research Council's guidance for the design and testing of complex interventions - interventions that 'comprise multiple interacting components, with additional dimensions of complexity including the difficulty of their implementation and the number of organisational levels they target' (Craig et al., 2008; Moore et al., 2015). This guidance recognises the importance of implementation factors, and context, in determining the ultimate impact of the intervention. However, it is common practice to first describe the quantitative results of a complex intervention, before a full analysis of the extraneous

factors that influenced them.

I will follow this approach, describing first the effect of HEADS-UP on different quantitative patient outcomes and processes of care, particularly in the subgroup of months where HEADS-UP was implemented with high fidelity [as specified in the study protocol - see 4.5.5]. In chapters 6 and 7, there will then follow a more detailed discussion of the environment into which HEADS-UP was introduced, and how it was implemented.

5.2 Aims

The aims of this chapter are:

(i) To characterise the quantitative use of HEADS-UP during the study period;

(ii) To use different multi-level modelling techniques to analyse the effect of HEADS-UP on patient outcomes and processes of care; and

(iii) To use a quantitative measure of ward-level HEADS-UP implementation in a subgroup analysis.

5.3 Methods

5.3.1 Study design & participating institutions

The intervention, study design, and patient inclusion criteria were described in detail in the last chapter [see 4.5]. Briefly, the HEADS-UP intervention was progressively introduced to seven wards, from two London hospitals. Characteristics of the participating hospitals are listed in table 5.1. The study was designed as a stepped wedge, cluster controlled trial. Interdisciplinary ward teams were assigned to implement the intervention at intervals, such that by the end of the trial all teams had adopted the intervention and had contributed both control- and intervention-group data. The order in which wards adopted the intervention was pragmatically guided by

logistical constraints, as described in previous stepped evaluations (Bion et al., 2013). Baseline data collection began in August 2013, with introduction of the intervention to new wards at two-month intervals from December 2013 [see figure 4.6]. Data collection continued until February 2015.

5.3.2 Study outcomes and data collection

The primary outcome of the HEADS-UP trial was eLOS. This was a binary variable, defined as an admission lasting at least 24 hours longer than the local average for similar patients, as classified by their HRGs (Health and Social Care Information Centre, 2016a) [see 4.5.4]. The study was adequately powered even for a small (2%) absolute risk reduction in eLOS [see 4.5.5]. Changes of this magnitude, which are still meaningful for an organisation, may be expected in interdisciplinary collaborative interventions (Wild et al., 2004; Jenq et al., 2016).

The remaining quantitative outcomes should be regarded as more exploratory. These included two patient-level outcomes: (i) 30-day readmission, and (ii) a composite of in-hospital death or death/readmission within 30 days of discharge. There were also two ward-aggregated outcomes: (iii) escalation events, and (iv) complications of care. Escalation events comprised referrals to the ICU outreach service[1], patient transfers to the ICU, and emergency calls[2]. Complications of care were defined as hospital-attributed pressure ulcers, Clostridium difficile (C.difficile) infections, and selected bacteraemias for which there is clinical surveillance[3].

[1] ICU outreach services are teams of senior practitioners with experience in critical care or acute care, facilitating interventions for deteriorating patients on the general ward. A referral to the ICU outreach service represents a stepwise escalation of patient care (National Outreach Forum, 2012).

[2] Emergency calls are calls to a standardised hospital switchboard number, requesting immediate assistance for a life-threatening emergency. These emergencies include, but are not limited to, cardiac and respiratory arrests. Staff may also use these calls to request assistance in the case of a patient's rapid, severe physiological deterioration.

[3] Meticillin-resistant Staphylococcus aureus (MRSA), meticillin-susceptible Staphylococcus aureus (MSSA) and Escherichia coli (E.coli).

Table 5.1: Characteristics of the participating institutions

	Hospital 1	Hospital 2
Institution type (inpatient admissions/ year)	Community hospital (52,000)	Academic teaching hospital (186,000)
Participating wards (n)	Acute medical unit (2) Gastroenterology / internal medicine (1) Heart failure / internal medicine (1) Geriatrics (1) Respiratory / internal medicine (1)	Geriatrics (1)
Existing interdisciplinary practice & approach to adverse event detection	Daily interdisciplinary 'board round', typically focused on patients' discharge requirements Online incident reporting system Unstructured mortality casenote reviews	Daily interdisciplinary 'board round', typically focused on patients' discharge requirements Online incident reporting system
Major organisational changes during the study period	Incipient institutional merger	Deployment of a new electronic health record

5.3.3 Statistical analysis

5.3.3.1 Adjustment for clustering

The analysis of stepped wedge trials is a developing area, with little consensus on the best approach (Eldridge and Kerry, 2012). As with any trial, covariates are incorporated into a regression model appropriate for the dataset (e.g., logistic regression for binary outcomes). Here, an additional adjustment needs to be made to account for the effect of clustering. Data points are correlated, not independent: the outcomes of patients treated on the same ward are more likely to be similar than those of patients on different wards.

Different techniques are available to account for clustering. Two conceptually different approaches will be discussed: mixed-effects models and generalised estimating equations (GEEs). Mixed-effects models use cluster-specific models, summarising the response of an individual over time, i.e., the effect of that patient moving from control arm to intervention arm, whilst remaining in the same cluster. GEEs use population-averaged models, calculating average changes in response over time (Eldridge and Kerry, 2012; Ma et al., 2012). Both techniques have advantages over repeated-measures ANOVA, which makes statistical assumptions that are unrealistic for longitudinal datasets (Ma et al., 2012).

The choice between mixed-effects models and GEEs depends on numerous competing trial features, which affect the robustness of their results. Their relative advantages and disadvantages will be outlined briefly, as it is prudent to use more than one analytic technique to confirm stepped wedge trial findings (Eldridge and Kerry, 2012).

5.3.3.2 The relative merits of mixed-effects models *vs* generalised estimating equations

Mixed-effects models are so called because they generally contain covariates with fixed effects (those that have the same effect across many patients) and random effects (those that vary substantially from one patient to the next) (Detry and Ma, 2016). Ward-level variance in patient

outcomes is represented as a random effect. As with GEEs, mixed-effects models require sufficient clusters to produce reliable estimates of treatment effects. However, when the number of clusters is small, as it is here, mixed-effects models are less biased than GEEs (Eldridge and Kerry, 2012). Primarily for that reason, mixed-effects models will be used for the main analysis, and GEEs for a secondary analysis.

Mixed-effects models may also be more efficient, in that they achieve higher study power than GEEs, albeit at the expense of higher type I error rates (Ma et al., 2012). Increasing sample size reduces the relative advantage of GEEs above mixed-effects models in terms of type I error rates (Ma et al., 2012). An important advantage of GEEs is their tolerance of misspecified correlation structures between data points (Overall and Tonidandel, 2004). However, this may not substantively affect results (Gardiner et al., 2009).

5.3.3.3 Definition of HEADS-UP implementation fidelity

For the per protocol analysis, a quantitative definition of implementation fidelity was required. Based on direct observations of HEADS-UP use during the pilot period, the implementation fidelity for a given month was defined as a combination of briefing frequency and team engagement. A categorical interaction term combining 'frequency*engagement' was incorporated into each statistical model, as well as a dummy code representing the HEADS-UP intervention itself. Frequency had three categories: high (≥75% working days that month), medium (50-75%), and low (<50%). Monthly engagement was defined as high (documentation of more than the median number of monthly issues) or low (below the median).

The interaction term therefore combined briefing frequency and engagement into a single representative coding of implementation fidelity. Four categories of implementation fidelity (high, moderate, low and poor) could be represented in the data [see table 5.2].

Table 5.2: Coding of monthly HEADS-UP implementation fidelity by engagement and briefing frequency

Engagement	Briefing frequency (% working days)	Implementation code	Implementation fidelity
Low	Low (<50%)	0	Poor
	Medium (50-75%)	0	
	High (≥75%)	0	
High	Low (<50%)	1	Low
	Medium (50-75%)	2	Moderate
	High (≥75%)	3	High

High engagement reflected documentation of more than the median number of monthly issues; low engagement, correspondingly, reflected documentation below the median.

5.3.3.4 Model specification

5.3.3.4.1 Mixed-effects models Analyses were conducted in Stata/SE v14.1. Mixed-effects models were used to evaluate the effect of HEADS-UP on each patient- and ward-level outcome. For the binary patient-level outcomes, binary logistic models were used ('melogit' command). Fixed effects were included for the HEADS-UP intervention; implementation fidelity; study month; the number of ward admissions that month; the patient's Charlson comorbidity score; age; and palliative care code. A random effect was included for ward. Analyses solely involving eLOS or readmission were restricted to those patients who survived to discharge; no patients were excluded on the basis of 'outlier' length of stay. Other analyses included all patients. The postestimation 'estat icc' command was used to estimate the intraclass correlation for eLOS.

For ward-aggregated counts data (complications of care and processes of care), Poisson loglinear models were used ('mepoisson' command). Fixed effects were included for season, time, the median Charlson score of patients on the ward that month, and the time accrued in that department by the team's most junior doctors (based on 4-monthly rotations between departments). A random effect was included for ward.

5.3.3.4.2 Generalised estimating equations For the secondary analysis of binary patient outcomes, the 'xtgee' command was used, with a binomial distribution and logit link. The same covariates were entered into the model, specifying them as continuous or categorical as appropriate, setting 'ward' as the panel variable. The analysis was run with an exchangeable correlation structure. A similar process was followed for counts data, using the 'xtpoisson' command to specify a population-averaged Poisson model with log link.

5.4 Results

5.4.1 Intervention and control group characteristics

12,077 eligible admissions took place during the study period, of which 11,445 could be included in the primary analysis. 4,927 admissions were in the control group and 6,518 admissions were in the intervention group. Monthly ward admissions were higher in the intervention months (77 admissions/ward/month *vs* 64 admissions/ward/month), as was unadjusted length of stay (2.2 days *vs* 2.0 days) [table 5.3]. eLOS could be assessed in 4927/4929 (99.96%) control admissions and 6543/6546 (99.95%) intervention admissions where the patient survived to discharge [see appendix C].

5.4.2 HEADS-UP use during the study period

HEADS-UP implementation data were available for 71/73 (97.2%) ward months. On average, wards conducted HEADS-UP briefings on 74.2% working days/month (range 15 – 100%), documenting 80.7 issues/ward/month (range 6 – 208). In the intervention group, 3607/6518 (55.3%) admissions were in high fidelity months (fidelity code = 3), 990/6518 (15.2%) admissions were in moderate fidelity months (fidelity code = 2) and 1921/6518 (29.5%) admissions were in poor fidelity months (fidelity code = 0)[4].

The following sections will describe the intention-to-treat and per protocol impact of HEADS-UP on the study's primary and secondary outcomes. In each section, the main analysis (using mixed-effects models) is reported before the secondary analysis (using GEEs).

[4]No months were coded as low fidelity (fidelity code = 1).

Table 5.3: Ward & admission characteristics in the control and intervention groups

	Control group	Intervention group
Ward characteristics		
Number of patient admissions	4,927	6,518
Ward admissions per month	64 (55-153)	77 (64-133)
Length of stay for study patients, days	2.0 (0.8 - 7.6)	2.2 (0.9 - 7.3)
Admission characteristics		
Age, years	67 (49 - 80)	65 (47 - 78)
Age ≥65 years [n (%)]	2707 (54.9)	3275 (50.2)
Female [n (%)]	2537 (51.5)	3391 (52.0)
Charlson comorbidity index	3 (0 - 10)	3 (0 - 8)
Palliative coding [n (%)]	50 (1.0)	94 (1.4)

Values are median (interquartile range) unless otherwise stated.

5.4.3 Primary outcome: excess length of stay (eLOS)

5.4.3.1 Intention-to-treat and per protocol analyses produced conflicting results

1279/4927 (26.0%) control admissions had eLOS *vs* 1714/6518 (26.3%) intervention admissions. In the unadjusted mixed-effects model (intention-to-treat), HEADS-UP had no impact on eLOS (odds ratio (OR) = 1.07, 95% CI 0.96 - 1.19, p = 0.219). In the adjusted intention-to-treat model, the intervention was associated with increasing eLOS (OR = 1.34, 95% CI 1.10 - 1.64, p = 0.004). However, this did not reflect whether the intervention had actually been used, and implementation fidelity significantly affected the result (χ^2 = 7.62, p = 0.0221).

5.4.3.2 High fidelity HEADS-UP implementation reduced excess length of stay

The percentage of admissions with eLOS varied with implementation fidelity: 904/1921 (47.1%), 196/990 (19.8%) and 614/3607 (17.0%) for poor, moderate and high fidelity respectively. An exploratory GEEs plot of the relationship between engagement, briefing frequency, and the probability of eLOS is shown in figure 5.1, using engagement as a continuous variable (the number of issues reported each month). This showed a progressive reduction in eLOS as fidelity improved, i.e., with increasing engagement at high frequency briefings. The relationship was not apparent for increasing engagement at low frequency briefings.

Point estimates at moderate and high fidelity favoured the intervention, with significant improvement, i.e., reduced eLOS, at high fidelity (OR = 0.78, 95% CI 0.66 - 0.94, p = 0.007) [table 5.4]. Of note, the ICC for eLOS in the mixed-effects model was 0.07 (95% CI 0.018 - 0.250).

These findings were replicated in the GEEs model. HEADS-UP implementation fidelity again significantly affected eLOS (χ^2 = 17.91, p < 0.001). The magnitude of the effect size for high fidelity implementation was higher using GEEs than the mixed-effects model, with an odds ratio of 0.72 (95% CI 0.62 - 0.84, p < 0.001). The GEEs models are reported in full in appendix C.

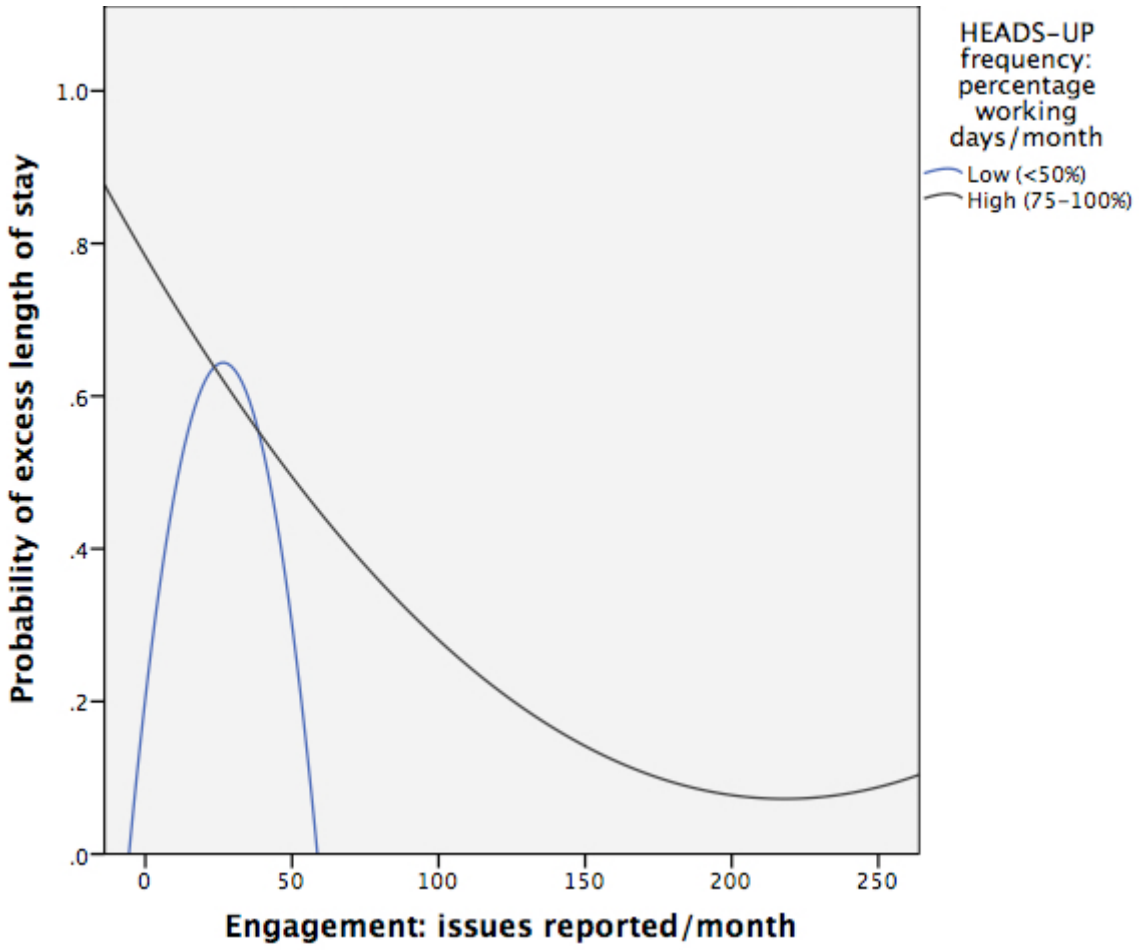


Figure 5.1: The relationship between engagement, briefing frequency, and the probability of excess length of stay

Table 5.4: Intention-to-treat and per protocol analyses for patient outcomes and processes of care (mixed-effects models)

Outcome	HEADS-UP	Control	Adjusted outcome ratio*					
			Intention-to-treat		Moderate fidelity**		High fidelity***	
			Outcome ratio (95% CI)	P value	Outcome ratio (95% CI)	P value	Outcome ratio (95% CI)	P value
Primary outcome								
eLOS [n (%)]	1714 (26.3)	1279 (26.0)	**1.34 (1.10-1.64)**	**0.004**	0.96 (0.77-1.19)	0.700	**0.78 (0.66-0.94)**	**0.007**
Secondary outcomes								
Readmission within 30 days [n (%)]	950 (14.5)	565 (11.5)	1.11 (0.87-1.41)	0.391	1.00 (0.78-1.29)	0.975	1.09 (0.90-1.33)	0.369
In-hospital death or death/readmission within 30 days [n (%)]	1662 (24.1)	1190 (22.9)	1.00 (0.82-1.22)	0.979	1.07 (0.86-1.31)	0.555	1.09 (0.93-1.29)	0.285
Escalation events/month**** [median (IQR)]	8 (3.5-16)	3 (1-6)	0.83 (0.62-1.10)	0.199	**1.35 (1.04-1.75)**	**0.024**	1.09 (0.88-1.35)	0.440
Complications of care/month [median (IQR)]	1 (0-2)	1 (0-3)	0.87 (0.51-1.48)	0.601	1.20 (0.62-2.31)	0.588	1.33 (0.78-2.26)	0.299

*Odds ratios for binary outcomes, and incidence rate ratios for continuous outcomes. **Medium frequency briefings with high engagement. ***High frequency briefings with high engagement. ****Emergency calls, referrals to the ICU outreach service, and ICU transfers.

5.4.4 Secondary outcomes

5.4.4.1 30-day readmission

In the mixed-effects model, HEADS-UP had no detectable effect on 30-day readmissions [table 5.4]. This remained the case regardless of changes in implementation fidelity. Results were similar with GEEs, HEADS-UP having no effect on the odds of readmission, either in the intention-to-treat analysis (OR = 1.10, 95% CI 0.87 - 1.38, p = 0.433) or in the per protocol analysis (χ^2 for implementation fidelity = 1.42, p = 0.492).

5.4.4.2 In-hospital death/death or readmission within 30-days of discharge

HEADS-UP had no detectable effect on the composite of in-hospital death/death or readmission within 30 days of discharge [table 5.4]. This remained the case regardless of changes in implementation fidelity. Similarly, there was no detectable HEADS-UP effect on this outcome in the GEEs model (OR = 0.99, 95% CI 0.82 - 1.19, p = 0.921; χ^2 for implementation fidelity = 2.46, p = 0.292).

5.4.4.3 Processes of care

Escalation events were higher in the intervention group than the control group (median of 8 *vs* 3 escalation events/month). There was some evidence that this may have been due to the intervention. In the mixed-effects model, there was no detectable intention-to-treat effect (OR = 0.83, 95% CI 0.62 - 1.10, p = 0.20). However, moderate fidelity HEADS-UP implementation was associated with significantly more escalation events per month (incidence rate ratio (IRR) 1.35, 95% CI 1.04 - 1.75, p = 0.024). The IRR for escalation events in the high fidelity HEADS-UP implementation group was also greater than 1, although in this group the increase was not statistically significant (IRR 1.09, 95% CI 0.88 - 1.35, p = 0.440).

In the GEEs model, the intention-to-treat analysis suggested a reduction in escalation events (IRR = 0.78, 95% CI 0.69 - 0.89, p < 0.001), with implementation fidelity having a significant impact

(χ^2 = 23.77, p < 0.001). Increases in escalation events were seen with both moderate and high HEADS-UP fidelity (IRR = 1.38, 95% CI 1.21 - 1.57, p < 0.001; and IRR = 1.10, 95% CI 0.997 - 1.20, p = 0.059, respectively).

5.4.4.4 Complications of care

Complications of care were similar in the intervention and control groups (median of 1 complication/month in each group). In the mixed-effects model, there was no evidence that HEADS-UP implementation affected complication rate, regardless of fidelity (IRR 0.87 for intention-to-treat; 1.20 for moderate fidelity group; 1.33 for high fidelity group; all 95% CIs crossed 1; all p values > 0.29; χ^2 for fidelity = 1.15, p = 0.56). These findings were replicated in the GEEs model [see appendix C].

5.5 Discussion

This study used one of the preferred QI methodologies - a stepped wedge, cluster controlled study design - to evaluate a novel interdisciplinary team care intervention. The HEADS-UP trial was adequately powered to detect appropriately small changes in its primary outcome. The calculated ICC was similar to the value predicted in the study protocol [see 4.5.5], and two complementary analytical techniques were used to evaluate the findings. However, the study had limitations, which will be discussed after a brief summary of the principal findings.

5.5.1 Summary of findings

5.5.1.1 Patient outcomes

High fidelity implementation of the HEADS-UP intervention reduced eLOS. This significant improvement was identified in per protocol, but not intention-to-treat, analyses. HEADS-UP did not improve - or worsen - 30-day readmissions, nor the composite of in-hospital death/death or readmission within 30 days of discharge.

5.5.1.2 Processes of care

There was some evidence that HEADS-UP increased the rate of escalation events. However, in the main analysis, statistically significant increases in escalation event rate were only associated with moderate (and not high) fidelity implementation.

5.5.1.3 Complications of care

HEADS-UP did not appear to affect the rate of complications of care, regardless of implementation fidelity.

5.5.2 Limitations

There are two main limitations in this study. First, opposing results from the intention-to-treat and per protocol analyses suggest there may have been important unmeasured confounding factors. Though some quality improvement efforts may indeed worsen outcomes, perhaps by distracting attention from existing good practice (Anthony et al., 2011; Nicholas and Dimick, 2011), this did not appear to be the case here. Low fidelity implementation may have been confounded by worsening health economy pressures and an increasingly high threshold for hospital admission later in the study period, when the intervention group (by design) accrued more data. Ward admissions are an imperfect proxy for this increasing workload: demands on staff increase as the numbers of admissions, transfers and discharges all increase (Needleman et al., 2011). However, ward admissions were the only workload measure consistently available from both participating institutions.

Second, under-use of HEADS-UP may have been a marker of wider team dysfunction and poor performance, rather than itself directly leading to poorer outcomes. Teams that participate more wholeheartedly in trials may be more innovative, with strong leadership, a readiness for change, and better managerial relations (McMullen et al., 2015). It is therefore difficult to entirely separate the intervention's effect from the characteristics of the teams that implemented

it best. Nonetheless, the statistical models incorporated baseline data from each area, and direct observation showed multiple teams using HEADS-UP meaningfully, even during periods of great pressure [see chapter 7]. It is unlikely, therefore, that the results are a simple reflection of teams' pre-existing practice or fluctuations in their workload.

5.5.3 Implications for further work

The adoption of HEADS-UP here compared favourably to that of other team initiatives (O'Leary et al., 2015b; Mayer et al., 2016), suggesting it merits further development. As with other team-based initiatives, the change in the primary outcome was seen only with high fidelity use of HEADS-UP (Mayer et al., 2016). Future work should therefore elicit strategies to improve the implementation of PCTS, as operationalised here with the HEADS-UP intervention. The calibre of implementation varied widely even between the wards of a single hospital; tailored support to help teams engage with this type of QI initiative may be required (Hamilton et al., 2014). Chapter 7 describes the implementation outcomes for HEADS-UP during the study period, which will inform future dissemination efforts. The integration of PCTS alongside other safety strategies also merits further investigation, as do the drivers for resolution of PCTS-raised concerns. Chapter 8 explores why only a subset of clinical concerns may have been formally resolved.

5.5.4 Conclusions

PCTS is a novel strategy to harness frontline staff experience in the pursuit of service improvement. This study provides empirical evidence that high fidelity PCTS implementation reduces eLOS, supporting the contention that healthcare safety and efficiency need not be opposing values (Tucker et al., 2008). The effects of this complex intervention cannot be fully encapsulated by changes in patient outcomes and processes of care. In chapter 6, I will now explore the mechanisms by which PCTS acted, its qualitative impact, and its effect on staff safety behaviours (incident reporting).

Chapter 6

PCTS mechanisms of action, qualitative impact, and effects on incident reporting

6.1 Introduction

The quantitative analysis in chapter 5 identified an improvement in eLOS, corresponding to high fidelity implementation of the HEADS-UP intervention. That analysis used numerical definitions of implementation fidelity, based on frontline briefing frequency and engagement. As a complex intervention [see 5.1], however, the use and impact of HEADS-UP were contingent on meaningful adoption by clinical and managerial staff. This cannot be fully encapsulated in a numerical summary. PCTS - which HEADS-UP operationalised - was also hypothesised to improve frontline teams' safety and teamwork climates, and to provoke service developments in response to ward teams' concerns [see figure 4.1]. These hypotheses too require additional investigation.

In this chapter, I describe more qualitative components of the mixed method evaluation of the HEADS-UP trial. Semi-structured focus groups, auto-ethnography, validated survey tools

and hospital databases are used to describe the effect of HEADS-UP on staff and the broader organisation, and its interaction with existing safety structures.

6.2 Aims

The aims of this chapter are:

(i) To explore staff perceptions of the effects of HEADS-UP, and describe the organisational changes it prompted;

(ii) To analyse the effect of HEADS-UP on staff safety and teamwork attitudes, using a validated survey tool; and

(iii) To relate HEADS-UP implementation to changes in staff safety behaviours.

6.3 Methods

The intervention and study design were described in detail in chapter 4 [see also the summary in 5.3.1]. Here, I will report how data were collected and analysed for the qualitative reports of HEADS-UP's usage and effectiveness.

6.3.1 Data sources and analysis

6.3.1.1 Auto-ethnography

Fieldnotes over a 20-month period recorded exchanges and emails with staff, personal observations, reflections on the broader context, wider governance proceedings, and other implementation challenges. These 'auto-ethnographic' insights have been widely used in organisational case study research, providing broad accounts of culture and practice (Watson, 2011; Shaw et al., 2014; McMullen et al., 2015).

As an embedded researcher [see 4.4.2], there were ample opportunities to create a rich dataset. I observed at close quarters the intervention, the context into which it was launched, and the competing demands placed on those involved. At hospital 1, the institution contributing the majority (6/7 [85.7%]) of the study wards, I became part of the small hospital quality improvement team, in addition to retaining clinical relationships within the department of medicine. During the early study period, links were established with senior hospital leaders, including the Chief Executive; the Director of Workforce & Development; the Medical Director; the General Manager for Medicine (later the Director of Operations); and the Director of Quality Improvement. In addition, the post required close collaboration with numerous 'hybrid' managers (e.g., senior matrons and clinical directors), who had ongoing clinical responsibilities alongside their managerial workload (Birken et al., 2012; Buchanan, 2013). Middle managers - reporting to the executive team - were also closely involved.

One facet of the role of embedded researcher is to expand the host organisation's knowledge capacity (Vindrola-Padros et al., 2016). The nature of the HEADS-UP intervention - as well as its link to the NIHR Imperial Patient Safety Translational Research Centre (PSTRC) - provided novel insights into broader issues of patient safety and quality at the hospital. I was therefore invited to attend regular meetings of a hospital-wide mortality review group (chaired by the Medical Director and attended by the Chief Executive), the hospital-wide patient safety forum, and a divisional quality and risk meeting. These experiences, along with email exchanges and personal interactions, were typed up formally as fieldnotes as soon as was practicable.

At the second participating institution - a far larger organisation - there was much less access to equivalently senior leadership. Nonetheless, appropriate links were made with departmental clinicians and managers. Those interactions were documented in a similar way, alongside personal reflections on the departmental safety and quality meetings that were held intermittently.

6.3.1.2 Semi-structured focus groups

Semi-structured focus groups were undertaken at hospital 1 in July 2015. The topic guide explored PCTS as a tool, as well as the implementation process that introduced it: implementation is discussed in detail in chapter 7. Participants were asked to discuss existing systems for identifying team concerns and improving patient care, to orientate themselves to the topic. They then reflected on their experiences of using HEADS-UP. Questions and follow-up probes scrutinised team-wide involvement in HEADS-UP; whether HEADS-UP affected perceptions of the ward's safety and quality; how it was used in staff interactions; and whether teams made changes based on their HEADS-UP briefings and feedback. The full topic guide is provided in appendix D.

Topic questions were reviewed and piloted within the research team involved in the study, and with researchers who had not been involved with HEADS-UP. The focus groups were advertised in clinical areas, by email, and in person. All ward staff were invited to attend, but certain staff groups were purposively targeted to ensure that key stakeholders (junior doctors, senior nurses, and service managers) were represented. Focus groups were led by experienced qualitative researchers[1], using the topic guide flexibly in view of the different roles and experiences within the two groups. The focus groups' discussions were recorded, and the audio recordings then professionally transcribed (PageSix Transcription Services, London, UK).

6.3.1.3 Analytical approach

Focus group transcripts and fieldnotes were managed with NVivo software (QSR International, Doncaster, Australia). Two researchers read and re-read the transcripts[2]. An inductive (theory-generating) approach was used, as recommended by Braun and Clarke (2006). Each researcher coded the transcripts individually, generating an initial coding frame, which was then discussed and refined. The transcripts were coded again, before a group of higher order themes was agreed. Field notes were searched for evidence to support or contradict the focus groups, and for other

[1] Focus groups were led by Professor Nick Sevdalis, Dr Max Johnston, Dr Louise Hull, Dr Stephanie Archer, and Ms Tayana Soukup.

[2] I conducted the analysis with Dr Stephanie Archer, Research Fellow.

themes not covered independently by focus group discussions.

6.3.1.4 Safety and teamwork climate surveys

Physicians and senior nurses were surveyed at baseline, and again six months after HEADS-UP introduction, using the safety and teamwork subsections of the Safety Attitudes Questionnaire (Sexton et al., 2006). Responses were recorded on a five point Likert scale, with reverse scoring for negatively worded items, and transformed to generate a score from 0-100 (Center for Healthcare Quality and Safety, 2015). Perceived workload was assessed with the NASA Task Load Index, on a visual analogue scale (Hart, 2006). Other survey items identified HEADS-UP participation, whether respondents had reported concerns differently as a result, and demographic data. The full survey instrument, and appropriate permissions, are provided in appendix D.

The effect of HEADS-UP was assessed in general linear models, incorporating mean teamwork or mean safety score as the dependent variable. HEADS-UP participation was defined as self-reported engagement in five or more briefings; workload, hospital and survey period were the other independent variables. Survey analysis was conducted in SPSS v22.

6.3.1.5 Incident reports

All incident reports pertaining to the study wards, submitted through online reporting systems during the study period, were analysed. The impact of HEADS-UP on the number of incident reports for each study ward, the number relating to non-falls incidents, and the number submitted by staff other than nurses were assessed. The reporter's initial assessment of incident severity, which is usually inaccurate, was not used (Macrae, 2016).

Numerical estimates of the HEADS-UP effect on incident reporting were generated with mixed-effects models and GEEs models, as described in chapter 5. Poisson loglinear models accounted for clustering of results by ward, and included covariates for season, time, median monthly Charlson score, ward admissions, and junior doctors' time in post [see 5.3.3.4]. Incident reporting analysis was conducted in Stata/SE v14.1.

6.4 Ethics, consent and permissions

Research and development authorities at both sites exempted the study from formal ethics review. Participants signed their consent to the recording and research use of the focus groups. The primary study was registered with the ISRCTN registry [see 4.5.8].

6.5 Results

The final dataset for this evaluation comprised 930 HEADS-UP briefings; 115 staff surveys; 2,175 incident reports; 2 semi-structured focus groups with 15 staff (junior doctors [n=3], nurses [n=8] and managers [n=4]); and 44 pages of ethnographic field notes.

The qualitative impact of HEADS-UP will be described first, followed by the results of the staff attitudes surveys, and then the evaluation of changes in incident reporting.

6.5.1 Psychological safety, with a hard edge of accountability

6.5.1.1 Equal participation from different staff groups

HEADS-UP was the first tool staff had used to provide a formal communication structure for routine work. Still, over the course of the study all staff groups had participated in - and indeed led - the briefings:

> 'The physio[therapist]s or the OTs [occupational therapists] are also flagging things up; that's how we found it useful... The discharges that did not happen, anyone [who] deteriorated, they flag it up and then we all learn from what went wrong.' *(Senior nurse)*

> 'They [consultants] contribute as much as anyone else... so their voice is being heard, because they suffer from the same frustrations as all of us.' *(Service manager)*

> 'It also highlighted issues to them [consultants], as well... Why that patient hadn't had a scan - so they could follow things through.' *(Senior nurse)*

6.5.1.2 A non-judgemental forum to raise concerns and learn as a team

Perhaps because the different staff groups were perceived as equal partners in the HEADS-UP briefings, the briefings formed a psychologically safe environment in which problems could be discussed openly, without fear of retribution. The style and timeliness of HEADS-UP feedback also contributed to a sense that HEADS-UP provided a non-judgemental forum for team learning:

> 'It was more of a constructive exercise, where everyone is then wanting the same outcome, so it wasn't like, "Oh this didn't happen, I am angry at you", it was more like we are all actually in the same team and we all want the same thing to happen... You all have the same end point and I think that is why it is quite safe to bring it up, because it is not confrontational, and it is not personal against someone, it is just what needs to be done.' *(Foundation doctor)*

> 'There were a few teaching sessions where things were raised, we went through things like ECGs [electrocardiographs] that were missed, and then you had a collective opportunity to think... It was a very good plot for learning, saying this happened and let's do some learning and try and prevent it happening again. That is quite helpful.' *(Foundation doctor)*

> '[A previous teaching session] was really kind of negative, this was kind of just approached very differently... It is very subtle, but the feeling about the feedback that you get is different.' *(Foundation doctor)*

This psychological safety was not limitless: there were still boundaries around what could be discussed. Although the briefings did prompt teams to consider positive notes from the previous shift, they rarely did so. When they did record something positive, their reflections often

concerned the overall management of patient flow, rather than specific actions or diagnostic processes that others could emulate:

> 'It did sometimes feel like we were just reporting all the negatives, and it would have been good for team morale to introduce some positives.' *(Foundation doctor)*

> 'What people do focus on is how many discharges you had, because that is what is prominent - wherever you are. 'We had seven discharges that day', that would be the sort of thing that went on it. But I am sure there are other things, that... could quite easily go on there.' *(Senior nurse)*

> 'Maybe it's just I am also still learning to flag up what went well the day before, because we are all good with, "Ah, this was frustrating yesterday..."' *(Senior nurse)*

This reflected a reluctance to delve too deeply into any one team member's performance. Professional identities were protected by assigning problems to other departments, acknowledging procedural complications as foreseeable, or linking delays to under-staffing. Thus, perceived threats to professional identity were largely deflected by a projection of blame onto other groups, or attributed to circumstances beyond the team's control. More introspective teams, with senior support and increasing psychological safety, did record issues like diagnostic errors, in which they had played a more overt role. However, even in those teams - where the briefings were implemented with greatest fidelity - there could be disputes as to whether certain problems were really problems at all:

> 'The senior nurse disagrees with medical teams' perceptions that the site moves [moving patients from one ward to another] were inappropriate. She thinks doctors expect patients to remain on the acute medical unit even when they are stable... She felt an inadequate handover perpetuated their concerns.' *(Fieldnotes)*

Thus, whilst psychological safety was an important factor in helping staff speak up about

their concerns, certain topics remained off-limits, and psychological safety did not guarantee agreement about what were reasonable concerns. These are important limitations.

6.5.1.3 A hard edge of accountability

The briefings did maintain a 'hard edge' for their effectiveness. Team members regulated their own behaviour, knowing that their actions (or inactions) might be flagged up at the next briefing, or in feedback. Junior doctors and nurses were both conscious of this internalised discipline ('panopticism'), acknowledging how it changed their own behaviour and how they could use it to their advantage:

> 'The group was more aware that if you perhaps missed something like that, it may be brought up later at a HEADS-UP meeting... It made you more accountable.' *(Foundation doctor)*

> 'It certainly works for the VTEs [venous thromboembolism assessments, completed by junior doctors]. There are days that we will have eight or 10 VTEs [to do], so if I write there 10 and give the list to the doctor, the following day I will come and there will be one or two, because they know the following day it will be mentioned again, with the consultant.' *(Senior nurse)*

> 'If you didn't then go and report the Datix [online incident report], then you would get an email saying this was mentioned on HEADS-UP and no Datix was completed, and so you had to go back...' *(Foundation doctor)*

Although the briefings emphasised accountability, the teams did not always explicitly agree exactly what would be done, and by when. This contributed to some difficulty in keeping track of the variety of issues they had raised; whether they had been escalated as planned; and if any further action was required. Similar problems have been encountered in other interdisciplinary initiatives, suggesting that teams may actually find comfort in systems that allow them to

avoid accountability. Shift patterns that lacked continuity exacerbated this problem, as briefing leaders neglected to confirm what had been resolved from previous HEADS-UP briefings. Simple adjuncts (like whiteboards) to carry over information from one day's HEADS-UP briefing to the next may have been helpful:

> 'I think it's more [than] just raising the issue... Give me a more detailed timeframe, resolution, how do you resolve it? Are you going to get back to me because it's important? And what will I expect?' *(Senior nurse)*

> 'It's just being flagged, flagged countless times by different people, same issues, but no-one is getting back as to what has been done...' *(Senior nurse)*

Others had more positive experiences where, held to account, their colleagues confirmed the outcomes of previous briefings:

> 'You get more feedback, so you raise concern[s], and then maybe the next day the same issue would crop up, and you actually then get feedback on what they are doing about it, or who they have escalated to.' *(Foundation doctor)*

6.5.2 Rapid resolution and meaningful managerial follow-up

6.5.2.1 Problems with the existing incident reporting system

Several staff members drew comparisons between HEADS-UP and the pre-existing processes for identifying and resolving ward-level problems. Previously, the hospital's online incident reporting system (Datix, London, UK) was considered the formal mechanism by which they could record problems. Staff did not find it easy to use the system for this purpose, and would not persist with reporting recurrent problems. Issues were so commonplace that they were no longer considered abnormal ('normalised deviance'):

> 'Some things staff have just got so used to that they don't Datix [report] it. So it's just become common practice...' *(Clinical risk manager)*

Managers had limited time to attend to the online reports that were submitted, and did not find them useful for identifying underlying problems. Long processing delays meant that even relatively simple problems were not reviewed for many weeks. As a result, reports that described patient harm were prioritised; where no harm had occurred, little was done:

> '[Before HEADS-UP] you could go down there [to the ward] and you could see people working very, very hard and you could also sense a lot of frustrations... but you weren't getting to what those frustrations were, and what those issues were.' *(Service manager)*

> 'On Datix it used to take weeks on end, whereas it might just be the hoist is broken, whose responsibility is it to make sure it is sorted out?' *(Senior nurse)*

> 'The only way [learning] would have happened would have been directly from a SI [serious incident][3]... it has got to that point of being a SI before it is then addressed and taught and learnt from.' *(Service manager)*

6.5.2.2 Triage of minor issues and thematic identification of deeper problems

HEADS-UP helped identify a route for more rapid resolution of practical problems. It provided an acceptable mechanism for staff to log issues into which which they had immediate insight:

> '[HEADS-UP] highlighted little things, that could be sorted out quite quickly, it made sure that somebody was allocated to deal with that on that day, or did you follow that through, did that get sorted out? So everything benefits, rather than just going on and on and on, and it never happens. Things get sorted out quicker.' *(Senior nurse)*

[3]Serious incidents are adverse events where the consequences to patients and families, or staff, are so great that a heightened response is justified. These may include, for example: unexpected deaths; serious avoidable injuries; or incidents that threaten public confidence in a healthcare service (NHS England, 2015).

> 'Things that have been reported on Datix that sometimes we have to investigate in depth... [with] HEADS-UP, at least right then and there you can actually sort it out immediately.' *(Senior nurse)*

In some cases, where there was no realistic prospect of an immediate definitive solution to a problem, HEADS-UP was then used to confirm a plan for mitigation. For example, staff raised the concern that the available magnetic resonance imaging (MRI) equipment could not accommodate their obese patients. This equipment could not be replaced at short notice, but having escalated the problem, an alternative plan still needed to be found:

> 'The bariatric patients that couldn't have a MRI... They ended up having the orthopaedics [team] still start the patient on IV [intravenous] antibiotics... You have to have a different plan, even if that investigation is not going to happen... So whether or not it was [osteomyelitis, a serious bone infection], she's on the right treatment already...' *(Senior nurse)*

The briefings, facilitation and feedback presented managers with clear, actionable topics for them to address. They were then better prepared to handle incipient risks to service quality:

> 'What I like about it is [that] it is instant and it is thematic very quickly, you understand what the issues are... Here is an opportunity to dump them [your daily issues] and suddenly a picture forms out of it... And not just here, but there, there, there and there... And it hasn't yet created a significant incident, but clearly there is risk associated with it...' *(Service manager)*

Not all managers were equally enthused. Where HEADS-UP had not provided data that appeared directly useful for them, middle managers did less to hold their service areas accountable for HEADS-UP performance [see chapter 8]. Still, the combination of more effective actions by frontline teams, and increased managerial involvement in at least some areas, led to a more general sense that things had changed for the better:

'[HEADS-UP] would make it clear what was actually deficient. And then I think things did start to change, and you did see the feedback coming through... So you could definitely see the evolution of it.' *(Foundation doctor)*

'There is a very broad range of stuff that comes out of HEADS-UP. I mean there are things that get raised repeatedly, and then sometimes you get feedback on a kind of wider level, something has changed, like they have changed the phlebotomy service or they have changed, they have allocated a new tool or something like that. But there is also the kind of day to day things within the team that start running better, and I think you do start seeing those but it is harder to put a specific, "this was raised, this was done, there was an outcome." It is a more kind of general change as part of HEADS-UP.' *(Foundation doctor)*

'I don't know if it is just HEADS-UP that has done it, but I do see improvements happening every day... I had an eight-month gap off the ward, and [when I returned] I really did feel that the feel of the hospital had changed.' *(Foundation doctor)*

6.5.3 Changing individuals: reversing normalised deviance, and catharsis

6.5.3.1 Reversing normalised deviance

Staff members described how HEADS-UP had changed their personal practice, prompting them to proactively address issues they may have previously ignored. This reversed the normalised deviance that had become ingrained:

'You wouldn't have identified that necessarily, because it is just part of... your daily life in the NHS, but when it is put like this [with HEADS-UP], it is highlighted, you have got an opportunity to really do something about it.' *(Service manager)*

'This particular patient is awaiting echo[cardiography] for the last three days... [I] would now question why is it still not done... [When] I wasn't doing it [HEADS-

> UP], I wasn't seeing the importance... [Now] I tell the Matron, or I tell my Bed Manager... or I go there myself... I go to the department and ask them, "This is what's happening, this is what I need."' *(Senior nurse)*

In addition, individuals used the information shared in the HEADS-UP briefing to actively coordinate their plans and decisions for the next shift:

> 'After HEADS-UP, before we actually all leave, I will tell the physio[therapist] and the OT [occupational therapist]: "These are your priorities for today, these are your priorities for tomorrow." So they already have a plan... Like I said, it gives you the structure that you need.' *(Senior nurse)*

Junior doctors also acknowledged how HEADS-UP briefings contributed to their decision-making. More cognisant that problems were occurring repeatedly, and aware of pressures on their colleagues, they adapted their practice:

> 'The group as a whole became more "present" to [what] happens maybe a couple of times in a week, and it was more in people's minds... For example if the nursing staff are short that day... when you were going around, just having that in your mind and thinking "No, we need to get all these blood tests done now", you would appreciate how the service could be best run. I think that made it more efficient.' *(Foundation doctor)*

6.5.3.2 Catharsis

The improved information sharing, interdisciplinary coordination, and individual efforts to share workload produced a more supportive team climate. In this atmosphere the HEADS-UP briefings became unexpectedly cathartic:

> 'You felt like you had got a lot off your chest by doing it, and if you had a bad day and nothing seemed to be going right because of so many errors that you

came across, it all felt a bit chaotic and scatty, the next day you could say these are the things that went wrong yesterday, and now I can clear that and start again...' *(Foundation doctor)*

'It is kind of a debrief as well, which is something you don't really do in medicine very much, especially on the doctors side. I know from having worked in psychiatry, you debrief all the time, you don't realise how useful it is and how much you learn from it, but we don't tend to. I suppose you see things and things annoy you and you go home and you deal with it, and then you come back in the next day. So it gives an element for that.' *(Foundation doctor)*

'It kind of makes you feel more heard, because otherwise you are just venting your concern, but you don't know if it has been actioned in any way. And then you can provide the objective evidence to move forward, to change something.' *(Foundation doctor)*

6.5.4 Proactive teams and a revealed gap in practice

6.5.4.1 Changes in incident management

The ease of HEADS-UP completion was a popular feature, compared to the multiple data required for an online incident report. Staff did recognise that certain events discussed in HEADS-UP briefings might need to be logged as formal incidents, for which the incident report format was more appropriate. However, no team had an agreed list of specific high priority events to record. This meant that translation of HEADS-UP-recorded incidents into online reports could be erratic:

'If something comes up... then it prompts someone to say, "actually no, I think [a] Datix [incident report] should be done about this"... and then you can gather that *(sic)* data in the already established format.' *(Foundation doctor)*

> 'Post-ERCP [endoscopic retrograde cholangiopancreatography, an intervention associated with a high risk of infection in some cases] antibiotics had not been given... The F1s [junior doctors] had not seen the report specifying they were needed. No DATIX [incident report] was done, even though [the registrar] specifically asked if one should be completed...' *(Fieldnotes)*

When incidents were reported, the HEADS-UP format provided a ready structure for incident investigators to start to collect relevant information. Issues were flagged up at the briefings as the incident reports were submitted:

> 'The people involved are already there, they are already able to explain to you, "This is what actually happened from our end"... [gives example]... It's easier for you to do your incident [investigation] report already because as the investigator you don't have to call everyone.' *(Senior nurse)*

The emphasis on clinical teams' accountability also led to a more subtle change in incident management, that was noted by middle managers:

> 'Very often before it gets to me, it has been dealt with and sorted out, by the team themselves, because it [HEADS-UP] helps them understand what is within their power to deal with... In the past, pre-HEADS-UP, if it was reported on Datix and I think to some degrees it still is, people think that is [the] end of the matter, "I have reported it, it is no longer my responsibility." But with HEADS-UP and Datix together, there is a different approach... We have discussed it, we have taken on the responsibility of this, we have done as much of this as we can, but here is the point where it has to be escalated and we want to make sure something is done about it, and this is our methodology for doing it. That is how it feels to me, anyway.' *(Service manager)*

6.5.4.2 A gap in practice, revealed

There was a wider recognition that the programme had revealed a gap in practice, one that teams had been unaware of. Self-confessed sceptics conceded that they had found HEADS-UP useful, even on wards where team function was already acknowledged to be good:

> 'There is definitely time, there is definitely time to do it, there is definitely a need to do it, but... you only know that once you have done it.' *(Service manager)*

> 'I don't think we would have known, unless we had done it, that there was a gap there.' *(Senior nurse)*

> 'I am [in] A&E [Accident & Emergency department rotation] now, you kind of miss [HEADS-UP] actually. You do kind of miss that opportunity.' *(Foundation doctor)*

Similarly, HEADS-UP was revealing service provision opportunities that had not previously reached board-level attention:

> 'HEADS-UP findings informed the discussion [in the Trust morbidity and mortality meeting, chaired by the chief executive], highlighting areas which needed faster examination and transformation, and bringing genuinely new information... [This was] useful in corroborating areas of intuitive concern, many of which had not been formally addressed either in incident reports or top-down initiatives... [It also] highlighted new targets for improvement.' *(Fieldnotes)*

6.5.5 Organisational change

There were a number of specific examples of tangible changes driven by HEADS-UP within the study period, predominantly at hospital 1 [box 6.1]. These were described as 'quick wins' by senior managers, with clinicians agreeing that they were at least a relatively rapid organisational response:

> 'We've been asking for this [equipment] for several years. This [HEADS-UP] only took five months.' *(Consultant - Fieldnotes)*

Box 6.1 Examples of service investment and organisational changes arising from HEADS-UP

Cost-neutral re-alignment of porter provision for radiology, doubling in-patient flow for X-rays;

Re-evaluation of phlebotomy service provision, ultimately generating a £20,000 investment in an expanded support service;

Additional phlebotomy training for nursing assistants, delivered internally;

Additional sonographer in echocardiography;

Intravenous infusion and nasogastric feeding pumps repatriated to central equipment library;

Development of electronic referrals and electronic reporting for medical specialty consultations, using existing electronic health record;

£5,000 investment in ketone meters for management of diabetic ketoacidosis;

Number of delayed therapy assessments incorporated into organisational status report.

Internal investment was not necessarily targeted specifically to participating wards, however. For example, expansion of the phlebotomy service (identified through HEADS-UP as no longer meeting patient needs) primarily benefited non-participant wards, which were deemed by senior management to be more in need of support. Nor was progress guaranteed by repeated ward-level reporting. Changes to structures or processes were more likely when there was an associated financial target; another organisational incentive aligning with improvement; and when clinicians and managers agreed the need for change. HEADS-UP data then enabled managers to make a convincing case for investment in service quality:

> 'It empowers me then as manager for that area to go forward with a business case with the evidence to say we need additional resource, this is the implications of

> it... And it works, so we have an additional [cardiac] sonographer... I think we saw very quickly a reduction in delays once that was sorted.' *(Service manager)*

However, progress proved challenging when a proposed change would impact workflow in more than one clinical area. Additionally, some of the concerns raised by frontline staff simply could not be addressed by middle managers, general managers, or even board members. Poor staffing levels, for example, were felt keenly in clinical areas, but reflected a broader national challenge. Creative workarounds and mitigation plans could only go so far in addressing a fundamental shortfall in staff. This was a source of frustration for teams completing the briefings:

> 'The staffing every single day was always... "we don't have enough nurses", so that just kind of became a slightly pointless box. And we had seen it the day before, the day before, the day before.' *(Foundation doctor)*

HEADS-UP also identified issues like unsafe inter-hospital transfers, which - whilst amenable to change - would have required a degree of close collaboration between a network of regional hospitals. The necessary cooperation and drive was not always forthcoming, even when existing processes did not meet pre-agreed standards:

> 'The Director of Operations is formulating an inter-hospital transfer agreement, based on London-wide agreed standards, referring to multiple cases flagged by HEADS-UP... She may need to invoke the Clinical Commissioning Group to sort out local transfer issues.' *(Fieldnotes)*

6.5.6 Safety & teamwork climates

115 surveys were completed; response rates were 71.8% (61/85) in phase 1 and 65.1% (54/83) in phase 2. 97.7% safety and teamwork items were completed; demographic data were incomplete in 14.8% (17/115) cases and HEADS-UP participation data in 3.5% (4/115) cases [table 6.1].

Unadjusted mean safety and teamwork scores were both higher after HEADS-UP participation, but not significantly so [table 6.2]. HEADS-UP influenced adjusted safety and teamwork

Table 6.1: Survey respondents and completion rates

	Phase 1	Phase 2
Professional background [n (%)]		
Consultant physician	8 (13.1)	5 (9.3)
Junior physician	28 (45.9)	35 (64.8)
Senior nurse	15 (24.6)	7 (13.0)
No answer	10 (16.4)	7 (13.0)
Median time on the ward [months]		
	4-6	0-3
Ward specialty [n (%)]		
Acute medicine	19 (31.1)	10 (18.5)
Gastroenterology & general medicine	13 (21.3)	11 (20.4)
Heart failure & general medicine	4 (6.6)	5 (9.3)
Geriatrics	16 (26.2)	14 (25.9)
Respiratory & general medicine	9 (14.8)	12 (22.2)
No answer	0	2 (3.7)
Participated in more than 5 HEADS-UP briefings [n (%)]		
Yes	13 (21.3)	37 (68.5)
No	47 (77.0)	14 (25.9)
No answer	1 (1.6)	3 (5.6)
NASA-TLX (workload) score [mean (SD)]		
	13.2 (2.7)	12.8 (3.1)
Unscored safety and teamwork items [n (%)]		
	33/1647 (2.0)	37/1458 (2.5)
Unscored NASA-TLX items [n (%)]		
	21/366 (5.7)	18/324 (5.6)

scores differently over time at each of the two hospitals (p=0.021 for safety; p=0.004 for teamwork). Participants' safety scores were lower in the early stages of HEADS-UP implementation (i.e., at phase 1 for hospital 1, or phase 2 for hospital 2, where HEADS-UP was only embedding into routine practice as the second survey period began). At hospital 1, after six months of sustained implementation, participant safety scores increased above those of non-participants (estimated marginal mean safety score 70.3 vs 61.0). Similar findings were observed for teamwork scores, with significant HEADS-UP-associated improvement over time only at hospital 1 (estimated marginal mean teamwork score 81.7 vs 70.0). At phase 2, 71.9% (23/32) of hospital 1 respondents agreed they had reported concerns differently as a result of the programme. Of note, higher workload scores correlated with lower safety scores (B= -1.52, 95% CI -2.59 – -0.44, p=0.006) but not teamwork scores (B= -0.21, 95% CI -1.10 – 0.68, p= 0.64).

6.5.7 Submitted incident reports

6.5.7.1 Total incident reports

More incident reports were submitted in intervention months than control months (18.8 incident reports/month *vs* 15.5 incident reports/month). In the adjusted mixed-effects model, HEADS-UP had no intention-to-treat effect on this outcome (IRR = 0.91, 95% CI 0.76 - 1.09, p = 0.305). Implementation fidelity significantly affected the result (χ^2 = 14.80, p < 0.001). High fidelity HEADS-UP significantly increased monthly incident reports (IRR = 1.33, 95% CI 1.15 - 1.55, p < 0.001). These findings were replicated in the GEEs secondary analysis (IRR for high fidelity HEADS-UP 1.34, 95% CI 1.19 - 1.51, p < 0.001).

6.5.7.2 High yield incident reports

In the mixed-effects analysis, similar improvements were seen with high yield incident reporting - whether defined as non-falls incident reports, or increasing engagement by non-nursing staff. Increases in high yield incident reporting were identified with per protocol, rather than intention-to-treat, analyses. For non-falls incidents, HEADS-UP overall did not increase reporting (IRR =

Table 6.2: Effect of HEADS-UP on safety and teamwork climate scores

		HEADS-UP		P value
		Participants	Non-participants	
Mean safety climate score (SD)				
		68.4 (13.6)	65.9 (16.9)	0.667
Mean teamwork climate score (SD)				
		80.2 (12.1)	79.6 (12.3)	0.524
Estimated marginal mean safety climate score [95% CI]				
Hospital 1	Phase 1	59.8 [51.4-68.1]	69.2 [64.6-73.9]	
	Phase 2	70.3 [65.3-75.2]	61.0 [50.8-71.2]	
Hospital 2	Phase 1	70.8 [51.1-90.4]	54.9 [46.2-63.7]	**0.021**[*]
	Phase 2	71.0 [56.0-86.0]	75.1 [63.5-86.7]	
Estimated marginal mean teamwork climate score [95% CI]				
Hospital 1	Phase 1	72.3 [65.3-79.2]	82.6 [78.8-86.5]	
	Phase 2	81.7 [77.6-85.8]	70.0 [61.6-78.4]	
Hospital 2	Phase 1	84.8 [68.5-100.0[**]]	74.5 [67.2-81.7]	**0.004**[*]
	Phase 2	88.2 [75.7-100.0[**]]	83.3 [73.7-92.9]	

[*] *p* value for model effect of the interaction term HEADS-UP participation[*]hospital[*]time

[**] The upper bound of the 95% confidence interval for estimated marginal mean climate scores was truncated at 100, the limit of the scale (Cowen and Ellison, 2006).

0.91, 95% CI 0.73-1.13, p = 0.400). Implementation fidelity was a significant factor (χ^2 = 16.50, p < 0.001), with high fidelity implementation significantly increasing non-falls reports (IRR = 1.41, 95% CI 1.18 - 1.69, p < 0.001).

For non-nursing staff reports, the same pattern was identified. High fidelity implementation of HEADS-UP significantly increased engagement with incident reporting, with more reports submitted each month by non-nursing staff (IRR = 1.60, 95% CI 1.08 - 2.36, p = 0.019). Intention-to-treat analysis identified no significant effect when implementation fidelity was overlooked (IRR = 1.04, 95% CI 0.62-1.74, p = 0.878; χ^2 for implementation fidelity 11.23, p = 0.004).

These results were robust to the secondary statistical approach, using GEEs. For non-falls incidents, high fidelity HEADS-UP increased monthly reports (IRR = 1.42, 95% CI 1.25 - 1.63, p < 0.001). Non-nursing staff reports also increased with high fidelity implementation (IRR = 1.54, 95% CI 1.07 - 2.23, p = 0.021). Intention-to-treat analyses showed no significant effect (p = 0.313 and p = 0.795 for non-falls incidents and non-nursing staff submissions, respectively). Implementation fidelity was a significant factor in each case (χ^2 =30.79, p < 0.001 for non-falls incidents, χ^2 = 11.44, p = 0.003 for non-nursing staff reports).

6.6 Discussion

This mixed methods evaluation of the HEADS-UP intervention expanded on its impact on patient outcomes, as described in chapter 5. The qualitative analysis suggested mechanisms through which HEADS-UP exerted its effects, as well as opportunities to refine the intervention for broader dissemination. Together, qualitative and quantitative data have provided empirical support for most, but not all, of the study's initial hypotheses [figure 6.1]. There was also evidence of the anticipated virtuous cycle, in which tangible changes arising from frontline HEADS-UP briefings motivated further engagement with those briefings [see figure 4.5].

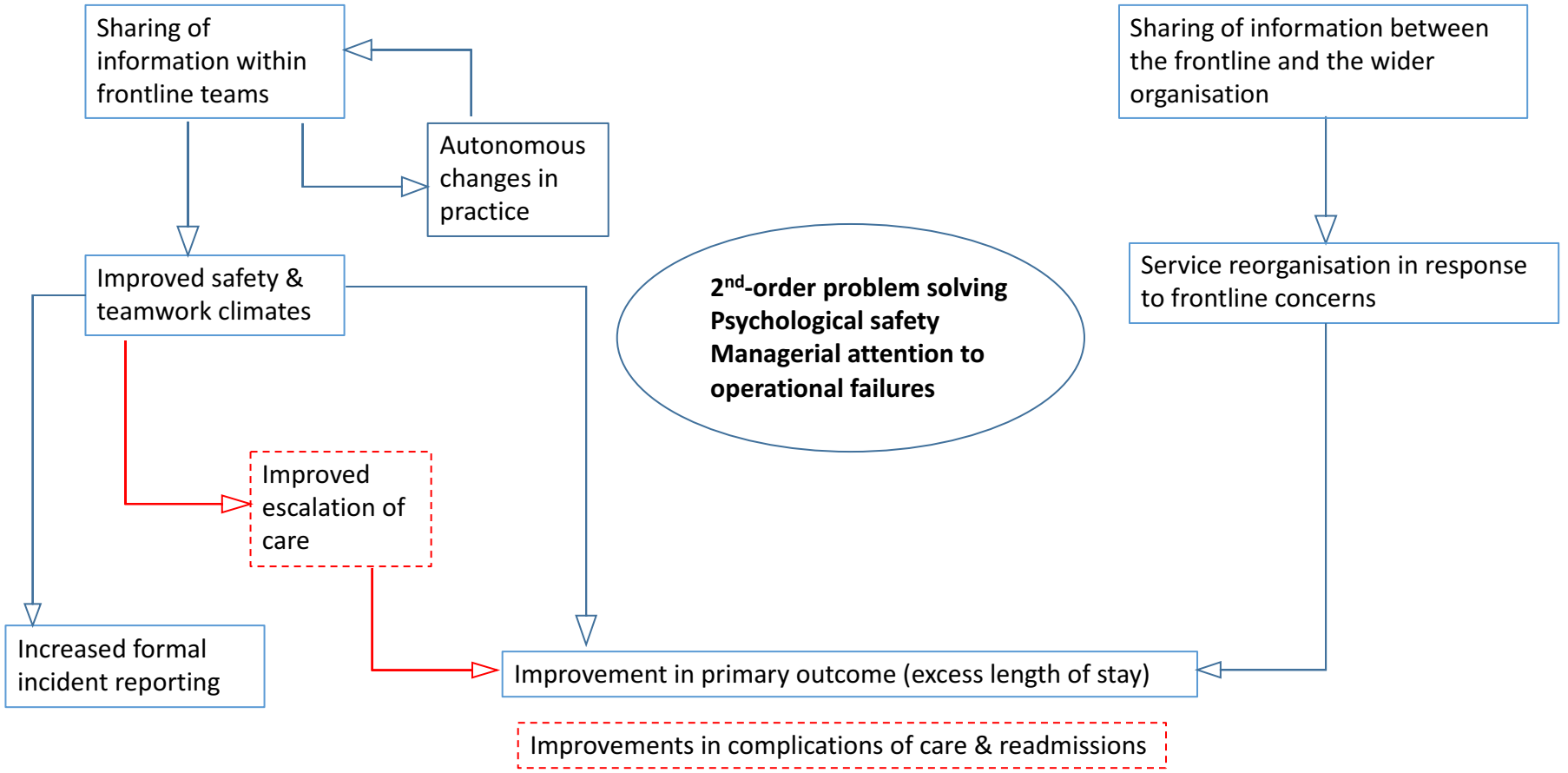


Figure 6.1: Empirical support for HEADS-UP study hypotheses

The figure summarises whether the study's hypotheses found empirical support [blue boxes], or were not supported (or only partially supported) by the data [dashed red boxes]. The three proposed mechanisms of action for PCTS [blue circle] were supported by the data.

6.6.1 Summary of findings

6.6.1.1 Mechanisms of action

The HEADS-UP briefings engendered participation from a broad range of clinical staff. This generated a non-judgemental forum in which concerns could be raised, and interdisciplinary learning could take place. A hard edge of accountability, of which both doctors and nurses were aware, encouraged staff to act on the issues that they had discussed. Minor issues were triaged to rapid resolution, whilst HEADS-UP provided useful deeper themes for managers to pursue. A lack of continuity between briefings limited the impact for some frontline staff, as did failures to secure explicit agreement on the plan of action. Managers needed directly useful data before endorsing the programme.

6.6.1.2 Individual and team effects

Individuals found themselves attending to issues they would have previously ignored, reversing the normalisation of deviance. Briefings took on a supportive, cathartic element, and teams became more proactive in autonomously managing incidents. Participation in HEADS-UP persuaded staff that similar programmes might be required for optimal interdisciplinary practice.

6.6.1.3 Organisational changes

HEADS-UP generated numerous, tangible developments, including cost-neutral service changes, internal service investments, and training for individuals. Some concerns could not be resolved at an organisational level, and reflected wider challenges that required regional or even national coordination to resolve.

6.6.1.4 Safety and teamwork climates

HEADS-UP was initially associated with lower safety and teamwork climate scores, but the scores significantly increased after sustained implementation. Similar temporal patterns have

been seen with other improvement interventions (Muething et al., 2012). Significant improvements were seen only at hospital 1, where a critical mass of wards had adopted the intervention.

6.6.1.5 Incident reporting

High fidelity HEADS-UP implementation increased total incident reports. There were significant increases in high yield incident reporting, as judged both by non-falls reports and increasing reports submitted by non-nursing staff. The intervention had no detectable effect in intention-to-treat analyses, which ignore the fidelity of implementation.

6.6.2 Limitations

There was disproportionate access to the teams and individuals who engaged with the HEADS-UP programme, as opposed to those who declined to meaningfully participate. This was the case both for auto-ethnography, and attendance at the focus groups. These enthusiasts may have had a more positive view of the programme than the wider group of eligible clinicians and managers. Selection bias has hampered mixed methods evaluations in previous reports (Goldenhar et al., 2013). The high uptake of quality improvement interventions by teams that are already performing well, with less to improve, has also been noted in other studies (Rotter et al., 2010).

However, the response rates to the surveys were high, suggesting that those results were likely to be representative of the broader safety and teamwork climates. Additional attempts to counteract any selection bias included leadership of the focus groups by researchers who had not previously worked with programme participants, and a deliberate search for conflicting opinions in the group transcripts and fieldnotes.

Some of the detectable impact of the intervention may have been reduced by contamination between groups. HEADS-UP-generated organisational support was not necessarily directed to the areas that had first raised concerns, and participants' incident reports may have been targeted to areas other than the study wards. These elements of the intervention could not be

controlled. Differences in the degree of organisational commitment to HEADS-UP also limited comparisons of its effect at the two sites. It is unclear whether HEADS-UP is by definition better suited to smaller, district general hospitals, or whether its greater impact merely reflected a higher number of participating wards.

Lastly, though HEADS-UP was delivered and evaluated in its own right (rather than in a package of multiple interventions), it did have multiple components. The relative contribution of each of these components (structured briefings, facilitation and feedback) is unclear from this analysis. Managers and clinical leaders may be interested in the likely impact of an intervention which is made still more frugal, i.e., reliant on frontline briefings alone.

6.6.3 Implications for further work

Future work should assess the importance of facilitation and feedback in the effectiveness of the HEADS-UP intervention. There are also opportunities to refine the intervention further, as staff identified some components of the intervention that did not work well. Building on the quantitative analysis in chapter 5, the data in this chapter lend further support to a broader dissemination of PCTS strategies.

6.6.4 Conclusions

A coherent set of mechanisms of action underpin the apparent effectiveness of PCTS. However, its effects are seen only with high fidelity implementation. In the next chapter, I characterise the study's implementation context, variations in HEADS-UP implementation, and the facilitators and barriers to its optimal use. This will help inform efforts to implement PCTS in other settings.

Chapter 7

PCTS: an implementation evaluation

7.1 Introduction

In the last chapter, I described how PCTS, as implemented in the HEADS-UP intervention, established a non-judgemental interdisciplinary forum that was able to hold individuals to account. It proved effective in provoking certain organisational changes, as well as changes in individual and team practice. These were reflected in improved safety and teamwork climates, and improvements in meaningful incident reporting.

The effects of the intervention were clearly dependent on the quality of its implementation: this was apparent in both quantitative and qualitative analyses [chapters 5 and 6, respectively]. Understanding the extent to which implementation was successful, and why, contextualises these results: perhaps a different implementation model might have proved more effective. A richer understanding of the specific implementation challenges for PCTS is also an important step towards a scalable strategy that can be usefully embedded in routine practice.

7.1.1 Characterising implementation

The quality, or success, of implementation is often conceptualised as implementation fidelity, i.e., the degree to which a programme's implementation adheres to its designers' intentions (Peters

et al., 2013; Slaughter et al., 2015). This fidelity can be detailed at both 'implementer' level (here, frontline briefings and managerial use of the data generated by them) and 'programmatic' level (the degree to which the programme's implementation strategies were used as preconceived). The World Health Organisation (WHO) defines an additional seven implementation outcomes (Peters et al., 2013), which provide a more comprehensive assessment of implementation. I draw on these broader implementation outcomes in this chapter.

7.1.2 Describing context

The implementation and effectiveness of any complex intervention are intricately linked to the context in which that intervention takes place. Often defined metaphorically, context might best be thought of as a 'garden, terrain or domain', determining whether QI efforts flourish or wither (Bate, 2014). The Medical Research Council holds that context 'includes anything external to the intervention that may act as a barrier or facilitator to its implementation, or its effects' (Moore et al., 2015). Though difficult to operationalise, context can be subdivided into strategic, cultural, technical and structural elements. Inattention to any one of these components reduces the likelihood of an intervention's long-term success (Shortell et al., 1998). The history of previous QI efforts also leaves an organisational imprint; this 'temporal context' too should be recognised (Bate, 2014).

There have been attempts to quantitatively measure the various elements of context, and the linkages between them. What is more important is the way in which people perceive that context, and make sense of it, in a dynamic and adaptive landscape (Dijk, 2009; Bate, 2014). Pettigrew et al. (2001) claimed that the real challenge was to examine 'how and why constellations of forces shape the character of change processes', rather than 'fixed entities with variable qualities'. Ethnography, participant interaction, and focus groups are recommended as key tools for this type of investigation (Bate, 2014).

These methods help describe both the inner (organisational) context and the outer context (factors beyond the organisation, such as social systems, laws and regulatory environments).

Contextual influences can also be grouped by the system level at which they they act: macro-, meso- or micro-. These come together to generate contexts which are largely receptive (favourable) or non-receptive (resistant) to change efforts (Pettigrew et al., 1992). This overview of local conditions establishes how context might affect all change efforts, before considering individual QI interventions and their specific requirements.

Here, I re-analyse the auto-ethnography and semi-structured focus groups described in chapter 6, seeking a better understanding of the implementation context; the facilitators and barriers for PCTS as an improvement strategy; and comprehensive implementation outcomes. These will be assessed with respect to hospital 1, which contributed the majority of the participating wards to the HEADS-UP trial.

7.2 Aims

(i) To describe the context in which the HEADS-UP trial took place, identifying salient characteristics that affected the use of the intervention;

(ii) To characterise HEADS-UP fidelity at implementer and programmatic levels during the study period;

(iii) To identify specific facilitators and barriers for PCTS; and

(iv) To evaluate other HEADS-UP implementation outcomes, as defined by the WHO.

7.3 Methods

7.3.1 Data sources

As described in 6.3.1, fieldnotes from intensive auto-ethnography over the study period were collated. In addition, two semi-structured focus groups were conducted in July 2015, assessing

the qualitative impact of HEADS-UP, as well as staff perceptions of its implementation. I will now describe how the implementation components of the focus group topic guide were established.

7.3.1.1 Topic guide development

The focus group topic guide was informed first by Greenhalgh's diffusion of innovations model (Greenhalgh et al., 2004). This model comprehensively describes factors influencing the adoption of innovative health delivery practices. Although the model has been widely used, it is not easily applied to controlled trials, and not all elements are equally relevant (McMullen et al., 2015).

A more recent report from the Health Foundation - *'Constructive comfort: accelerating change in the NHS'* (Allcock et al., 2015) - was thought to be more directly applicable. *'Constructive comfort'* synthesises numerous contextual frameworks, drawing on relevant literature, professional testimony, and the Health Foundation's own decade-long experience of funding improvement programmes. It concludes that there are seven factors for successful change at any level of the NHS, particularly for local organisations [box 7.1].

These success factors operationalise relevant features of the inner and outer context, translating similar components of other models for change in healthcare. However, to my knowledge, the *'Constructive comfort'* framework has not yet been empirically tested. The implementation analysis topic guide [appendix D] therefore drew on this framework for two reasons. First, it is an evidence-based tool that distils context specific to change management in modern NHS practice. Second, there was a novel opportunity to assess the utility of the framework itself.

7.3.2 Analytical approaches

As described in 6.3.1.3, fieldnotes and focus group transcripts were managed in NVivo. Two complementary approaches were used, based on an initial reading of the data by two researchers[1]. For the fieldnotes, a deductive approach was used, coding and aggregating evidence according to the seven factors in the *'Constructive comfort'* framework. This generated an understanding of

[1] I conducted the analysis with Dr Stephanie Archer, Research Fellow.

Box 7.1 Seven evidence-based success factors for change in the health system (Allcock et al., 2015)

Committed and respected *leadership* that engages staff
A *culture* hospitable to, and supportive of, change
Management practices that ensure execution and implementation
Capabilities and skills to identify and solve problems
Data and analytics that measure and communicate impact
Resources and support for change
An *enabling environment* which supports and drives change

the local implementation context, i.e., those factors likely to affect any improvement endeavour within it.

Similar deductive approaches were then used to evaluate implementation fidelity. [Quantitative measures of fidelity, based on the frequency of briefings and the documentation of issues within those briefings, were described in 5.4.2.] For programmatic fidelity, a recently-published panel of 73 implementation strategies (Powell et al., 2015b) was used to identify those strategies that had been deployed during the study period. Fieldnotes and transcripts were then screened for evidence of fidelity to those strategies.

By contrast, the focus group transcripts did not map well to the *'Constructive comfort'* framework. An inductive (theory-generating) approach was therefore used to analyse the focus groups, as per the methodology described in the previous chapter (Braun and Clarke, 2006). The researchers individually coded each transcript to generate coding frames, which were subsequently refined and re-coded, to produce agreed *de novo* higher-order themes. These dual approaches (deductive and inductive) were brought together to summarise the WHO's implementation outcomes [table 7.1][2].

[2]Table 7.1 is reproduced, with permission, from: Implementation Research in Health: A Practical Guide. Peters D, Tran NT, Adam T. What is Implementation Research, p.30, copyright 2013. Available from http://www.who.int/alliance-hpsr/resources/implementationresearchguide/en/ (accessed 22/07/2016).

Table 7.1: WHO implementation outcomes (Peters et al., 2013).
Reproduced with permission

Implementation outcome	Working definition	Related terms
Acceptability	The perception among stakeholders that an intervention is agreeable	Comfort, relative advantage, credibility
Adoption	The intention, initial decision, or action to try to employ a new intervention	Uptake, utilization, intention to try
Appropriateness	The perceived fit or relevance of the intervention in a particular setting	Relevance, perceived fit, compatibility, trialability, suitability, usefulness, practicability
Cost	The incremental cost of the delivery strategy. The total cost of implementation would also include the cost of the intervention itself	Marginal cost
Coverage	The degree to which the population that is eligible to benefit from an intervention actually receives it	Reach, access, service spread or effective coverage [focusing on those that need an intervention and its delivery at sufficient quality, thus combining coverage and fidelity], penetration [focusing on the degree to which an intervention is integrated in a service setting]
Feasibility	The extent to which an intervention can be carried out in a particular setting or organisation	Practicality, actual fit, utility, suitability for everyday use
Fidelity	The degree to which an intervention was implemented as it was designed in an original protocol, plan, or policy	Adherence, delivery as intended, treatment integrity, quality of programme delivery, intensity or dosage of delivery
Sustainability	The extent to which an intervention is maintained or institutionalised in a given setting	Maintenance, continuation, durability, institutionalisation, routinisation, integration, incorporation

7.4 Results

The relevant data set included 930 HEADS-UP briefings, 2 semi-structured focus groups with 15 staff (junior doctors [n=3], nurses [n=8] and managers [n=4]), and 44 pages of ethnographic fieldnotes. The results will be reported as follows: local context, as defined by the *'Constructive comfort'* framework; implementation fidelity; specific facilitators and barriers for PCTS; and other implementation outcomes.

7.4.1 Local context

The local context was one in which enthusiasm for change, and receptivity to it, were undermined by a lack of improvement resources; the outsourcing of change management skills; significant turnover of experienced staff; and an emphasis on standardisation and central control. These factors will be discussed in more detail.

7.4.1.1 Resources and support

'Constructive comfort' asks whether there are sufficient resources and support for improvement. Yet, over much of the study period, there appeared to be insufficient health economy resources for frontline teams to comfortably deliver standard care, let alone invest in change. This sustained pressure was manifest in repeated hospital declarations of a 'black alert', indicating that bed capacity had been exceeded, affecting both elective procedures and emergency work. The strain on staff was at times immense, with clear implications for the quality of care they could provide:

> '[The staff member] disclosed severe stress due to workload, admitting tearful breakdowns in front of other clinicians, and reiterated the desire to resign. It was striking that an experienced, capable colleague had been brought to this level by workload pressures...' *(Fieldnotes)*

> 'If you told me I'd missed a microcytic anaemia [a laboratory abnormality that may indicate a serious underlying condition, such as colon cancer], I wouldn't be surprised. You can't keep up your [energy] levels for that many patients.' *(Registrar - Fieldnotes)*

> 'A long-awaited meeting with A&E (Emergency department) and AMU consultants, general manager and clinical director has been cancelled due to bed pressures. They are battening down the hatches... The Trust is challenging the [regional] decision to close nearby A&Es. This is a challenging environment for quality improvement, to say the least.' *(Fieldnotes)*

Staff shortages, as described in the previous chapter [see 6.5.5], formed a constant backdrop to ward care. These shortages persisted despite efforts to recruit additional nurses, nationally and even internationally; there were also shortages of non-clinical support staff. Managers too found themselves stretched. Senior clinicians and managers openly admitted that, with no dedicated staff to work on improvement, they had no slack capacity to attend to it themselves:

> 'The general manager discussed how there is little improvement capacity in the department. There are no funded or protected sessions, it's all done *ad hoc*. A business case for a support person or dedicated session could be looked on favourably.' *(Fieldnotes)*

> 'They need dedicated, paid time to do safety work - it is not something that can be added on to an existing staff member's job description without something else giving way.' *(Consultant - Fieldnotes)*

The limited resources for QI work were further strained by competing priorities and long-standing commitments. Preparation for an inspection by the CQC was particularly arduous, the volumes of information needed for regulatory assurance requiring significant managerial time. The opening of a new ambulatory emergency care centre also placed demands on the

available managerial resources. In the face of these pressures, managers tended to invest their time (and funds) in responding to current demands, rather than planning future improvements. As the study period progressed, attention then turned to the forthcoming organisational merger. Again, preparations for this were intensive, leaving clinicians and managers little option other than to withdraw from other projects.

Nonetheless, QI investment was forthcoming at times. Where a case could be made that patients and staff were poorly served by current arrangements, *and* that focal investment might ease overall inpatient flow, middle managers and board executives were receptive. Examples of financial commitments achieved through HEADS-UP are included in box 6.1. Indeed, at times it seemed easier to marshall a financial investment (even in this setting) than a more fundamental change in pathways or processes [see 7.4.1.2].

7.4.1.2 Capabilities and skills

With little time available to them for quality improvement, it was difficult to evaluate whether senior staff truly had the required capabilities and skills to deliver change. As they would readily admit themselves, they were more preoccupied with day-to-day 'firefighting'. Strategic operational changes could not be dealt with by these overstretched permanent staff: change management was outsourced to a striking extent. External contractors were employed to assess inpatient flow and discharge delays, establish novel care pathways, and deliver numerous transformational changes at the interface between hospital and community.

These external contractors described extensive experience of change management, and successfully achieved several of their goals. However, it is debatable whether their skill sets contributed to the hospital's overall receptiveness to change. The contractors' remit was limited to their pre-agreed projects, and they did not participate in routine governance, safety and quality fora. Their skills and insight in pathway design and project management were not available to those seeking improvements in escalation of care, or sepsis management. The perception of a lack of improvement capacity suggests that these contractors were not seen as

an improvement resource for clinical departments or managers [see 7.4.1.1].

When clinicians and managers faced difficult challenges, the project management skills and interpersonal abilities deemed crucial in *'Constructive comfort'* - such as conflict management and negotiation - were not always evident. There was little internal support to help navigate antagonistic reactions to change, for example. The hospital's improvement team had more experience in governance than QI *per se*:

> 'There was a lot of resistance from these consultants, variously denying that this was a significant problem, or that they should be expected to deal with it as a department... I was surprised by the relatively passive stance of Dr X [senior clinician] and Ms Y [senior manager]. It was unclear whether their passivity reflected blurred ownership of the meeting, a pragmatic acceptance of their colleagues' highly antagonistic stance, or uncertainty about the importance of the proposed changes.' *(Fieldnotes)*

For more junior clinical staff, a central clinical leadership programme introduced quality improvement techniques, and the tools to deploy them. The programme had limited capacity, accepting 10-15 applicants per year, most of whom were junior doctors. Although a small number of allied health professionals also participated, there was no hospital-wide expectation that staff should be equipped with these improvement skills. The projects which participants chose had nominal senior sponsorship; in practice, there was little coordination between the projects and wider organisational priorities, and they had limited visibility. Other junior doctors developed their improvement skills through schemes such as the Royal College of Physicians' 'Learning to make a difference' programme, but this was not a core requirement for their job progression.

7.4.1.3 Leadership

In the long build-up to the organisational merger, many senior staff moved to other institutions, reducing the sense of coherent leadership at the hospital. The study period (December 2013 - February 2015) saw three directors of nursing, three general managers for medicine, two deputy general managers for medicine, two directors of workforce and development, and two divisional directors. Other key middle managers and senior clinical managers also left the hospital, or found themselves with additional responsibilities to cover departing colleagues [box 7.2].

With senior leadership in such flux, it was difficult for them to engage with staff on a patient-centred vision for change. Many of the incoming leaders were external appointees, building relationships entirely anew, and then moving on again. Even when long-standing staff remained, a major preoccupation was operational stability during the merger period, whilst new processes and procedures were established for the larger organisation. Paradoxically, the merger - by definition an enormous organisational change - was actually a stable focal point for rapidly transitioning senior staff. Those senior leaders were less able to support discrete ward-level programmes, even those that aligned with their own values.

7.4.1.4 Culture

Safety culture is pragmatically defined in *'Constructive comfort'* as 'the way things are done around here': a 'healthy' culture harnesses staff commitment to patient care with a positive attitude to change. More formally, safety culture is the 'product of individual and group values... that determine the commitment to an organisation's safety management' [box 1.2]. Whether or not organisational culture was healthy in this setting proved difficult to decide, not least because different units within the organisation varied so widely in their attitudes. Significant variation within hospitals has been widely documented elsewhere [see 1.5.1], and contrasts between neighbouring teams - and between senior managers and frontline staff - were prominent here.

Positive elements of culture were evident: senior clinical managers took clear pride in

Box 7.2 Senior staff turnover during the study period

The following roles changed hands during the study period (or shortly after it), or were covered by other staff due to departures:

- Chairman (2);
- Director of nursing (3);
- Director of workforce and development (2);
- Medical director (2);
- Divisional director of surgery (2);
- General manager for medicine (3);
- Deputy general manager for medicine (2);
- Service manager for medicine (2);
- General manager for clinical support services (prolonged leave);
- Head of information governance (unfilled post);
- Lead for incident reporting (unfilled post);
- Matron for emergency care (role redefined);
- Service manager for A&E (2); and
- Portering service manager (3).

maintaining high standards. For example, an experienced matron made sure to support a new nurse in her dispute with a junior doctor, insisting that high risk medications must be prescribed diligently:

> 'Verbal prescriptions for insulin... That's not how we do it here.' *(Matron - Fieldnotes)*

Other teams worked to combine meaningful local audits with service development, drawing on relevant regional publications for inspiration. Still others demonstrated great willingness to reflect on their own practices as a department, and initiate trials of change. Yet these positive traits did not permeate evenly throughout the organisation. Certain teams were consistently difficult to engage in quality improvement efforts, and some middle managers had become nihilistic about the possibility of improvement. When it came to the underlying reasons for delays and inefficiencies, senior managers disagreed with the assessments of their colleagues at the frontline, with a distinct lack of common ground:

> 'We have two chances of getting [this done] - fat chance and no chance! We are monitoring [this] on a daily basis but... the situation is far from ideal and we have escalated up the chain - not that it has helped.' *(Support service manager - Fieldnotes)*

> 'The general manager's perceptions of discharge delays are very different to the ward teams': when delays are appropriately escalated, her experience is that they are rapidly resolved. Problems with wrong equipment, transport delays, etc., are more likely to be issues with the wards' requests than true problems with other services...' *(Fieldnotes)*

Cultural attitudes to risk taking depended not only on the team involved, but the specific proposal at hand. Major service changes, such as the creation of an ambulatory care centre, were perhaps seen as inevitable developments, given increasing population demand. Changes to clinical treatments or protocols may have been perceived as more threatening or risky; their competing risks were more immediately flagged up. Whether one clinical risk was then

considered more acceptable than the other was influenced by the external environment, and by historical factors.

For example, the hospital antibiotic protocol relied heavily on gentamicin, a nephrotoxic drug; procedures to reduce or mitigate gentamicin toxicity had failed on several occasions, resulting in preventable kidney injuries. This did not prompt a meaningful reassessment of the alternatives to gentamicin, microbiologists valuing its advantages[3] more highly. The ongoing risk of further acute kidney injuries was seen as a justifiable one.

Increases in C.difficile infections, a potential adverse consequence of a move away from gentamicin, would be highly visible, clinically important, and financially penalised; they were reportable infections identified through systematic surveillance. Preventable kidney injuries were neither monitored in the same way, nor subject to equivalent regulatory or financial censure[4]. In addition, the prioritisation of infection control above preventable kidney injuries may have reflected previous successes in this area, which staff were unwilling to jeopardise - an element of temporal context:

> 'Some years ago we were one of the worst Trusts in the country for C.diff[icile] and a lot of work went into devising and implementing an antibiotic policy with narrow spectrum agents [to reduce the risks of hospital-acquired infections]. As with a lot of treatment options, it is often a matter of exchanging one patient safety risk for another... [Gentamicin] is an essential part of the antibiotic formulary, and... I believe the junior doctors need support to prescribe [it] safely...' *(Microbiologist - Fieldnotes)*

Clinicians could also be quick to judge proposed changes that they perceived to be similar to previous initiatives. Different teams recalled how 'Lean' QI interventions had been deliberately

[3] Advantages of gentamicin monotherapy, compared to other antibiotics, included limited resistance and lower risks of hospital-acquired C.difficile infection.

[4] Interestingly, acute kidney injury was soon to become a priority in its own right in the NHS, the focus of a national campaign - albeit without the punitive mechanisms or surveillance techniques used for hospital-acquired infections [see https://www.thinkkidneys.nhs.uk/aki/ (accessed 18th July 2016)].

discontinued by newly arriving organisational leaders. Pneumonia care bundles - designed and launched with a local QI collaborative - had fallen out of use with a lack of organisational support. This historical influence presented a barrier to innovation, with authoritative staff dissuading their colleagues from engaging in further change efforts.

Safety culture, then, was more a patchwork of differing team attitudes and practices than any one organisational characteristic. Safety attitudes and behaviours reflected externally imposed priorities as well as internal motivations and values. Although there is an assumption of 'a shared mindset, common mission or values espoused throughout an organisation' (Krein et al., 2010), senior managers disagreed with frontline staff on where improvement was needed. Indeed, there were so many simultaneous changes to the senior leadership team during the study period [see 7.4.1.3] that - even if an organisational culture were maintained - they had little time to absorb its values, let alone enact them.

7.4.1.5 Management practices

Organisational structures and processes were well defined. Each clinical division had a hierarchical structure with explicit levels of responsibility and accountability. There were numerous standard operating procedures, encompassing - amongst other things - handovers, ward rounds, and the use of surgical safety checklists. When senior leadership required governance changes to be disseminated through the various divisions, these changes were assigned to the general manager and clinical director in each area to oversee, which they typically did promptly.

However, other facets of operational management were less rigorously executed. The hospital intranet was difficult to navigate, with numerous links to out-of-date clinical guidelines. Poor document management was highlighted as a concern by the CQC in the hospital's 2014 inspection. In addition, safety and governance meetings - such as departmental mortality reviews to identify preventable deaths - were often poorly attended. Although the issue was taken up with repeat offenders, there appeared to be little appetite to resort to performance management or meaningful sanctions. This reluctance may have reflected the tacit understanding that these

clinicians had been asked to institute new procedures, for which they received no training nor supportive resources. Castigating them for under-performance might have been unfair - although many of their colleagues managed to do what they had been asked, despite the same constraints. This approach did work to some extent: most clinicians were ultimately cajoled into participating, or at least sending a representative in their place. Nonetheless, some clinicians vehemently resisted change; managers tended to work around this, rather than confront it robustly [see 7.4.1.2]. This undermined the policies and procedures which had been so carefully delineated.

7.4.1.6 Data and analytics

'Constructive comfort' asks whether detailed, timely data are available at all levels of the local system, as well as staff with the skills to interpret them. The hospital had a small central team of information analysts: much of their time was taken up with routine reporting, formulating patient safety scorecards and hospital mortality reports. They were also asked to provide data to support new hospital initiatives (such as the development of an ambulatory care centre), often at short notice. The data provided by this central team were typically aggregated to hospital level, whether for patient safety (e.g., venous thromboembolism, falls, and pressure ulcers) or mortality (SHMI, HSMR, and crude mortality). Information sources included the hospital's own data warehouse, and the healthcare information company CHKS, with which the hospital had a contract. Information analysts would attend the hospital's safety forum and mortality review meeting to discuss their findings with senior leaders. Aggregated hospital statistics were graphed to show trends over time, typically with a three month delay to allow for coding and the recording of relevant outcomes. Senior staff were familiar with statistical process control charts, and the concepts of common cause and special cause variation.

By contrast, more specific ward-level data were manually generated by ward teams themselves, typically on an 'audit day' taking place once each month, and then reviewed at monthly divisional meetings. Similarly, any data required for specific unit-level improvement programmes

were generated by the staff involved; there was no direct interaction with the central information team. These data were tabulated, with no graphical indication of trends over time or control limits. Most ward managers and clinicians had not been formally introduced to those concepts [see 7.4.1.2].

There was a clear distinction, therefore, between the data and analytical skills available to senior leadership, and those available to frontline staff and ward leaders.

7.4.1.7 Enabling environment

There was considerable emphasis on standardisation and central control, both from the external bodies to which hospital leaders were held accountable, and within the hospital itself. Key institutional priorities were established in negotiations with the local Clinical Commissioning Group, whose expectations appeared (at best) stringent. Visits from other arm's-length bodies, such as the Trust Development Authority, were more supportive. Nonetheless, the consequences of failure (as judged by these external bodies) were readily apparent, several regional chief executives having resigned in the wake of critical regulatory reports.

The hospital was expected to focus on patient safety issues which were largely externally mandated, such as venous thromboembolism and pressure ulcers. There was little room for the hospital to establish (or find resources for) other priorities. Whether intentionally supportive or not, external bodies made multiple demands of the hospital, be they statistical performance updates requiring validation, or a thorough regulatory inspection. There appeared to be little headspace for hospital leadership to carve out improvement priorities and plans of their own volition.

Within the hospital too, teams were expected to comply with initiatives launched by the senior leadership team. There was limited provision to support 'bottom-up' improvement ideas, through the clinical leadership programme [see 7.4.1.2]. In the absence of systematic support for staff-driven improvement, these ideas often gained little traction. Some ward leaders did experiment with changes more informally, usually within the parameters of the hospital's wider

priorities; for example, individual surgical wards developed their own routines to review pressure ulcer assessments and care plans at the beginning of a shift. Although experimentation was not deliberately encouraged *per se*, success stories were disseminated through safety and quality improvement fora. Overall, there was perhaps greater confidence that strong central messages and standardised procedures would contribute to improvement, more than delegation and ward-level empowerment. This may have reflected the senior leadership's dominant familiarity with governance structures above other ways to implement QI [see 7.4.1.2].

7.4.2 Implementation fidelity

7.4.2.1 Implementer level: briefing fidelity

As described in chapter 5, ward teams varied considerably in how often they used HEADS-UP, and how extensively [see 5.4.2]. They also differed in adapting their workflow for the intervention. Two wards incorporated HEADS-UP into an existing interdisciplinary meeting; two wards adopted a joint daily HEADS-UP briefing; and two wards duplicated briefings to accommodate attendance by multiple medical teams. Briefing attendance was dictated by existing interdisciplinary practice, with only doctors and senior nurses consistently expected to participate. Individual staff members' participation was dependent on their on-call commitments and patterns of leave.

Briefings were heavily influenced by pre-existing team dynamics. Consultants' engagement was largely consistent with their previous participation in institutional quality improvement initiatives. Some showed no inclination to be involved; others were proud of how HEADS-UP had been embedded in their ward work:

> 'Our new mantra is HEADS-UP-DATIX [incident reporting] in the same breath.'
>
> *(Consultant - Fieldnotes)*

Nurse-led briefings varied in style: some were reluctant to pursue follow-up actions, believing it unnecessary. Others consistently required their teams to act on the discussion, and were

sufficiently enthusiastic to ask for help introducing HEADS-UP to new clinical areas, beyond the study protocol. In some cases, teams would go well beyond the limits of the pro forma provided, documenting extensive discussions on numerous continuation sheets. In low fidelity briefings, by contrast, staff completed the HEADS-UP proforma summarily, or with only part of the team present, and did not explicitly share information:

> 'The ward manager waved the pro forma at doctors in a friendly way, asking "any problems you were aware of?", rather than going through each category... She subsequently completed some further details on the pro forma on her own, including patients awaiting discharge and some who had clinically deteriorated... She suggested to me that the rest of the team were already aware of these issues, or that they did not require any specific action.' *(Fieldnotes)*

Despite this variability, the quality of HEADS-UP briefings was commended by the CQC, the independent regulator of health and social care services in England, in November 2014. This was (from the study's perspective) an unplanned, impartial evaluation of HEADS-UP implementation, the CQC inspectors observing HEADS-UP briefings and describing them as 'outstanding practice', their highest accolade:

> 'Staff we spoke to were all aware of the scheme and were very positive about it... We observed [HEADS-UP] being used and staff were confidently highlighting issues.' *(Care Quality Commission final report)*

7.4.2.2 Implementer level: managerial fidelity

HEADS-UP data were increasingly incorporated into organisational risk management and governance processes. The programme was invited to contribute regularly to the hospital's mortality review group, chaired by the medical director and the chief executive, and to the division of medicine's quality and risk group, chaired by the general manager.

Some middle managers were able to use robustly-generated ward data to support ongoing plans for service development and investment, and to develop new strategies outright. This formed a virtuous circle in which these managers, seeing the benefits of the intervention, increasingly displayed their commitment to it [as hypothesised; see figure 4.5]. Others - hybrid managers as well as pure managers [see box 8.1] - did little to hold their service areas accountable for HEADS-UP performance, perhaps believing that the programme aligned poorly with their existing priorities, or would not prove directly useful for them. On occasion, HEADS-UP generated data specific enough to raise concerns, but not sufficiently detailed to allow further investigation. This was a source of some frustration for managers who were actively reviewing the data.

Discussions with senior managers revealed other conflicts and tensions. With limited time and resources, those managers wanted to validate ward staff reports more thoroughly before committing themselves to action. One senior leader was also concerned that HEADS-UP would reduce the number of formal incident reports submitted, the National Reporting and Learning System having already flagged a very low reporting rate compared to other institutions of similar size:

> 'I would like to get some feedback on [these issues] first... Currently they are only perception[s], and following investigation it may be that the risk changes, or the issue is solely around communication of processes already in place.' *(General manager - Fieldnotes)*

Board members questioned how best to detect a meaningful signal amongst the noise of staff reporting (whether through formal incident reports or HEADS-UP), and how then to prioritise the allocation of limited resources for improvement. They foresaw a challenge in maintaining and strengthening a culture of safety, if additional resources to resolve valid staff concerns were not ultimately forthcoming.

At the same time, the possible value of the data was appreciated, multiple senior managers

enquiring if the programme could be rolled out to more units simultaneously. Towards the end of the study period, board members were sufficiently invested in the programme to want to establish if and how it could be sustained [see 7.4.4.7].

7.4.2.3 Programmatic fidelity

A summary of implementation strategies, and the study's fidelity to them, is shown in table 7.2. Programmatic implementation challenges might be classed in three groups: contextual challenges, difficulties specific to the intervention, and tensions between implementation and evaluation.

7.4.2.3.1 Challenges from context First, challenges stemmed from the local context. For example, the adequacy of programme resources, building an improvement coalition, and train-the-trainers strategies were limited by the absence of an established improvement infrastructure with adequate resources, skills and stable leadership. Inadequate programme resources, and the lack of an existing framework for QI, particularly impacted on the delivery of feedback. Formal presentations were delivered to three ward teams at regularly scheduled quality meetings. Scheduling clashes, or the absence of an appropriate forum, precluded regular presentations for the remaining three ward teams.

In mitigation, staff on all study wards were sent regular summary emails, and individuals were given opportunistic feedback in person throughout the study period. Where feedback could be delivered formally, it appeared welcome, staff proactively rearranging presentations that had been postponed due to service pressures:

> 'There is a real value in having [someone] external review and feedback the HEADS-UP information.' *(Consultant - Fieldnotes)*

Where wards did not have an existing quality forum, the programme was reliant on the efforts of ward and service managers to help with ongoing training and data dissemination. The degree to which they did so varied considerably:

Table 7.2: Implementation strategies & programmatic fidelity

Implementation strategy	Enactment	Limitations
Access new funding	Matched funding from hospital and PSTRC	Funding may have been inadequate for the programme's needs, given the existing improvement infrastructure
Audit & feedback	HEADS-UP usage data provided to teams in person & by email	Timely data provision limited by manual collection & variable existing quality fora
Build a coalition	Partnerships cultivated with clinical and non-clinical staff	Staff turnover did not allow for a stable coalition
Educational outreach visits	Clinical teaching sessions, using HEADS-UP-identified concerns & incidents to promote its use	Scheduling clashes precluded delivery to all participants
Develop academic partnerships	Hospital-PSTRC collaboration in programme design & delivery	Tension between delivering the innovation & evaluating it
Facilitation	Additional support and troubleshooting for daily briefings; follow-up for the concerns raised	Limited access to team members who may have benefited most from facilitation
Identify early adopters	Canvassed willingness to introduce HEADS-UP in the early stages of implementation	Some of the planned early adopters' actual involvement was limited by their prolonged leave and ward changes
Involve executive boards	Senior executives gave their support, and reviewed HEADS-UP implementation data through existing quality & safety governance structures	Stepped implementation meant that HEADS-UP was not fully embraced as an institutional imperative until later in the study period
Promote adaptability	Elements of the briefing adapted for each clinical area	Proprietary format of the pro forma's design limited rapid iteration
Staged implementation scale up	Gradual roll-out of HEADS-UP to multiple units	Stepped wedge schedule had to adapt to wards' readiness to begin implementation
Train-the-trainer strategies	Ward managers & service managers introduced to HEADS-UP, and asked to introduce it in turn to their teams	Variable acceptance of this responsibility

Abbreviations: PSTRC - Patient Safety Translational Research Centre

> 'We were not being told of the feedback, so I didn't really get what was happening... I was just asked to fill out the form with the doctors, with the OT [occupational therapist] and the physio[therapist]; that was it. When I [moved ward]... I had the full training, it was explained to me properly, so it was different and then after a month you get your feedback...' *(Ward manager)*

7.4.2.3.2 Challenges from the intervention itself A second set of barriers to programmatic fidelity were due to features of the intervention itself. The proprietary format of the briefing pro forma restricted wards to make only intermittent changes, whereas rapid iterations would have been more helpful. Similarly, the paper data collection of the intervention made rapid data feedback more difficult.

7.4.2.3.3 Challenges from a tension between implementation and evaluation Third, the tensions between rigorous evaluation and effective implementation were manifest. The study's internal validity required limited contamination between the control and intervention groups. However, this may inadvertently have stopped wards from learning from each other, which is considered an effective way of promoting good implementation. The stepped implementation meant that HEADS-UP was not fully embraced as a departmental imperative until later in the study period, limiting executive involvement. It seemed that a critical mass of participation was required at ward level, which this gradual implementation plan delayed:

> 'There is a tension again between how useful this is to the hospital (which wants it to be on all wards immediately for data generation) *vs* the scientific [study] concept...' *(Fieldnotes)*

7.4.3 Facilitators and barriers for PCTS

Four superordinate themes emerged from the focus group discussions: (i) existing processes for identifying and managing concerns; (ii) the reception for HEADS-UP; (iii) pre-requisites and

facilitators for its meaningful use; and (iv) barriers.

7.4.3.1 Dissatisfaction with pre-existing processes helped generate an acceptable and appropriate intervention

7.4.3.1.1 Formal incident reporting system Few managers had time to coordinate meaningful incident investigation and follow-up. Managers were overwhelmed by unilluminating reports, and so prioritised incidents which had already caused significant harm. Little was done to address problems at an earlier stage, or to resolve issues for which staff had devised a work-around. Even for serious incidents, the time lag before subsequent action could be considerable. As a result, the staff submitting reports felt markedly detached from their colleagues who generated the response. Conversely, incident managers perceived an abrogation of responsibility on the part of the ward team, who themselves took little action to address what they had reported:

> '[The incident report] just goes off into cyberspace... You feel like you are doing it for somebody else, it is just figures, for things to be marked, for eventually somebody to put that all together and to come up with a solution that I won't be affected by.' *(Foundation doctor)*

> 'If it was [web] reported... People think that is [the] end of the matter, I have reported it, it is no longer my responsibility.' *(Service manager)*

7.4.3.1.2 Concurrent informal system for managing concerns Staff preferred to escalate certain concerns verbally. This was an entirely separate, informal system for managing concerns, which staff would deliberately choose to use instead of web reporting. Their choice was influenced by the nature of the incident, time pressures, and the immediate availability of someone to talk to, as well as the awareness that web reports were subject to process delays.

Multidisciplinary teams capitalised on this informal system to generate learning, using existing meetings for feedback or teaching, but only erratically. Again, this learning tended to be precipitated only by the most serious incidents, after a patient had come to harm:

> 'I would only perhaps raise concerns... or flag something up to one of my consultants, if it was of such a severity, a serious incident or it impacted seriously on patient care. I think one to one, face to face conversation you expect something back, and you will approach somebody who you felt had the power to actually do something about that... If it was something I was concerned about, you are more likely to have that informal conversation rather than expecting something from a [web report].' *(Foundation doctor)*

HEADS-UP therefore offered several advantages. It was an alternative route to identify staff concerns that were not well recorded with the formal web report system; it proposed better use of the existing informal mechanisms through which concerns were escalated; and it moved the initial nidus of conversation about action back to the team who had first reported the issue. Both frontline and managerial staff were conscious of their dissatisfaction with existing reporting mechanisms, and the intervention's explicit goal to make better use of frontline knowledge resonated with them. Clinical staff were often interested to hear more about HEADS-UP after early presentations: 'We need that here' was one such reaction.

7.4.3.2 The reception for HEADS-UP: significant reservations, mitigated by positive experiences

Despite its acceptability, HEADS-UP represented another intervention in a long line of QI initiatives. Conscious of the resource and opportunity costs, and under intense service pressures, many staff reacted cautiously or negatively to its introduction in practice, with signs of intense change fatigue:

> 'Initially it was like, "Oh God, what are we doing now?"' *(Ward manager)*

> 'How long are we doing this for?' *(Charge nurse - Fieldnotes)*

Staff were sceptical how HEADS-UP would bring about change, or the type and extent of that change. However, the wards that overcame their initial hesitation found their perceptions

of HEADS-UP altered quickly, and incorporated it into their normal workflow – even to the point where staff felt that removing it again might be difficult:

> 'Actually, when we got into it, it became routine, it took ten minutes really... To go through the form, decide what was happening, who was doing what, any other issues, how can we address this, and it is done.' *(Ward manager)*

> 'I think [our ward] was very good, because it was very formalised, everyone knew what time it was going to happen, and more or less all the time everyone was there, including physio[therapy] and OT [occupational therapy].' *(Foundation doctor)*

> 'I think you would have a hard job taking [HEADS-UP] away from some places now.' *(Foundation doctor)*

7.4.3.3 Facilitators and pre-requisites for meaningful use

Box 7.3 summarises the factors affecting HEADS-UP implementation, as identified by focus group participants. In addition, high performing units tended to have several characteristics clustered together that supported HEADS-UP:

- An existing, regularly scheduled interdisciplinary quality forum for feedback;
- The explicit senior expectation that HEADS-UP would be incorporated into daily practice;
- A history of engagement with quality improvement interventions;
- Empowered nursing staff; and
- Sub-teams that collaborated rather than working in isolation.

The facilitators and barriers will now be outlined in more detail.

Box 7.3 Facilitators and barriers to high fidelity implementation

Facilitators	Barriers
Training	Inadequate staffing
Expectation setting	Maladaptive interdisciplinary norms
Dedicated facilitation role	Service pressures
Feedback	Shift work
Organisational support	

7.4.3.3.1 Training and expectation setting Staff in low-fidelity areas perceived greater training requirements than those in high-fidelity areas, who found the briefing requirements easier to navigate. Without protected time for team training, teams that were already well placed to adopt HEADS-UP (through structures, practice or personality) largely retained that advantage through the study period:

> 'It is very self-explanatory what each part entails, so as long as someone is going through and leading the discussion through it, it works.' *(Foundation doctor)*

An agreed start time, dedicated multidisciplinary attendance, and a specific focus on accountability helped establish an appropriately formal briefing tone. This established collective responsibility and a shared purpose, mitigating any changes in leadership style from one day to the next. Where HEADS-UP was used best, it was underpinned by a senior clinician's explicit expectation that it would be done reliably. However, a lack of agreement about reporting priorities hindered the translation of briefing-identified concerns into formal incident reports.

7.4.3.3.2 Facilitation and feedback Although meaningful HEADS-UP briefings depended on local ownership, there was a need (as one risk manager put it) for a dedicated programme facilitator 'to bang the drum a bit'. This was important to maintain staff engagement with the programme, but also specifically to identify and link key points arising from HEADS-UP in

different areas. However, this may have inadvertently blurred the responsibility for addressing the safety issues that were then identified. An important component of the facilitation role ultimately involved follow-up with clinicians, managers, and executives to explore how these safety concerns could be addressed:

> 'I'd say it would have to be one person rather than a person in each area, to be able to spot those themes coming through and to be able to collate all the information... [Web reports] sometimes [go] to ten different people... and everyone is trying to do their own thing in their own department.' *(Foundation doctor)*

Feedback to participating clinical teams was widely acknowledged to be important, reinforcing clinicians' commitment to the programme. Opportunities to reflect on episodes of good performance (positive deviance) were welcomed, the teams recognising that they did not routinely do so:

> 'I think people forget they can learn equally as well from things that went well and things that haven't gone well, and we are missing potentially an opportunity to learn from... "Why did it go well? Can we replicate that and spread that?"' *(Service manager)*

Many staff members were eager for HEADS-UP to continue, but were conscious that it might lose relevance if supportive facilitation and feedback were not maintained. Senior managers were more optimistic that HEADS-UP was sufficiently embedded to continue without a dedicated lead. No further resources were obtained to support it; instead, local champions were nominated in each area. Frontline briefings were still continuing several months after the study period formally ended, albeit with concerns about reduced effectiveness in the absence of a robust mechanism for issue escalation and resolution:

> 'They're still using it religiously.' *(Consultant - Fieldnotes)*

7.4.3.3.3 Organisational support The piecemeal introduction of HEADS-UP was a pragmatic choice, recognising the evaluation benefits of a stepped wedge study and the small study team's logistical constraints. Nonetheless, fidelity might have been improved by a simultaneous roll-out to all of the intervention units, appropriately resourced, building a more collaborative participant community. This was recognised by staff as well as the research team [see 7.4.2.3.3]:

> 'In retrospect, in hindsight, I think as an organisation we could have supported its implementation more... I think if other organisations were implementing it, then they can look and learn from us, to say actually it really is of value and it is worth supporting from the very beginning, don't let it just trickle in, come in with a big bang, make it happen.' *(Service manager)*

7.4.3.4 Barriers to PCTS

7.4.3.4.1 Staffing and service pressures Consistent multidisciplinary attendance at any team meeting was a major challenge. Staffing shortages and intense service demands could also compromise meaningful interdisciplinary discussion, even after the HEADS-UP briefing had started. In some cases, however, ward nurses demanded these meetings be prioritised:

> 'You will have to put your foot down at times... The locum [doctor] has no choice but to stay with me until [HEADS-UP is] finished [laughs]. Same with my OT [occupational therapist] and physio[therapist]... We won't finish until everyone has put out their opinions. [On another ward] however, I've noticed our physio and our OT, because they are short [staffed]... will try to leave earlier... So I will say to them, "No you have to wait, I need to speak to you... because it's already flagged up that these particular patients [were] not discharged yesterday."' *(Ward manager)*

Shift patterns also played a role in determining what clinical staff wanted to focus on during their time at work. Sporadic shifts, providing less continuity of care, pushed staff to focus on

more immediately pressing patient care issues. Discussing pervasive service problems seemed less relevant during these nomadic appearances on the ward:

> 'I'd be days off and then doing nights on the weekend... The thing about [short] staffing... or if there's any delays in phlebotomy or in echo[cardiography] or any services, I don't really – I didn't really mind that, I was more in to the deterioration of the patients or why were they not discharging... So those were the things that I looked at.' *(Ward manager)*

7.4.3.4.2 Interdisciplinary norms Counterproductive interdisciplinary norms (e.g. tardy attendance at existing team meetings) challenged HEADS-UP implementation. One registrar reported being told not to attend pre-existing meetings on his ward as they were of 'no educational value'. However, a more productive interdisciplinary structure could even mitigate the difficulties of staffing shortages and high intensity workload. How existing meetings were run affected both HEADS-UP, and the workload for the remainder of the day:

> 'Nurse X's handover could take 45 mins to get through all the patients. People would get upset if she made them do HEADS-UP after that.' *(Ward manager)*

> 'So [medical staff] missed the discussion on a lot of the patients... and then that causes just a knock-on effect. I don't think they value the [existing] meeting and then ironically they'll be the first to complain that they weren't told about something.' *(Risk manager)*

7.4.4 Other implementation outcomes

Thematic strands from these analyses can be drawn together, to describe each of the implementation outcomes defined in table 7.1.

7.4.4.1 Acceptability

Managers and clinicians were dissatisfied with the effectiveness of the formal and informal systems for managing concerns. HEADS-UP offered improvements in information recording and follow-up, making better use of existing mechanisms for escalating concerns.

7.4.4.2 Adoption

Managers and senior clinical leaders were keen to employ HEADS-UP. Frontline staff were interested in the potential benefits, but expressed intense change fatigue.

7.4.4.3 Appropriateness

The methodology and goals of the HEADS-UP programme resonated with clinical staff, although some were sceptical how it would bring about change in practice.

7.4.4.4 Cost

This was not formally assessed.

7.4.4.5 Coverage

Frontline staff coverage was restricted by existing practice, which primarily limited briefing attendance to doctors and senior nurses. Not all staff received all feedback modalities, and managerial access was constrained by staff turnover and major competing priorities.

7.4.4.6 Feasibility

Staff used HEADS-UP briefings up to 100% of working days each month. Frontline teams were able to complete timely briefings, and incorporate them into normal workflow.

7.4.4.7 Sustainability

Briefings were better institutionalised than feedback and facilitation. Managers planned to maintain the programme by placing HEADS-UP briefing percentages in ward-level nursing scorecards, and establish ward champions and ward responsibility for data collection. Frontline briefings were ongoing well beyond the end of the study period, but ward staff expressed some concerns about their continued effectiveness, given the lack of an overall programme lead.

Although a HEADS-UP facilitator was generally acknowledged to be helpful, continuation of this role was made difficult by uncertainty about the post-merger organisational governance structure. Dedicated consultant time for governance and quality improvement was funded, but no candidate was immediately forthcoming. The place of HEADS-UP in the new, larger, post-merger organisation was ultimately unclear. However, the obstetrics division, which already had a clearer structure for governance and safety, did introduce their own version of HEADS-UP with consultant leadership.

7.5 Discussion

This evaluation explored the implementation of PCTS at hospital 1, where the majority of the study wards were located. Features of the local context influenced the capacity for implementation, with varying fidelity at briefing, managerial and programmatic levels. Auto-ethnographic fieldnotes and semi-structured focus groups identified specific facilitators and barriers for this novel safety strategy, and the programme's other implementation outcomes. These suggest opportunities to improve implementation in other contexts, to augment the programme's effects.

7.5.1 Summary of findings

7.5.1.1 Local context

Local enthusiasm for change was hampered by extreme service pressures, and few specific resources for improvement. At a senior level, which was in a state of flux, change management had been largely outsourced. The improvement efforts of junior staff were not systematically coordinated. Data management and analysis were concentrated at higher organisational levels.

Rather than having a single culture, teams adopted different standards and attitudes to change, dependent on the innovation itself, historical institutional challenges, and the success or failure of previous initiatives. Close oversight by hospital regulators and commissioning bodies limited the organisation's ability to establish its own priorities. This in turn was reflected in a relatively top-down management culture.

7.5.1.2 Implementation fidelity

Briefing implementation was predicated on existing interdisciplinary workflow. Consultant engagement with the briefings varied widely; similarly, there was a spectrum of nurse engagement, to which pre-existing differences in team dynamics and handover skills may have contributed. Independent regulators were impressed by the quality of the briefings they observed.

Implementation fidelity on the part of middle managers may have depended on the perceived personal relevance of HEADS-UP data. Senior managers were conscious of the need to validate concerns and prioritise the allocation of limited resources, but were eager to see HEADS-UP adopted on more units.

Eleven implementation strategies were used. Programmatic fidelity to these strategies was constrained by contextual factors, challenges inherent to the design of the intervention, and tensions between implementation and evaluation.

7.5.1.3 Facilitators and barriers for PCTS

Dissatisfaction with pre-existing processes helped staff to perceive HEADS-UP as an acceptable and appropriate intervention. Due to an intense change fatigue, it was still received warily, but wards that overcame their initial hesitation then embraced it in their normal workflow. These wards tended to have strong senior backing for HEADS-UP, alongside an existing interdisciplinary quality forum, a history of engagement with QI interventions, empowered nursing staff, and sub-teams who worked together rather than as entirely separate entities.

Facilitation, feedback, and appropriate organisational support were highlighted as important for the programme's success. Specific training, and establishing expectations for the briefing, were suggested to improve fidelity. Excessive clinical workload, counterproductive interdisciplinary norms, staffing issues, and inconsistent shift patterns all jeopardised potentially useful briefings.

7.5.1.4 Other implementation outcomes

PCTS was a largely acceptable and appropriate intervention in this setting, which staff were willing to adopt - although frontline staff were initially more sceptical than their managers. Briefings proved to be feasible additions to existing interdisciplinary workflow.

However, the intervention's coverage and sustainability were more challenging. Although the incorporation of HEADS-UP into existing workflows eased its adoption, the design of those same workflows restricted briefing attendance, and the delivery of feedback. Access to managerial staff was constrained by staff turnover. Briefings were sustained at ward level, but the merger reduced managerial engagement, precluding a long-term plan to sustain the programme's facilitation. The cost-effectiveness of PCTS was not assessed.

7.5.2 Limitations

Both data sources for this evaluation were reliant on access to the staff (managers and clinicians) who implemented the HEADS-UP programme. It proved easier to gain access to those who engaged with it; the evaluation may less accurately represent the views of those who did not. Implementation outcomes may be contested: quantitative measures are poorly developed (Lewis et al., 2015), and the number of related terms attached to each WHO definition reflects the complexity of the concepts. The evaluation may therefore have been particularly prey to confirmation bias; the blurred line between researcher and study participants is a consistent challenge for this model of research (Marshall et al., 2014).

I took several steps to counter any inherent bias. Focus group moderators were briefed, but had not previously worked with the participants; transcripts were coded independently by two researchers prior to analysis; and emergent themes were grounded within the experience of the focus groups and supplemented by auto-ethnography, rather than vice-versa. Although auto-ethnography can only be 'accurate' in the sense that it is one person's document of record, I feel that it provides a meaningful perspective on the implementation experience over the study period. The richness of the data mitigates their potential flaws. Integration of research staff with clinical teams is not only unavoidable in this setting; indeed, it is vitally important, if those teams are to feel comfortable discussing failings in care (Wong et al., 2015).

Two complementary analyses (inductive and deductive) were required, as focus group discussions were not readily mapped to the *'Constructive comfort'* framework, as had been planned originally. There are several possible reasons for the apparent mismatch. *'Constructive comfort'* is an ambitious document, specifying policy-level levers for change as well as novel advice for local change management. It describes success factors 'for change at any level of the health system, but particularly locally in organisations' (Allcock et al., 2015). However, assessing these factors (e.g., organisational leadership, culture, and quality improvement capacity) still requires a broad, pan-organisational view that frontline staff may not be privy to: focus group discussions

largely centred on microsystem observations. *'Constructive Comfort'* may be more relevant to policy-level changes than local quality improvement programmes, or it may be that discussants need to include the senior organisational leaders who are better able to reflect on its components.

Another possibility is that this framework, like Greenhalgh's diffusion of innovations model, is less well suited to the trial setting than to routine changes in practice (McMullen et al., 2015). Future studies should take this into account, when selecting a theoretical basis for their empirical work. Øvretveit (2014) has previously recommended other frameworks to explain context. For complex, multiple component interventions (like PCTS), Øvretveit cited Damschroder's 'Consolidated framework for implementation research' (Damschroder et al., 2009), and the PARiHS model (Rycroft-Malone et al., 2013). An earlier version of the PARiHS model had the advantage of a validated instrument for QI context assessment in the NHS, the Context Assessment Index (CAI) (McCormack et al., 2009).

7.5.3 Implications for future work

For complex interventions like PCTS to achieve their maximum efficacy, more comprehensive implementation campaigns are typically needed. For example, in an intervention to improve handovers, Starmer et al. (2014) used a two-hour training workshop; a one-hour simulation session; a computer module for independent learning; a faculty development programme; direct-observation tools for immediate participant feedback; and campaigns for process and culture change, with dedicated promotional materials.

This evaluation found substantial opportunities to improve the fidelity, coverage and sustainability of the HEADS-UP intervention. These deficits could be addressed by broadening the intervention, to include a wider range of implementation strategies. Specifically, implementation might be more effective with additional components for team training; optimisation of workflow and briefing participation; establishing more formal expectations for the briefing and managerial follow-up; and digitisation to eliminate the manual element of data collection. An

example of an electronic application to ease HEADS-UP data collection[5] is shown in appendix E. This would allow greater focus on effective facilitation, which was flagged repeatedly as an important component of the intervention. More structured approaches could also match implementation strategies to the identified barriers and facilitators for PCTS (Powell et al., 2015a).

Implementation efforts are often overlooked when trial effects are generalised to the real world: trials typically establish interventions' efficacy, not effectiveness. The HEADS-UP trial was designed to be a more pragmatic study of PCTS effectiveness (Loudon et al., 2015). Nonetheless, this evaluation suggests how the intervention could be optimised further, without compromising its real-world utility. A further evaluation of the updated intervention would be warranted: reliance on the quantitative and qualitative results achieved so far [chapters 5 and 6] would be premature. Whether the intervention could have been better optimised prior to its assessment in any trial is discussed further in chapter 10.

7.5.4 Conclusions

The design and introduction of HEADS-UP achieved several key implementation goals. Other implementation outcomes were heavily influenced by existing interdisciplinary practices and the local context, which the implementation model had not fully addressed. Managerial commitment to the programme was an important aspect of implementer fidelity, determining which safety and quality concerns would be resolved, and how. In the next chapter, I explore how managerial engagement in QI might be improved.

[5] The app was designed with Matthew Harrison, Senior Designer at the HELIX centre, and implemented for iPad by Victor Lesk at Digital Stitch, a spin out from Imperial College London.

Part IV

Future targets for quality improvement on medical wards

Chapter 8

The model of alignment

8.1 Introduction

In chapter 7, I described how the novel safety strategy of PCTS achieved several desirable implementation outcomes, such as feasibility, appropriateness, acceptability, adoption, and - to an extent - fidelity. The programme did not produce changes in structures or processes in a linear fashion, however. Rather, there were unpredictable advances, particularly when clinicians' and managers' interests directly aligned.

The degree to which managerial engagement in QI affects programme outcomes has not been widely explored. More often, the difficulties of scaling up QI programmes are attributed to problems with clinical engagement, or the broader context in which the programme takes place. Neither clinical engagement nor context is well defined (Taitz et al., 2012; Nilsen, 2015), yet the two often become de facto explanations for QI failure. Moreover, the numerous complex models examining context and QI might overwhelm researchers and clinicians [see 7.1.2]: there is limited 'how-to' support for those actually implementing change (Dixon-Woods et al., 2011; Taylor et al., 2011; Kaplan et al., 2012; Dixon-Woods et al., 2013; Harvey et al., 2015). Here, I

The work described in this chapter has been published as: Pannick S, Sevdalis N, Athanasiou T. Beyond clinical engagement: a pragmatic model for quality improvement interventions, aligning clinical and managerial priorities. BMJ Qual Saf 2016; 25:716-25. DOI: 10.1136/bmjqs-2015-004453

discuss whether the dominant focus on clinical staff has been misplaced, and propose a novel, pragmatic model for the development and selection of effective, durable QI interventions.

8.2 Aims

The aims of this chapter are:

(i) To explore definitions of clinical engagement;

(ii) To discuss specific challenges of clinical engagement in QI;

(iii) To describe the role of managerial staff in QI implementation; and

(iv) To generate a pragmatic model of aligned clinical and managerial incentives in the pursuit of QI.

8.3 Methods

This chapter draws on a selective narrative review of publications describing clinical and managerial engagement in hospital improvement efforts. The review was not intended to be a systematic one; rather, it was guided by the previous discussions of context [see 7.1.2], and informed by the auto-ethnography and implementation evaluation of the HEADS-UP study [see 6.3.1.1 and 7.4]. Given the substantial changes in healthcare management in the United Kingdom over the past 30 years (Griffiths, 1983), the review focuses on more recent literature. This generates themes of greater relevance to modern NHS practice.

Pertinent publications were initially identified in governmental reports, health services research journals, quality and safety journals, and specialist reports from charities such as the King's Fund and the Health Foundation. Hand searches of their reference lists, and exploration of similar work by those authors, were then augmented by topic suggestions from academics in the field. A similar approach was taken in a recently published narrative review with equivalent

scope (Ocloo and Matthews, 2016). A novel model to explore clinician-manager alignment in QI was then produced, again drawing on the practical experience of implementing a novel intervention. Finally, this model was retrospectively applied to the HEADS-UP study, to test its utility.

8.4 Results

8.4.1 Engagement: a reciprocal commitment from staff and their organisation

There is no universal definition of engagement: it may be an attitude, behaviour, an outcome — or all three (MacLeod and Clarke, 2011). Schaufeli et al. (2002) describe engagement as an employee's positive motivational state, characterised by 'vigour, dedication and absorption'. A broader, more cooperative, position is that engagement is a two-way phenomenon, with an onus on the organisation to establish conditions encouraging engagement and the opportunities for it to be manifest.

Clinical engagement, then, involves staff actively contributing 'within their normal working roles to maintaining and enhancing the performance of the organisation, which itself recognises this commitment in supporting and encouraging high quality care' (Spurgeon et al., 2011). This working definition makes it clear that real engagement is a very different entity to staff acquiescence, for which it is often confused (Clark, 2012; Milliken, 2014).

In QI, however, clinical engagement has been summarised simply as staff's 'active involvement', with no recognition of the possible dialogue between clinicians and those seeking to improve their performance (Wilkinson et al., 2011). That no organisational contribution is expected may go some way to explaining why clinical engagement has been problematic. Although there is specific literature pertaining to physician engagement, in this discussion (unless specified otherwise) I group clinical healthcare professionals together. QI interventions are typically interdisciplinary, and securing greater engagement of a single staff group is not an end

in itself, only a step towards an 'organisational culture where all staff feel valued and involved' (Clark, 2012).

8.4.2 Clinical engagement can be improved by co-design, local modification and strategic selection of QI interventions

Healthcare professionals have been reluctant to involve themselves in QI initiatives. This is especially apparent in periods of sustained organisational turbulence, but is a long-standing, multifactorial and international problem (Wilkinson et al., 2011). Doctors are disproportionately hesitant to participate in safety behaviours like incident reporting, and active resistance from senior staff remains the most common barrier to the successful implementation of interdisciplinary safety checklists (Evans et al., 2006; Farley et al., 2008; Lingard et al., 2008; Fourcade et al., 2012; O'Connor et al., 2013; Treadwell et al., 2014; Russ et al., 2015). The narrative of 'automatic' clinical resistance to new initiatives, or 'change fatigue', is seemingly widely accepted within the QI literature. Circumventing this fatigue is considered a major triumph, even fundamental to QI success. With few exceptions, local QI breakthroughs are attributed to good clinical engagement; conversely, failures are seen only through the prism of inadequate clinical buy-in (Dixon-Woods et al., 2011; Dixon-Woods et al., 2013).

Not all interventions are subject to the same clinical disengagement barrier, however: some programmes might have lower thresholds for participation (Taylor et al., 2011; Øvretveit, 2011; Øvretveit et al., 2011). Conversely, problems with the introduction of a specific improvement strategy do not necessarily indicate a wider reluctance to change practice. The intervention's characteristics, at least in part, determine its reception.

Iterative co-design of the HEADS-UP intervention with physicians, to maximise its face validity, mitigated much of their expected resistance [see 8.4.6]. The process of co-design may also (in itself) improve staff ownership of the intervention. Other strategies may also have immediate appeal: interventions that used peer facilitation, wider reporting options and feedback, or engaging whole teams to identify problems all seemingly fell on fertile ground (Weingart

et al., 2000; King et al., 2006; Schuerer et al., 2006; Evans et al., 2007; Nabors et al., 2011; Sujan et al., 2011; Lear et al., 2013; Nabors et al., 2014; Dijkema et al., 2015; Provenzano et al., 2015).

Even when QI strategies do not appeal intuitively to clinicians, generating clinical engagement need not prove an insurmountable challenge. Defining the 'soft periphery' of a QI programme —the elements that should be flexibly adapted to optimise the programme's acceptance, without invalidating the entire intervention — is key (McMullen et al., 2015). Making the effort to appropriately modify QI tools for the context in which they will be applied (e.g., creating separate versions of surgical safety checklists for different specialties) then makes those tools much more palatable for clinicians.

In fact, local adaptation is the most commonly cited factor affecting checklist implementation, more so even than resistance from specific clinicians (Russ et al., 2015). This reflects each organisation's responsibility to create the opportunities for meaningful engagement: active clinical involvement is more likely when QI tools have been purposefully tailored, and when there is protected time for training to use them (Wilkinson et al., 2011; Russ et al., 2015). In contrast, unmodified checklists are unlikely to be used as intended, nor improve patient outcomes — regardless of hospitals' reported compliance (Leape, 2014; Urbach et al., 2014).

While co-design and local modification do improve the adoption of QI interventions, the financial and opportunity costs of pre-existing efforts represent a major challenge to any new initiative. Relentless organisational change, with little sense of an overall strategic direction, also contributes to a general ennui. Clinicians, believing that each 'fad' will soon be replaced with another, feel there is 'little point in investing heavily in any one initiative' (Wilkinson et al., 2011). The strategic selection of a limited number of QI interventions, appropriate to the organisation's capacity to implement them, is therefore crucial (Burnett et al., 2010; Harvey et al., 2015).

Experts have identified 22 patient safety strategies with a sufficient evidence base to recommend their widespread adoption: organisations may choose to focus on these first, with a view to their specific local needs (Shekelle et al., 2013a; Shekelle et al., 2013b). However, if organisa-

tions are to select only the QI targets that they have the capacity and willpower to pursue, other potentially valuable initiatives need be deferred in the interests of preserving engagement and momentum. Regulatory bodies have an important role here: they should give institutions the time and space to develop these focused improvement strategies (Jones and Woodhead, 2015). It may seem odd to decry the slow pace of improvement, and yet advocate a more deliberate, institution-specific approach. With space for self-determination, however, organisations that strategically shape their QI attempts go on to see wider benefits, tackling deep cultural issues that go unaddressed with more haphazard approaches (Jones and Woodhead, 2015).

8.4.3 Beyond clinical engagement: the role of managers in QI

Improving clinical engagement is only part of the solution to ineffective QI: 'administrative engagement is equally important, or disillusionment... ensues' (Milliken, 2014). Quality of care is not a leading priority for many hospital boards, however (Jha and Epstein, 2010; Jha and Epstein, 2013; Parand et al., 2014). Although board-level attention to quality issues has been associated with clinical quality, how this commitment translates, in practice, into front-line action remains unclear (Jiang et al., 2009; Jha and Epstein, 2010; Birken et al., 2012; Jha and Epstein, 2013).

A recent survey study provided a key insight: board and middle management practices are linked, and correlate strongly with hospital performance on clinical quality metrics (Tsai et al., 2015). Certain board characteristics were specifically linked to middle management styles: board attention to quality was associated with management practices that monitored it, and the use of quality metrics at board level corresponded to good operational management and target setting. If good management is truly associated with clinical quality, the role of managers in QI deserves further attention.

Yet managerial participation in QI interventions is often assumed, rather than analysed in detail. Although senior hospital executives may participate constructively in collaborative safety programmes, active managerial involvement usually goes no further than the 'expres-

sions of support' described in many QI reports (Parand et al., 2013; Reames et al., 2015). More detailed evaluations describe difficulties recruiting executives to work with QI teams, even as part of major safety initiatives (Dixon-Woods et al., 2013). When they do engage, managers have a different outlook on quality and safety programmes to clinical staff, perceiving different components of the programme to be valuable and holding more positive views of the overall results (Parand et al., 2011). Meaningful input from managers is important for the design, monitoring and evaluation of QI interventions; simply obtaining their permission to proceed is not enough.

There is little published work on the role of managers in QI, the majority of which relates only to senior (board-level) managers, rather than the middle managers under their supervision (Birken et al., 2012; Parand et al., 2014). Importantly, most improvement initiatives fail to specify how they engage these middle managers, with whom front-line staff interact directly and regularly. Middle managers are a particularly heterogeneous group, with diverse professional backgrounds, often promoted on the basis of a technical skill set rather than any specific leadership or management ability (Federico and Bonacum, 2010; Birken et al., 2012). Many have 'hybrid' clinical and administrative duties, with an inherent tension between those roles [box 8.1] (Doherty et al., 2010; Buchanan, 2013). Their decisions are necessarily 'constrained, contested and political', but favour knowledge drawn from experience rather than research findings (Walshe and Rundall, 2001; Dopson et al., 2012; Edwards et al., 2013).

Little more is known about the cognitive biases that affect middle managers' judgements, but enthusiasm for QI is not automatic. For example, they may feel the operating costs of QI programmes are not justified by any potential future benefits (Morgan et al., 2015). There remains a pressing need for research into how healthcare managers balance their multiple fiscal, statutory and service responsibilities (Parand et al., 2014).

Box 8.1 Middle managers, hybrid managers and pure managers

Middle manager
'Any manager two levels below the CEO and one level above line workers & professionals' (Huy, 2001)

Hybrid manager
A manager whose role combines clinical & managerial responsibilities

Pure manager
A manager whose responsibilities are entirely managerial, with no clinical workload

8.4.4 Influencing middle managers to facilitate effective QI: status, incentives and resources

It appears, then, that managers contribute to organisational quality; that their active involvement in QI has been taken for granted rather than proven; and that their decision-making relating to QI is likely to be complex, with conflicting priorities that are not easily resolved. Yet middle managers, in particular, are uniquely placed to facilitate effective QI. They have the power to accelerate or impede the implementation of innovations, mediating organisational messages for front-line staff, but also upwardly influencing their seniors to draw attention to the high-level support needed for specific QI programmes (Floyd and Wooldridge, 1997; Birken et al., 2012; Birken et al., 2015). Acting as information brokers, translating organisational strategy into actionable tasks, and promoting innovative practice, middle managers can convince clinical staff to prioritise QI implementation among numerous competing demands (Birken et al., 2012).

Harnessing middle managers' ability to broker organisational and front-line attention to a QI programme may prove essential to its success: proactive commitment from middle managers does influence effective QI implementation (Birken et al., 2013). Although many advocate a clinician-led, 'bottom-up' approach to improvement (Braithwaite et al., 2009), staff-driven initiatives that do not align well with strategic priorities have only limited impact or longevity (Proudlove et al., 2008; Rycroft-Malone et al., 2013; Morgan et al., 2015). Clear tensions emerge

when QI efforts are delegated entirely to clinicians without support for their direction and goals (Morgan et al., 2015). Without more senior support, front-line staff are unable to marshal the resources required to spread change, and managers have an important role to play in navigating cross-departmental obstacles (Parand et al., 2014; Stewart et al., 2015). Managers who effectively facilitate QI, without micromanaging it, are well appreciated by front-line staff. The subsequent pace of change may be slow, but a combination of top-down and bottom-up implementation results in a lasting impact (Stewart et al., 2015).

Senior managers play a role in determining middle managers' commitment to QI. These senior managers should directly emphasise QI as an organisational priority, incentivise QI commitment in performance reviews, and — vitally — make the necessary resources available (Birken et al., 2015). In addition, encouraging middle managers to leverage the human resources and performance reviews at their own disposal also improves their commitment to QI implementation. Interestingly, a performance-related human resources management framework for clinicians has recently been described and implemented, encountering little of the expected physician resistance (Trebble et al., 2015). Transparent negotiation of QI goals at each organisational level may therefore be feasible and necessary for high quality implementation.

8.4.5 A model of alignment for successful QI

I suggest that neither clinicians nor managers can make meaningful QI progress in isolation: their collaboration is fundamental to sustainably embedding practice innovations. The choice of a QI intervention, and its implementation model, both need coordination between clinical and managerial teams. I have discussed, in the preceding sections, how each group might be motivated to take part in this process. Yet their efforts need to be aligned, if QI is to form a significant part of their workload, and not be overwhelmed by other priorities (Tsai et al., 2015).

I highlight some examples of QI programmes in which the degree of collaboration between front-line and managerial staff may have contributed to the ultimate outcome [table 8.1]. In trials reporting significant improvements, investigators ensured there was adequate managerial

participation — or took on managerial roles themselves — to complement clinical involvement (Pronovost et al., 2006; Haynes et al., 2009; Haugen et al., 2013; Haugen et al., 2015). Similarly, implementation and spread of a QI intervention in a real-world setting was best accomplished with the co-leadership of top-level administrators and front-line champions (Stewart et al., 2015). Where managerial engagement was lacking, interventions did not improve outcomes significantly, or systems defects did not prove amenable to the efforts of clinical teams alone (Bion et al., 2013; Dixon-Woods et al., 2014b; Urbach et al., 2014).

With this in mind, I propose a simplified model for successful QI interventions (figure 8.1). This model emphasises, foremost, that QI interventions aiming to change healthcare providers' practice should aim to meet the aligned needs of staff at multiple levels in the organisation. Failing to coordinate these interests renders interventions susceptible to failure, regardless of enthusiasm and engagement at the other organisational levels. In fact, the degree to which an intervention recognises, makes use of, or conflicts with existing staff priorities is fundamental to its success, and should not be considered in the accompanying implementation strategy only. This pre-emptive consideration of where an intervention is likely to garner support, and the conflicts that need to be resolved to allow wholehearted participation, reflect the 'practical wisdom' thought to be a critical element of successful QI (Dixon-Woods, 2014a).

Throughout this review, and reinforced by the practical experience in the HEADS-UP study, I identified specific facilitators that coordinate clinical, middle management and senior management participation in QI. To build a useful model, I then separated these factors into incentives (establishing each group's QI participation as a core expectation of their work) and actions (specific actions by that group that make QI implementation more effective). I also highlight important barriers to aligned QI, again identified from the narrative synthesis. For facilitators and barriers, I focused deliberately on modifiable factors, with a view to building a valid model that has immediate application in practice. Importantly, the inclusion of managers as core members of the QI team may augment what is actually 'modifiable': changes that remain frustratingly out of reach for clinical QI teams may fall within the remit of an expanded clinical-managerial

Table 8.1: Managerial collaboration in selected QI interventions

QI intervention	Implementation phase	Managerial collaboration	Outcome
Surgical safety checklist	Trial (Haynes et al., 2009; Haugen et al., 2015)	Systems changes facilitated by the local investigator - essentially fulfilling a dedicated managerial role. Hospital administration/leaders required to 'support the intervention'	Reduced in-hospital complications
	Scale up (Urbach et al., 2014)	No assessment of managerial involvement in mandatory checklist implementation. Meaningful local implementation unlikely to have taken place	No significant change in patient outcomes
Programme to reduce central line infections	Trial (Pronovost et al., 2006)	Programme targeted middle managers and senior hospital leaders as well as front-line staff. Chief executives wrote 'commitment letter' to the programme team. Nurse manager led the project locally; project team also included a hospital executive advocate	Reduced infection rates
	Scale up (Bion et al., 2013)	Chief executives agreed organisations would participate, and that a director would join the local project team. In practice, most units struggled to involve executives	No improvement compared with controls
Programme to detect & mitigate organisational weaknesses	Proof of concept (Dixon-Woods et al., 2014b)	Executive sponsor for each site team. Managerial staff less often directly involved as project team members	System defects not tractable to small clinical teams' QI methodology
Programme to improve interprofessional coordination	Scale up (Stewart et al., 2015)	Spectrum of managerial involvement. In 'bottom-up' hospitals, administrators delegated and served as resources. In 'top-down' hospitals, managers primarily drove the change effort	Co-leadership of top-level administrators and front-line champions best facilitated implementation and spread of the intervention

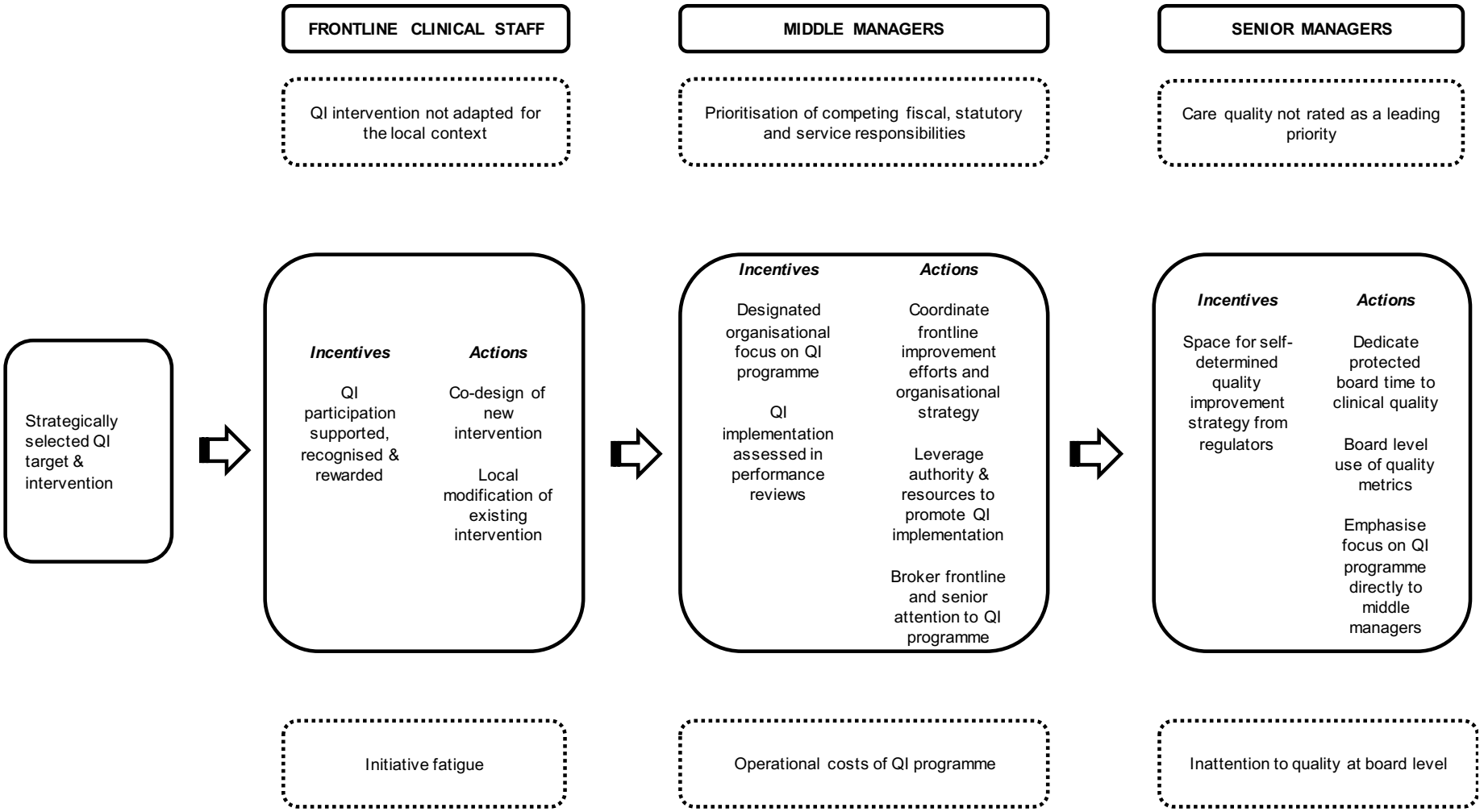


Figure 8.1: The model of alignment. Alignment of staff interests in QI is facilitated by deliberate incentives, the recognition of competing priorities and barriers to involvement [dotted boxes], and actions to address them.

group (Dixon-Woods et al., 2014b).

How could this model be used in practice? Prospectively, clinicians and managers jointly establish and prioritise the challenges facing their service. Multidisciplinary tools systematically collect data relevant to front-line care delivery problems (from staff and patients) (Sheard et al., 2014; Pannick et al., 2015). Teams then assess and rank the apparent safety threats, e.g., with streamlined versions of tools like Healthcare Failure Modes and Effects Analysis or Hierarchical Task Analysis (Dixon-Woods et al., 2014b; McElroy et al., 2016). These tools inform the strategic selection of high priority targets for local improvement efforts.

Interventions are co-designed or adapted with clinicians, with a focus on the quality strategies for which effectiveness and implementation evidence is strongest. Clinicians' participation is supported, recognised and rewarded, perhaps as part of a formal performance management process. Middle managers coordinate alignment of the improvement efforts with organisational goals, and are themselves heavily incentivised to see QI facilitation as a core role of their own. Influencing their supervisors to attract organisational support and resources for QI efforts, middle managers' interest in quality is further reinforced by protected board time for quality issues, and board-level use of quality metrics. Board members may need to robustly engage with other stakeholders in the local healthcare economy to generate (and protect) an institution-specific quality strategy.

The model can also be used retrospectively, to describe how interventions were implemented in practice, and to explain their effects. I now use the model to describe the HEADS-UP intervention in a different light, highlighting mutable influences on staff over the course of the trial. Definitive progress on a number of key issues occurred only when front-line staff and middle managers agreed the need for change.

8.4.6 Applying the model of alignment to analyse HEADS-UP implementation at the primary study site

8.4.6.1 Strategic selection of quality improvement (QI) target and intervention

Poor engagement with incident reporting was an organisational concern. However, there was little capacity to enact an improvement programme. No other goals were set aside to prioritise the HEADS-UP programme. As an experimental intervention, the efficacy of HEADS-UP and its implementation strategy was unknown.

8.4.6.2 Incentives and actions for front-line clinical staff

Front-line staff expressed initiative fatigue at the beginning of the study, but the face validity of the HEADS-UP tool (and its co-design) mitigated much of their expected resistance, and HEADS-UP was incorporated into normal workflow. QI participation was not formally rewarded, but some junior clinical staff were able to exploit their involvement in HEADS-UP to help with career progression. Others reflected that they found HEADS-UP useful for their own practice and that it improved the quality of interdisciplinary care for their patients, which may have been perceived as a benefit or reward.

8.4.6.3 Incentives and actions for middle managers

Middle managers were not personally incentivised to participate in HEADS-UP, although an endorsement by the CQC during the study period prompted senior managers to designate a greater focus on the programme. Where HEADS-UP generated information to bolster middle managers' existing business plans and develop new ones, those managers coordinated staff-identified opportunities for service development with linked organisational priorities. They gave HEADS-UP their personal backing, encouraging its use, although no additional resources were available to accompany this support. In contrast, middle managers for whom HEADS-UP was less directly useful did less to hold their service areas accountable for HEADS-UP performance.

HEADS-UP piggybacked onto existing staff meetings and governance structures: no protected time was made available for involvement in dedicated training, analysis or feedback. Middle managers' time and attention was strictly limited, and competing priorities (eg, a forthcoming merger) often precluded meaningful progress with this QI intervention.

Interestingly, resolution of the persistent issues raised in the HEADS-UP briefings appeared to depend less on each one's inherent safety threat, than on agreement between clinical staff and managers of the need for change.

8.4.6.4 Incentives and actions for senior managers

The organisation focused heavily on clinical quality, investing in external consultants to help develop new clinical services and improve the efficiency of existing ones. Board members also dedicated substantial time to clinical quality issues. In fact, clinical quality was the focus of numerous committees and subcommittees, with a complex, devolved governance structure. However, the regulatory environment did not allow for a self-determined quality strategy, with quality priorities established largely by the local healthcare commissioning body and a national quality inspectorate. Board-level meetings were awash with clinical quality metrics, among which the 'softer data' emerging from the HEADS-UP briefings had a less certain place. As the organisation moved to merge with another institution, fewer resources were available for continuous QI in the interim.

8.5 Discussion

8.5.1 Summary of findings

An emphasis on clinical engagement in QI, rather than the managerial structures needed to support it, may have limited the ability to replicate QI successes. Clinical engagement can be optimised with deliberate intervention design, but a degree of reciprocal change at the organisational level is also required - if improvement is to be sustained. Managerial participation

in QI interventions is not typically emphasised in their published reports, but there is increasing evidence of its importance. Specific incentives and actions for clinicians and managers at each organisational level may improve the implementation of strategically-selected QI interventions.

8.5.2 The novel contributions of this model

I feel the proposed model of alignment is a useful, novel concept for a number of reasons. First, it emphasises the need to go beyond clinical engagement. Second, it highlights the role of non-clinical staff in sustaining effective QI. Most importantly, it integrates separate literature streams on clinical and managerial influences on QI, lending itself to the prospective design of interventions as well as the retrospective analysis of why they achieved their goals or not. Interestingly, multilevel interventions (explicitly addressing patient, professional and organisational factors, for example) show the most consistent improvements in process and clinical outcomes (Pinnock et al., 2015). The model of alignment suggests how a similarly multifaceted approach might be incoroporated into the design of any new QI intervention. Other authors have recently raised concerns about 'colliding' QI interventions, conceived in isolation but effectively competing in a limited marketplace (Pendharkar et al., 2016). This model encourages a broader analysis of the environment into which any new QI intervention is launched.

I hope this model will prove useful for future interventions, which need to more explicitly link their assumptions with underlying theory (Davidoff et al., 2015). Critics of implementation research have argued that its theories are no more helpful than common sense (Oxman et al., 2005; Bhattacharyya et al., 2006). Yet common sense is itself an informal, 'lay' theory, relying on implicit assumptions that are difficult to challenge (Nilsen, 2015). This review offers something different: a focal point for the design and evaluation of future attempts to improve healthcare delivery. The extent to which QI aligns the interests of front-line staff and their managers has not previously been explored in this way.

8.5.3 Limitations

The model does not negate the value of existing models, which remain well placed to structure QI reporting and evaluation (Taylor et al., 2011; Kaplan et al., 2012; Harvey et al., 2015). The model of alignment was derived from a narrative review of the literature, drawing on prior systematic reviews (Wilkinson et al., 2011; Parand et al., 2014), and tested retrospectively against the experience of implementing a single intervention. It requires more empirical validation, particularly in the prospective design of novel interventions. Lastly, though parsimony is necessary for a 'good' theory of context (Davidoff et al., 2015), the model of alignment may be too simplistic for some analyses.

8.5.4 Implications for future work

This review sets out in some detail how the organisational contribution to QI, and readiness for it, are important contributors to clinicians' engagement. Organisational improvement capability therefore influences the results of QI programmes, both directly and indirectly. This was recognised by the Health Foundation, which has now issued a checklist for organisational leaders seeking to build and sustain improvements[1]. A similar emphasis on establishing the organisational structures to support QI (including aligned incentives for improved performance), and a specific organisational commitment articulating the managerial role in QI, was described.

The need to ground QI programmes in an appropriate organisational structure is therefore better established. This has implications for the introduction of complex, novel safety strategies like PCTS. In the last chapter, I suggested that a broader range of implementation strategies would improve the effectiveness of PCTS: these implementation strategies largely targeted the use of the briefings at ward level [see 7.5.3]. By contrast, the work in this chapter suggests that further improvements to PCTS would be seen with more fundamental organisational changes. A degree of organisational development is likely to be a pre-requisite for successful PCTS

[1]The checklist is available at http://www.health.org.uk/sites/default/files/Organisational%20checklist.pdf (accessed 15th June 2016).

implementation. This would represent a distinct change of approach in QI - in the UK, at least - which has been more proactive in developing stand-alone interventions than the organisations adopting them.

The next generation of guidelines for QI reporting will emphasise narrative understanding (Davies and Ogrinc, 2015), and this should extend to fuller descriptions of whether clinical and managerial priorities coincided (or collided) within the context of the intervention. Although there are few hard barriers to either group's participation in QI, competing demands force clinicians and managers to rationalise their efforts, and in some cases consciously relinquish other priorities. Developing effective and sustainable QI interventions may depend on our ability to align the two groups' divergent interests.

8.5.5 Conclusion

Harmonising managerial and clinical efforts to improve performance may be challenging. Nonetheless, this foundational alignment is potentially a robust strategy to support the sustained implementation of QI interventions. The role of patients in these ward improvement efforts is not clear. In the next chapter, I explore medical inpatients' views of the metrics commonly applied to ward performance.

Chapter 9

Patient-prioritised quality metrics for medical wards

9.1 Introduction

In the last chapter, I found that the alignment between clinical and managerial interests influenced the implementation of improvement interventions on medical wards. The two groups - clinicians and managers - may have very different priorities. Clinicians may consider themselves solely responsible for the delivery of clinical care, and its immediate outcomes; pure managers [see box 8.1] may focus on patients' timely access to a service, or its financial stability. Despite their varying incentives, both groups typically need to engage for meaningful progress on any performance measure.

One way to establish common ground for clinicians and managers is to define key outcomes which then become priority targets for collaborative improvement. Compared to more subspecialised teams, finding these key outcomes may be more difficult for medical wards. With such variety in their workload, operationalising quality is challenging [see 1.5.3]. Moreover, the most commonly collected ward outcomes are not particularly affected by changes in the delivery of interdisciplinary team care (ITC) [see 3.5.1.2 and 3.5.1.3].

In this chapter, I ask how patients might help establish the key outcomes for medical wards. Patients can be involved in this process in two ways. First, they can help define the priorities for quality measurement, i.e., the domains that outcome measures should seek to capture. Second, patient-reported outcomes and patient-reported experiences are themselves important measures of the quality of care (Nelson et al., 2015). Here, I explore patients' ideas about care quality on the ward, and their attitudes towards the quality measures already publicly displayed there.

Most studies of how patients might use performance metrics explore their use in elective decisions, such as hypothetical choices between healthcare facilities or providers (Damman et al., 2012; Elbel et al., 2014; Byham and Madathil, 2015; Damman et al., 2015). Similarly, sponsors of public outcome reporting hope to inform consumers' choice of provider, even if consumer-directed report cards are typically ill-suited for this, and few consumers use them in this way (Sinaiko et al., 2012; Shaller et al., 2014). With few choices between acute providers, inpatients and their carers are highly unlikely to use performance measures as would typical 'consumers'. The ward setting is the context which least resembles the purchase of goods or services in a competitive marketplace: lessons from consumer engagement efforts outside of healthcare may be of little relevance. Indeed, even patients who are comfortable accessing complex information resources at home should be treated as 'situationally-impaired' in the hospital environment (Morris and Karlson, 2011). The physical debilitation of an acute illness, and the emotion accompanying new diagnoses, could further impede information processing, and intense anxiety can lead to active information avoidance. Alternatively, provision of the right information, in the right context, could potentially improve the participation of patients and carers in care, perhaps even improving their outcomes (Shaller et al., 2014).

The interview study described in this chapter is based on the research question 'What measures of medical ward performance are meaningful for patients and their carers?' I hypothesised that patients and carers would prioritise information relating to their daily care above generic metrics of previous ward performance; that the metrics which would resonate most with them would be those perceived to be 'actionable'; and that the value of ward metrics could be

improved by contextualising them with action plans and follow-up information.

9.2 Aims

The aims of this chapter are:

(i) To describe current practice in the public display of ward performance metrics in an NHS setting;

(ii) To identify the aspects of ward care that relate to patients' and carers' perceptions of a 'good ward'; and

(iii) To explore patients' and carers' perceptions of formal metrics of ward care quality.

9.3 Methods

9.3.1 Establishing current practice in the display of performance metrics

Ward information displays in the acute medical wards and geriatric wards at a tertiary hospital in London were assessed with a standardised instrument. This documented the position of ward displays; the performance measures included in them; how those measures were defined; the context provided (in terms of previous performance, a target, or an acceptable benchmark); and whether any actions were suggested for patients. Free text notes recorded any adjacent information or advice also on the display boards.

Examples were photographed with a digital camera. These images provided insights into the time and priority that the displays were afforded in practice: images are powerful conduits for the feel and texture of environments (Rose, 2014). Visual materials 'reveal what is hidden in the inner mechanisms of the ordinary', providing perspective on everday practices (Rose, 2014). This 'visual sociology', or visual research method, also allows the researcher to reflect on what they encounter in their fieldwork. In doing so, photographic meaning is constructed; one must

be aware that photos are not themselves unmediated or unbiased, but dependent on the viewer (Schwartz, 1989). As with auto-ethnography, these documents of record are not undisputed [see 7.5.2]. Rather, their value lies in triangulation with other data, in this case the objective categorisation of their contents, and interviews with patients and carers. The use of the photos in the interviews themselves is described below.

9.3.2 Semi-structured interviews

9.3.2.1 Topic guide development

Semi-structured interviews were undertaken with general medical inpatients and their carers at this tertiary hospital. The interviews were based on a topic guide, co-developed with local patient and carer representatives[1], who in turn canvassed their patient and carer networks for opinions and feedback. The topic guide [see appendix F] included demographic details, patterns of recent admissions, and questions such as 'What information would you need to decide whether a ward was good or not?' Initial drafts of the topic guide included the modified information-seeking sub-scale of the Autonomy Preference Index (API) (Ende et al., 1989; Simon et al., 2010). This scale establishes the degree to which the patient feels they should be informed about their diagnoses and treatments. The topic guide was re-evaluated after the first three interviews, resulting in the removal of the API information-seeking subscale, which could not be readily completed by participants in this context. Participants found it difficult to distinguish the intent of the API questions, which focused on their own experience and illness, from the broader theme of the interview, i.e., the general performance of the ward. In any case, inpatients have consistently been shown to have a high desire for information (Ende et al., 1989; Wilkinson et al., 2008; Chung et al., 2012).

[1]The topic guide was co-developed with Fran Husson (patient representative) and Margaret Turley (carer representative) from the NIHR Imperial PSTRC.

9.3.2.2 Photo-elicitation

During their interviews, participants were shown laminated copies of the photos described above [9.3.1]. Photographs were used to examine ward practice with the participants and delve deeper into their own views. Using photographs as interview stimuli in this way has been termed photo-elicitation (Harper, 2002). Photo-elicitation can reduce misunderstandings and power differences between researcher and participants (Robinson, 2002), increase engagement, and lessen the extent to which the interviews are governed by the participants' moods (Collier, 1957).

9.3.2.3 Inclusion and exclusion criteria

Participants (patients and carers) were aged 18 years or more, and able to provide informed consent. Patients were excluded if they were physiologically unstable, had major cognitive or communication difficulties, or did not speak English.

9.3.2.4 Data collection

Ward managers and charge nurses were informed that the study was taking place. Ward staff (doctors, nurses, physiotherapists and occupational therapists) were asked to suggest patients or carers who would be physically capable of taking part in an interview. The interviews took place at patients' bedsides, between February and May 2016, using a similar methodology to a recent study with hospitalised medical patients (Tobiano et al., 2016). Interviews were audiotaped, and then transcribed verbatim, with informed consent.

9.3.3 Data analysis

Interview transcripts were managed in NVivo (QSR International, Doncaster, Australia). Two researchers read and re-read the transcripts[2]. An inductive (theory-generating) approach was

[2]I conducted the analysis with Dr Stephanie Archer, Research Fellow.

used, as recommended by Braun and Clarke (2006), and described earlier in chapter 6. Each researcher coded the transcripts individually, generating an initial coding frame, which was then discussed and refined. The transcripts were coded again, before a group of higher order themes was proposed. A third round of analysis - individually, and then with consensus - confirmed the validity of these metathemes and the aggregation of coded transcript fragments within them.

9.3.4 Sample size

Recruitment continued until thematic saturation was achieved. Thematic saturation was operationalised here as the point in data collection and analysis when new information produces little or no change to the coding schema. Development of this coding schema was discussed at three stages, and recruitment ceased when both researchers agreed that little new information was emerging.

Relevant literature was consulted to confirm that this decision was appropriate. Although a relatively large sample was initially envisaged, thematic saturation can occur within the first twelve interviews; basic elements for metathemes can be found even within the first six interviews (Guest et al., 2006). The participant sample was necessarily somewhat heterogeneous - general medical inpatients being defined by their location rather than any specific shared characteristic - which may increase the sample size required. Nonetheless, the two researchers agreed that the dataset held sufficient 'information power', i.e., sufficient information relevant to this study and its goals (Malterud et al., 2015). The adequacy of the sample size was therefore confirmed both by repeated reassessments of the emerging data, and by guidance from previously published work.

9.3.5 Ethical approval

Ethical approval for this study was granted by the Westminster Research Ethics Committee (16/LO/0196) and the hospital's joint research compliance office (16SM3129) [see appendix F]. Participants were provided with information sheets about the study, and given the opportunity

to ask questions, before being asked to give their written consent.

9.4 Results

The results will first describe current ward practice in performance metric displays, and the interview participant demographics. The two overarching themes from the interviews will then be discussed: the key components of high quality care, and perceptions of the quality metrics on display.

9.4.1 Current practice in the display of performance metrics

9.4.1.1 Choice and location of performance metrics

Performance and quality metrics on public display included measures of hand hygiene and cleaning, hospital-acquired infections (MRSA and C.difficile), nurse staffing, pressure ulcers, falls, and patient feedback. Wards typically dispersed these measures across at least two display boards, which were usually found in different locations. Although each ward displayed all the quality measures listed, wards varied in their display choices. For example, some wards included staffing metrics on their 'patient information board' [figure 9.1]; others included patient feedback (responses to the 'Friends and Family Test'[3]) in this location, displaying staffing levels elsewhere.

A common finding was that individual display boards contained combinations of messages for patients (e.g., visiting hours, nutrition support, and NHS eligibility criteria) and staff (e.g., the results of a recent organisational engagement exercise). There was little to indicate which information was directed to each group. Possessive pronouns (our and your) and pronouns (we and you) were used interchangeably, within the same display, to refer to both patients and staff [figure 9.1]. Quality and safety boards were located at the entrance to the ward, or adjacent to the nurses' station; no performance information was visible from patient beds.

[3]The mandated 'Friends and Family Test' asks the question 'How likely are you to recommend our service to friends and family if they needed similar care or treatment?' (NHS England, 2014b).



Figure 9.1: Typical 'patient information board', including staffing metrics, messages to staff, visiting hours and patient instructions.

9.4.1.2 Representations of quality

9.4.1.2.1 Hand hygiene and cleaning Performance measures for hand hygiene and cleaning were displayed as percentages achieved in the last audit. Definitions of the numerators and denominators used to calculate the percentages were not provided.

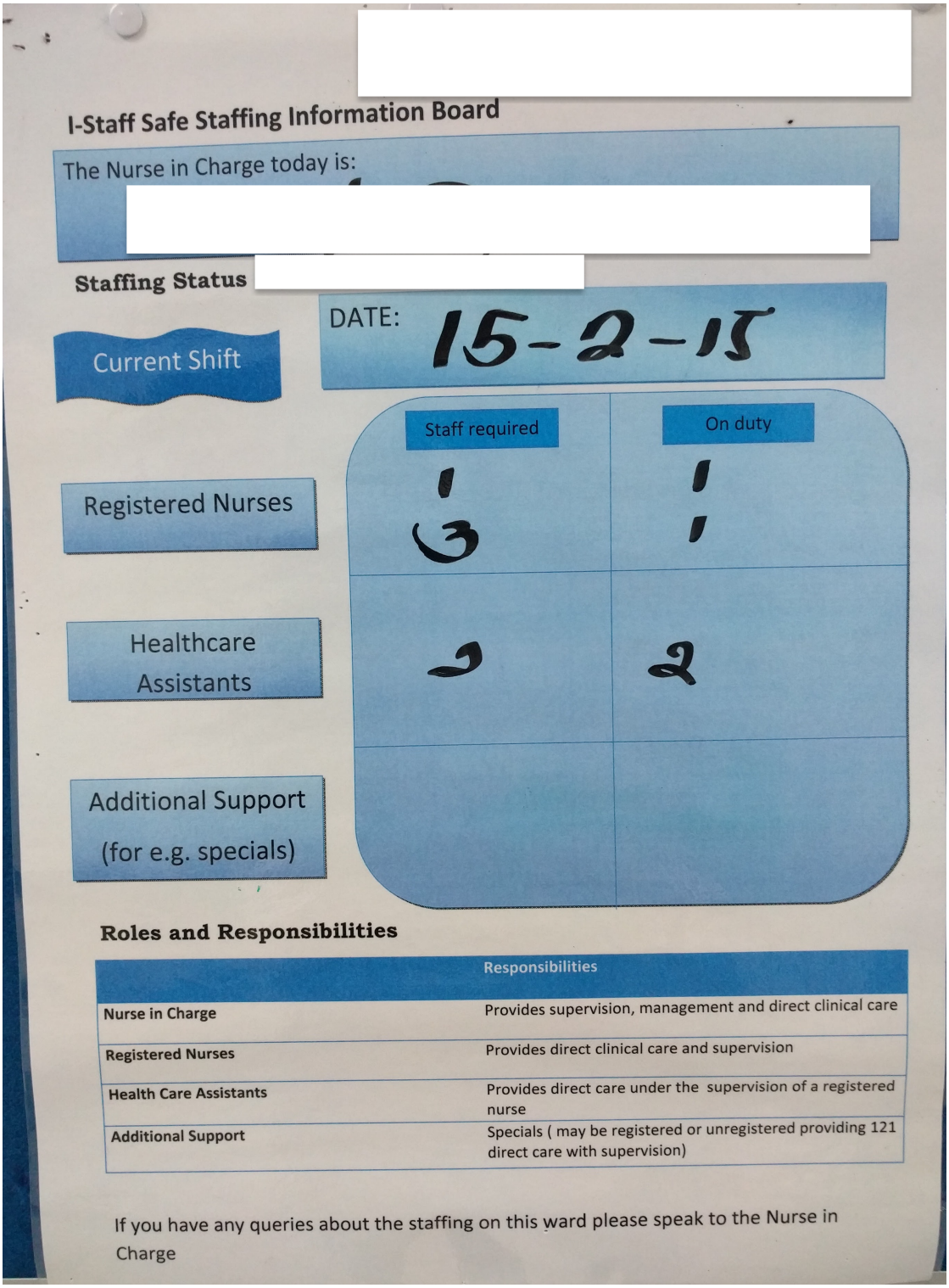
9.4.1.2.2 Staffing Nurse staffing levels for the current shift were displayed in a table, comparing - at each level of seniority - the numbers of staff required with those actually on duty. The roles and responsibilities of staff at different grades were listed [figure 9.2].

9.4.1.2.3 Hospital-acquired conditions Ward performance with respect to hospital-acquired infections, pressure ulcers and falls was displayed in each case as the date of the last recorded event [figure 9.3]. Definitions of each performance measure, or how they were assessed, were not provided.

9.4.1.2.4 Patient feedback Patient feedback was displayed as a star rating, summarising the most recent results of that ward's 'Friends and Family Test'. The star rating was accompanied by the percentage of patients who agreed they would recommend the ward to family or friends who needed similar treatment. Examples of previous patients' comments were also given. The percentage required to achieve each star rating was not provided [figure 9.4].

9.4.1.3 Contextualisation of quality data and patient engagement

All quality measures were displayed as a single, static measure of performance, i.e., the most recent result available. In no case were previous results given, that would have indicated a performance trend over time. Nor was there any indication of a performance target, or acceptable benchmark. No patient-actionable information was given for any of the performance measures, other than a suggestion to speak to a senior nurse for more information about staffing on the ward. Where performance or available resources might be taken to be sub-optimal (e.g.,



I-Staff Safe Staffing Information Board

The Nurse in Charge today is:

Staffing Status

DATE: 15-2-15

Current Shift

	Staff required	On duty
Registered Nurses	3	1
Healthcare Assistants	2	2
Additional Support (for e.g. specials)		

Roles and Responsibilities

	Responsibilities
Nurse in Charge	Provides supervision, management and direct clinical care
Registered Nurses	Provides direct clinical care and supervision
Health Care Assistants	Provides direct care under the supervision of a registered nurse
Additional Support	Specials (may be registered or unregistered providing 121 direct care with supervision)

If you have any queries about the staffing on this ward please speak to the Nurse in Charge

Figure 9.2: Shift-by-shift staffing levels, comparing numbers required to numbers present

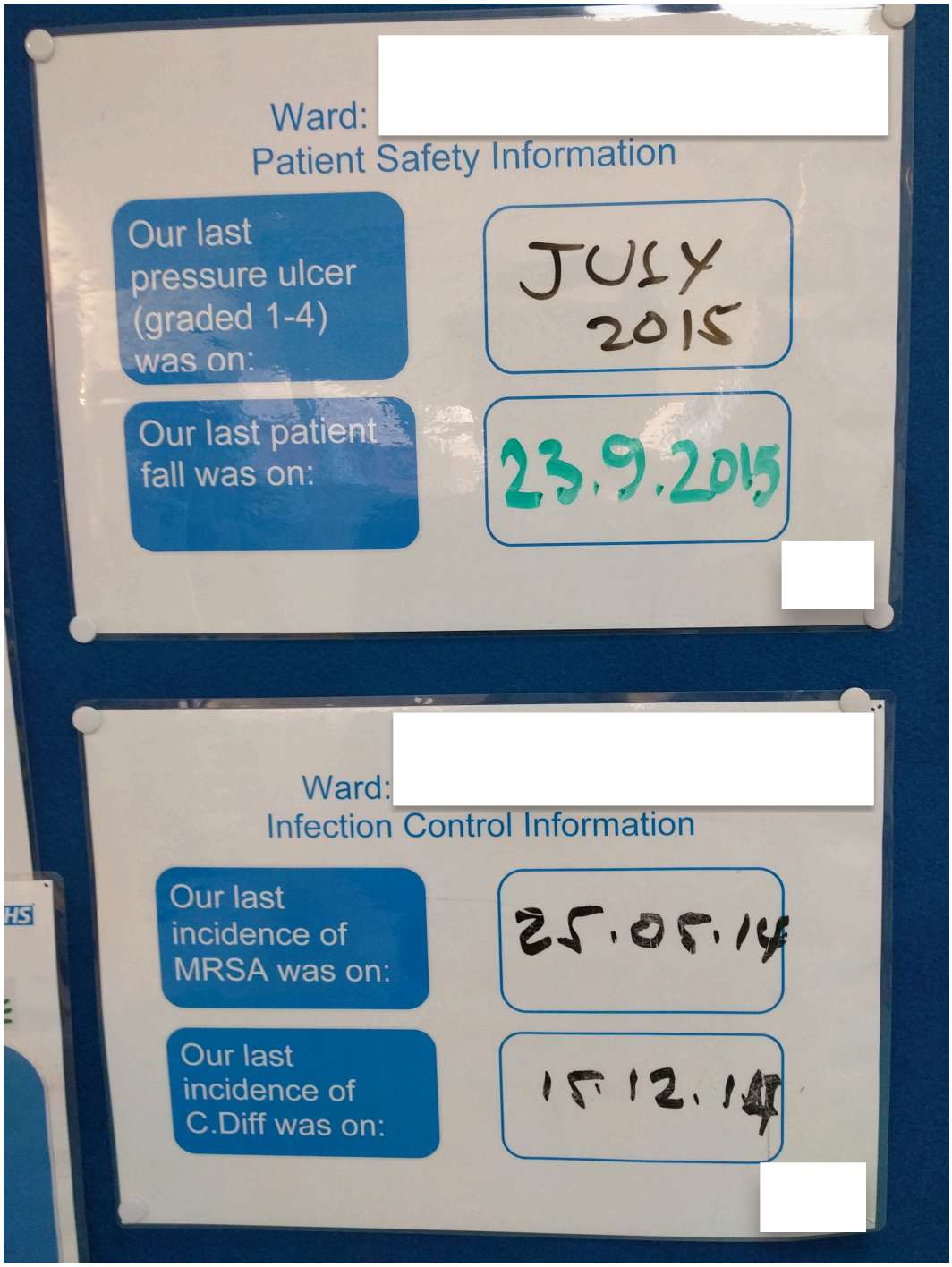


Figure 9.3: Hospital-acquired problems, displayed as the date of the last recorded event

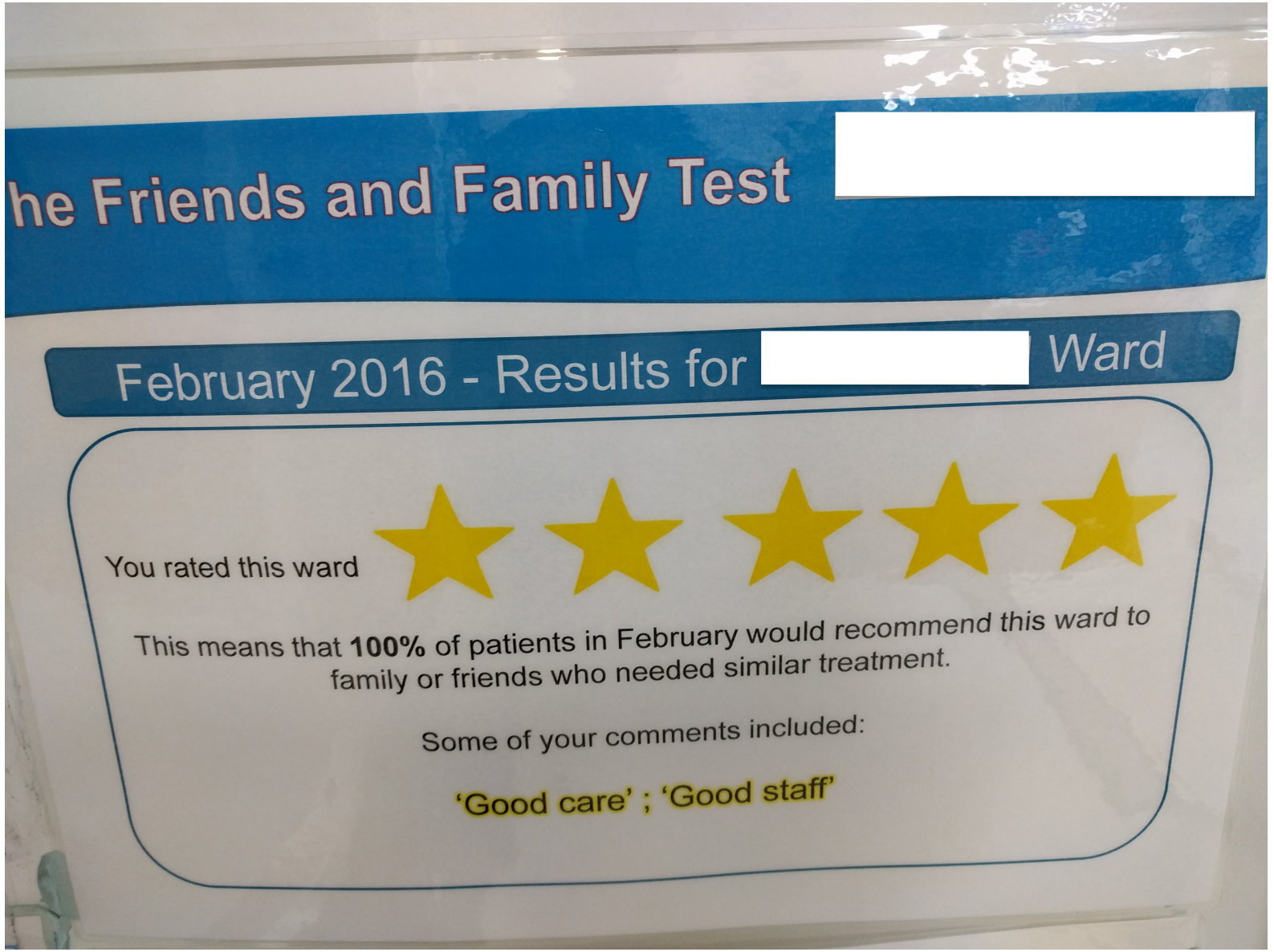


Figure 9.4: 'Friends and Family Test' results displayed as a star rating

fewer staff on shift than expected), there was no indication of a mitigation plan to reduce the likelihood of harm. Monthly quality and safety priorities for the ward could include processes that would improve performance metrics (e.g., 'MRSA compliance'). However, these priorities were not explicitly linked to previous performance, nor were they necessarily displayed near to the relevant metric.

9.4.2 Interview participant demographics

Fourteen participants were interviewed - nine patients and five carers. 50% of participants (patients and carers) were female. Nine interviews took place on the acute medical wards and five on medicine for the elderly wards. Interviews lasted a median of 23 minutes (range 11-48 minutes). Patients had a median age of 75 (range 57-86), with a median length of stay of approximately five days; their admissions were most commonly due to disorders of the cardiovascular system or abdominal complaints. 71% of participants spoke English as a first language. Half of the patients depended on family or community support, and one-third of them had undergone other hospital admissions in the preceding six months.

9.4.3 Perceptions of quality in this setting

Interview discussions centred on two overarching themes: key components of high quality ward care for patients and carers on medical wards, and their perceptions of existing quality metrics [figure 9.5]. The key components of high quality care will be discussed first: communication, staff attitudes, and hygiene.

9.4.3.1 Communication

9.4.3.1.1 Appropriate updates for patients and carers High quality communication with staff about individual treatment plans was greatly valued. Recognising that they were entirely dependent on staff to keep them abreast of forthcoming investigations, treatments and transfers, participants were keenly aware of effective communication:

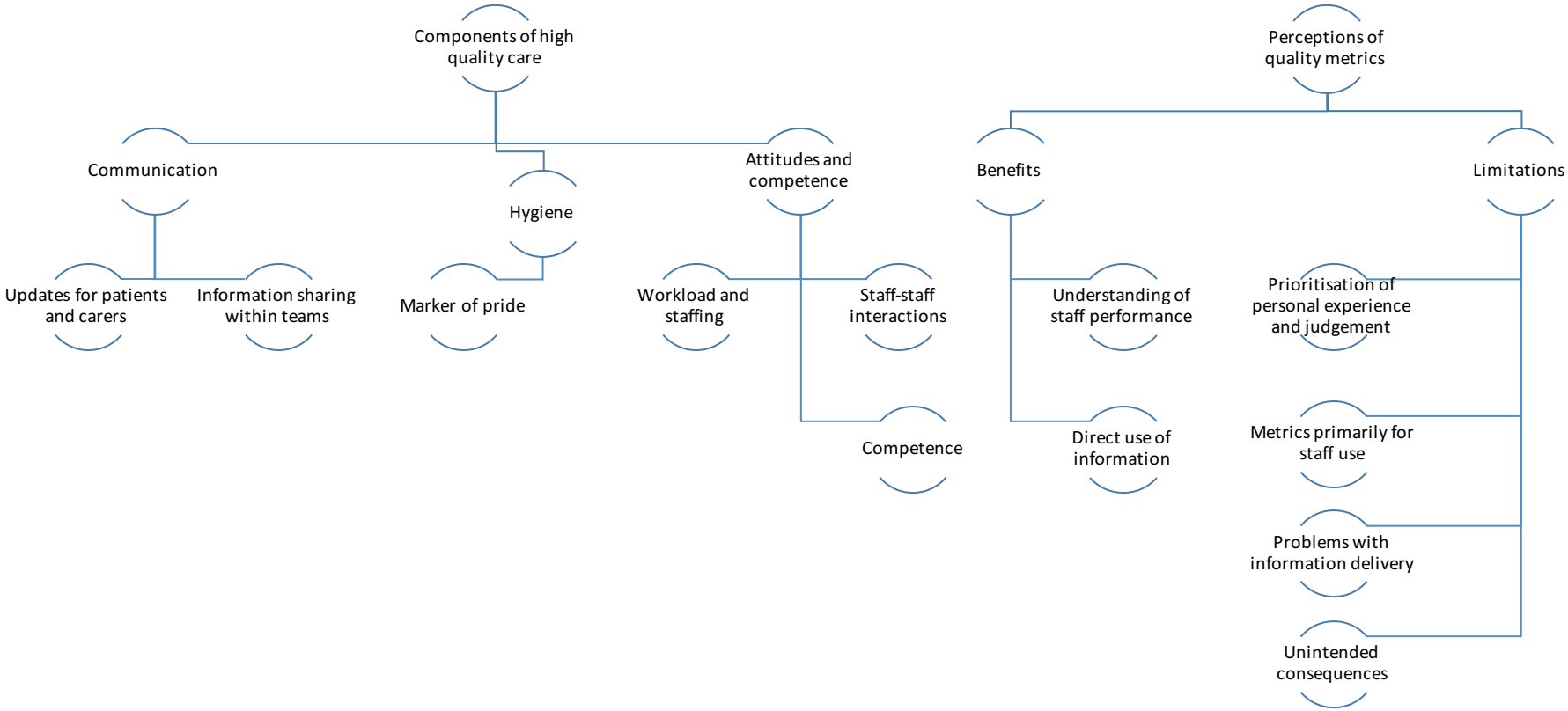


Figure 9.5: Thematic map for patient and carer interviews

The thematic map shows the two overarching themes from the patient and carer interviews: components of high quality care, and perceptions of quality metrics. These are broken down into subthemes.

> 'So what I would suggest is more attention from the side of the doctors, because if they come, they talk to each other, but they never tell you what's going to happen... I say always that [when] you see your doctor and when he comes to visit you, make it clear to you what he is telling to his colleague.' *(Patient 2)*

> '[What's important is] simply having staff that communicate with the patient... If you want an example of a good nurse, you would choose [nurse X]... If the patients interact with him, he responds immediately. It does not mean that he is running around like a fool for everyone... [If] he is asked a question, he tells you that he has recognised the fact that you have asked the question. He responds to it by saying, if he does not know the answer, then he will find out for you. And then he does what he says he is going to do, and he finds out for you... Every human being simply needs to know where the goalposts are... And be kept informed. And as long as you are [informed], then you do not have any reason to complain.' *(Patient 3)*

As treatment plans would frequently change, often for reasons outside of their teams' control, staff were merely held accountable for keeping patients updated as much as was reasonably possible. There was an expectation of an update within an appropriate timeframe, the duration of which depended on the precision of previous updates, and allowing for hospitals' unpredictable workload:

> 'I know it is not always possible that definitive information is available. But as long as you are informed to the ability that they can inform you. You cannot have any gripes about that. If someone says to you, "Look, you may go home tomorrow", I am big enough and ugly enough to know that it may be the day afterwards, or the day after that. And I can make that judgement. But if someone says to you, "We are going to take you to MRI [magnetic resonance imaging] at 2pm", I would expect someone to come within a certain timeframe of 2pm to take me to MRI. It is simple. They have informed [me] that is what is going to happen... In the same

> respect, I am not stupid enough to think - or precious enough to think - that I am the most important thing that is happening here. And if something happens in the meantime, and there is some form of emergency, and so on and so forth...' *(Patient 3)*

9.4.3.1.2 Information sharing within teams Participants also reflected on the value of effective information sharing within the multidisciplinary team. As much as it was important for patients and carers to be told about changes, so too their views and concerns needed to be shared with professionals. For them, the ability to speak to one team member, and have that conversation disseminated to the rest of the team, was a key feature of good performance:

> 'I have found you'll be speaking to one person - and it could be a nurse or a doctor or anybody else - and at the end of the day, *everybody* knows what I'm talking about... So you can communicate with [just] one person. Sometimes [in] other hospitals, you go to the doctor or the staff doesn't know what has happened to you and what has to be done, whereas here you speak to one staff and it's passed [on]... It's a vital thing.' *(Carer 1)*

> 'The communication is really good. So all of the nurses and doctors tell each other what's happened...' *(Carer 2)*

> 'This ward [team] works all together... They all seem to know what each other's doing.' *(Carer 3)*

The majority of the comments about how information was used and shared within teams came from carers, rather than patients. This perhaps reflected the role of carers in the ward environment, where they are frequently an important information source for the interdisciplinary team - for example, regarding the readiness of the home environment for the patient's return. Carer-provided information is therefore important for the overall coordination of care and treatment planning; how teams dealt with that information was clearly valued.

9.4.3.2 Staff attitudes and competence

The general approach of the ward to its patients and carers was, to some extent, manifest in its rules and how they were policed. Strict enforcement of visiting hours was not felt appropriate, or welcoming:

> 'They're much more relaxed now about visitors and visiting times... Once upon a time, it was all very strict, and that is so nice, that it's much more relaxed now.' *(Carer 5)*

Beyond those rules, however, was the much deeper impact of staff behaviours and individual interactions. Staff attentiveness, or 'service', influenced whether patients felt they were on a good ward. Considerable attention was paid to *how* staff went about their work, not just what they did. Delivering care in a positive manner distinguished high performing wards. Adjectives like 'jolly', 'respectful' and 'helpful', or 'abrupt' and 'wishy-washy', were not so much seen as individual personality attributes, as they were features of work performance:

> '[A] good ward is to be helpful to patients, being more human than a machine, you understand?' *(Patient 2)*

> 'I think it's the attitude of people [that makes a good ward]. It's the main thing.' *(Patient 6)*

> 'A good ward is... helpful staff with a smile on their face.' *(Carer 1)*

> 'If you see a gloomy face it doesn't help. Attitude is very important.' *(Carer 2)*

The manner of care delivery - rather than the resources available for it - defined the care experience to an extent, and could even influence patients' own actions:

> 'The personal factor is that you can get into a ward environment, [with] the same situation, the same support... but you found that the personal factor is not the same. There is a difference.' *(Patient 2)*

> 'Even the cleaners who come and clean around, they would say good morning to you... It makes a difference... It does mean that I'm on a ward with people that are nice, and if I did have a problem, I would be able to approach them. Because sometimes people can be very hostile and you don't want to approach them, you know... It's mainly down to the staff again, it's down to the staff.' *(Patient 4)*

However, with the potential for significant change between one shift and the next (even in the same setting), the concept of a fundamentally 'good ward' could be challenged:

> 'There's a difference from night and day. Sometimes you find the night, they are lovely people. Sometimes you find the night, they're problem people.' *(Patient 2)*

> 'Where it changes more than anything else is at night, when you have a complete change of staff. Sometimes the night staff that come on are absolutely fantastic, and are very engaged. But sometimes they are entirely the opposite. It is like, "Well we are just here to get you through until the morning, when the people that are looking after you come back."' *(Patient 3)*

With such dramatic fluctuations in the perceived quality of care, averaged measures of care appeared less relevant. This posed a challenge for ward metric displays [see 9.4.5.1.1].

9.4.3.2.1 The impact of workload and staffing levels Practices changed from one shift to another, as well as attitudes. This reflected pragmatic differences in workload and staffing levels, as much as it did a new set of staff members:

> 'They try to keep the patient as clean as possible, but this happens at night. The day people don't bother (laughs)... You see, during the day there's lots of movement, lots of traffic, so doing it in the night, everything is quiet, everything is sleepy, so they can spend the time cleaning you and looking after you a lot more...' *(Patient 2)*

There was an understanding that the quality of care received by any one patient was determined (at least in part) by the general workload on the ward. Allowances were made for this: the fact that staff might be stretched did not cause them to be judged unfavourably. Even open displays of inappropriate behaviour could be excused:

> 'Because what I have seen is everything going good, you know. Each time they're giving you everything... Only when they're very busy they're running here and there. So I don't mind.' *(Patient 1)*

> 'Sometimes I see the ward, where they're a little bit understaffed, when things get busy obviously, so there's a lot of pressure put on the staff, you know it's understandable... Obviously they can't come straightaway when you ring the buzzer... You can see that they're actually very tired people, they needed a good rest, and that's why the whole thing gets on top of them, they're overworked.' *(Patient 4)*

9.4.3.2.2 Staff-staff interactions One contrasting view was that a focus on staffing levels was in fact unwarranted, the appreciation of junior staff and healthcare assistants by other staff being a more pressing concern for high quality care:

> '[You might think] the more staff, the better the person feels, and that is *not* how I feel... Everything depends on the lower level[s] of staff we've got working in the ward... and their position [should be] respected by the doctors and the more senior people... On one or two occasions, there was somebody who I felt was not being appreciated... They did all [sorts of tasks], and nobody seemed to recognise that they were doing something like that... I would absolutely disagree that numbers are the solution. I don't believe that for a minute.' *(Patient 7)*

More generally, interactions between staff were keenly observed. These interactions directly influenced perceptions of the quality of care. Indirectly too, inferences about care quality could be drawn from how staff turned to each other for support, or the leadership on the ward:

> 'The way that the staff themselves interact with each other... You can instantly see whether they are happy working in the place where they are, and that radiates through, again, to the patients.' *(Patient 3)*

> 'If they have a good nurse in charge, who's doing a good management job of the other nurses and healthcare assistants, student nurses or whatever, then that will mean that hopefully they will function well as a team. So you've got to ensure there's a good quality nurse in charge on the day and night shifts. I know within an hour whether I've got a good nurse in charge or not. [It's] crystal clear. It's just bloody obvious.' *(Patient 8)*

> 'Two nurses were having a fight with each other, and that's not very good for the rest of us. And of the course the supervisor was asking them to be quiet, because they were shouting and screaming at each other.' *(Patient 4)*

Staff did not need to be at loggerheads for patients to be concerned about unproductive working relationships. Over-familiarity could be equally concerning, although cultural expectations of staff behaviour did vary, with patients and staff from different backgrounds:

> 'It is almost like they were a bunch of friends working together. So no one was checking on [us] - no one was trying to keep the standard raised. It was like "We can get through this doing the absolute minimum."' *(Patient 3)*

9.4.3.2.3 Staff competence Just as attitudes, workload and team ethos could vary from one shift to the next, patients perceived variations in the technical competence of staff. This was important because the idea of ward quality was very much bound up with the care that individuals had received, rather than being an abstract concept that would affect all patients equally. Personal histories of preventable adverse events, or successful treatments, influenced judgements about ward quality, recognising that staff were fallible, and that junior staff were less able to form a definitive diagnosis or treatment plan:

> 'You have doctors - sometimes they have knowledge of what they're doing and sometimes they don't. They could make mistakes. Not just the nurses, but also the doctors. I tell you, one nurse here... [describes swollen hand and thrombophlebitis resulting from a misplaced intravenous cannula].' *(Patient 2)*

> 'You realise in this situation you don't have [an experienced doctor coming to you] and looking at your situation face to face, but you have all these young students who are trying to find out what the best situation is. Eventually you get annoyed with them - because although they are doing very much, for you it's very little (laughs).' *(Patient 2)*

> '[Am I on a good ward?] Well, the medical team are doing the necessary tests... I've got confidence they've thought through the issues and are coming up with a treatment plan.' *(Patient 8)*

The prioritisation of the personal experience of care above aggregate summaries of ward performance is discussed further in 9.4.5.1.

9.4.3.3 Hygiene

Hygiene standards were held to be important for two reasons. First, good hygiene was *de facto* evidence of a ward that was providing safe care, with little unwarranted risk of infection. Second, it served as a deeper marker of staff pride, diligence and attention to detail, all of which were reassuring:

> 'I just recently came from India, and two relations were in hospital there, and there's no comparison. The hygiene - there's nothing there, so to enter a hospital is frightening... Here, there's no problem like that. You just enter happily and you know that you will be looked after, [without] infection which is quite dangerous... We're not worried that much...' *(Carer 2)*

> 'The cleanliness aspect, I think, is... more important than possibly people realise... It sets out a marker if you like... if the mindset of the ward is, you know, "We are proud of the place that we work in." So it is a fairly good marker of how that ward will actually be.' *(Patient 3)*

Nonetheless, the display of hygiene metrics within the ward was still somewhat problematic, particularly as the reported measures could clash with the personal experience of care [see 9.4.5].

9.4.4 Benefits of existing quality metrics

Potential benefits of the quality metrics on display included an increased understanding of staff performance, and opportunities for patients and carers to use the information.

9.4.4.1 Increased understanding of staff performance

Real-time information on staffing levels was felt to be helpful, in that it provided meaningful context for staff performance. Staff attitudes, and actions, would often be viewed in a more positive light when it was clear that overall staffing levels were low. Staffing data also helped set realistic expectations of the care patients might receive:

> 'When I'm getting poor service on a particular day, at least I can see that there might be a good reason for it... I would be more understanding, if I had to wait twice as long for help, if I knew that there was only half the number of staff there should be.' *(Patient 8)*

> 'When I saw the amount of staff that you're suppose to have on the ward, there were not half the staff. So the other staff that turned up were constantly busy, running back and forth, and you can see how much stress they were [under]. But they were doing a good job... You can see the nurse who has turned up is doing a really good job.' *(Carer 1)*

In addition, results indicating good performance could have a positive effect on visitors:

> 'Lots of five star results make you feel happy and secure, which is good... I certainly don't mind seeing them.' *(Carer 5)*

This latter phenomenon showed that quality measures were not only a reflection of previous performance, but could actively affect ongoing judgements about care. This is explored further in 9.4.5.4.1.

9.4.4.2 Patient and carer use of quality information

Patients and carers might also use quality data to modify their own actions, or prompt them to seek more information. This would particularly be the case if specific advice and guidance were provided alongside the quality metric. Infection control data appeared particularly resonant, participants recognising that staff, patients and carers might all help uphold hygiene standards. Feeling that they could do something helpful, participants focused on their own hand hygiene, hoping that staff would 'do the same' *(Carer 2)*:

> '[Handwashing data] should be aimed at everybody. Everybody... Obviously, yes, the staff are going to be dealing with you. Patients as well because we all touch things.' *(Patient 4)*

> '[If quality displays showed a recent infection], we'd be more careful when we came in... You would be thinking of wearing masks and gloves and things like that, because we don't want to catch it, and especially the patient could die or catch it. Because they are at their last stages. They are so weak...' *(Carer 2)*

The ability of patients and carers to engage with quality metrics was helped, in some cases, by at least a vague understanding of the medical terms used in quality displays:

> 'Because that's in the press [MRSA rates], I suppose people do want to know that, don't they? All of this you read in the papers of people being in hospital - they

went in with one thing and they came out with that... You don't want to get worse. They're meant to be making you better.' *(Patient 5)*

However, the medical terminology was not clear to all, and the potential for patients and visitors to engage with quality metrics was considerably limited by a number of factors. These will now be discussed in more detail.

9.4.5 Limitations of existing quality metrics

Patients and carers identified the following limitations to their engagement with existing quality metrics: their prioritisation of personal experience and judgement; problems with the practical delivery of the information; and unintended consequences of quality reporting.

9.4.5.1 Prioritisation of personal experience and judgement over published figures

Patients and carers, still in the process of receiving care, felt little need for measures that ultimately related to previous performance. Their ongoing experiences of care were not only adequate to form a judgement about service quality, but would override any other evaluation arising from a formal metric. In that light, performance metrics (whether objective measures, or derived from other patients' feedback), became irrelevant:

'Why would I be interested? By the time I would have got around to seeing one of those [quality metric] boards, I would probably have formed my own opinion anyway... Put it this way. Whatever those figures are, when I go into that toilet, I wash the seat before I sit down on it. I would not care if that said 1000%... We have a personal judgement of that... I would have seen [those hand hygiene figures]. But how much notice I took of them? ...I can tell you, from my experience, how many people use that [hand gel] at the end of the bed...' *(Patient 3)*

'I think you've just got to look around yourself and see what sort of service they have.' *(Carer 1)*

> 'I use my own judgement. If I'm satisfied: that's it.' *(Patient 9)*

Reported quality metrics could clash with an individual's experience of care. In addition, they were not necessarily helpful in predicting future experiences. Even near-real-time updates in staffing levels would not necessarily help patients anticipate what care they would receive. With these discrepancies in mind, there were questions about what purpose the metrics actually served:

> 'It [the staffing metric] does not mean a thing... One of the staff came in, in the evening, and said, "There is one less of us tonight." But they were better that evening than the night before, when there were more of them... The numbers really do not tell the whole story.' *(Patient 3)*

> 'It's not only in hospitals. I mean, you do a *Skype* call and you get an assessment afterwards. You walk into a place and you can press how many stars you thought it [merited]. Absolutely everyone has just gone mad. But I don't mind assessments. What I do mind: what are they for, exactly?' *(Carer 5)*

9.4.5.1.1 Inpatients have a necessarily self-centred view of care quality There was also an understanding that quality metrics, which summate overall performance, masked important variations in care. From an individual's perspective, the personal receipt of excellent care was the priority, whether or not it reflected a typical standard of care on that ward. There was therefore a conflict between the importance of an 'average' standard, and what had been received at the individual level, which may have been much better or much worse. For that reason, other patients' experiences were held as less relevant:

> 'Most people's experience in hospital is an entirely selfish or self-contained one.' *(Patient 3)*

> 'My cousin died... [because] the hospital put the food pipe in the lungs.... So the rating there [the "Friends and Family Test" result]? I can't compare that.' *(Carer 2)*

There appeared to be a distinction between what care quality meant to patients and carers, and the domains that quality metrics were purporting to measure. Indeed, some disputed that quality metrics were capable of having much meaning for patients:

> 'To actually completely quantify that [service calibre] would be very, very difficult.' *(Patient 3)*

> 'If we want information, we ask for it and we get it. As long as [my relative] is alright and getting looked after, I'm not really bothered about nothing else. If she's getting well looked after, the nurses are lovely, their care is great... that's all we are concerned about.' *(Carer 3)*

> 'As long as we are happy that she's had her one-to-one person, like the Sister, when they take over, introduce themselves to her... We don't check none of that, no... No, [it's not for relatives and patients].' *(Carer 4)*

This alluded again to the general feeling that quality measures were of limited use to patients and their carers, whose inward focus in this setting was natural.

9.4.5.2 Quality metrics for staff use, not patients'

Nonetheless, the potential value of some quality metrics for organisational use was acknowledged. Knowing that the organisation was responsibly using quality metrics for improvement was itself an important consideration:

> 'I think [patient feedback] is [important] both [for patients and hospital staff]... Because it makes me as a patient have a bit more faith in the hospital... People taking notice.' *(Patient 4)*

In practice, though, participants felt that staff had to have ownership of the data. The organisational responsibility for quality improvement could not be delegated away from staff, even to fully engaged patients and carers:

> 'One would hope the information was displayed, and if some of the scores were low... staff would recognise that to pull their socks up... If their performance isn't as good as it should be, staff have got to up their game.' *(Patient 8)*

> 'In my opinion, [the information] is for the staff. [So] that they can see how they are doing. Because this is not something that the patient is actually in control of. The hospital staff are in control of whether they have washed their hands... A patient would like to assume that that had all been happening anyway... It is not something that you... should be thinking about, "Are they washing their hands?". Because if you are thinking about that, I would not come in in the first place.' *(Patient 3)*

One patient described how he perceived some quality metrics to have little chance of generating improvement - based on the manner in which those data were collected. Seeing any sort of quality data immediately disregarded by frontline staff, as they were collected, was disheartening:

> 'The nurses don't know what to do with it. I had two [surveys] this morning, one for the food, which was low level, nobody was interested in that... The lady who did the food [survey] did it on a piece of paper, and I knew that piece of paper would be 50% likely not to get through. And then the nurse who'd seen me do it said, "That's not important." That's what I want to avoid... Then half an hour later, somebody came to do the ["Friends and Family Test"]... [They] had this iPad, and I could see she was taking it seriously doing it... There's no point in being very keen on something if nobody can administer it, and people say, "That's not important, don't worry about that." That's useless.' *(Patient 7)*

There was further frustration that the display of various quality metrics might imply that patients could act on them - choices that were not actually available. This is discussed in more detail in 9.4.5.4.2.

9.4.5.3 Problems in information delivery

9.4.5.3.1 Inadequate contextualisation The meaning of the quality metrics on display was often unclear. Without a context, definition, comparator, trend or benchmark [see 9.4.1.3], patients struggled to find the relevance of a single figure:

> 'Obviously as a member of the public I want the minimum, but I have nothing to compare it with. So if you [say], "We've not had one [infection] for three years", I can't compare that with anything. So it doesn't mean anything to me...' *(Carer 5)*

> 'That [single figure] doesn't mean anything. That doesn't inform. It could be an increase... but it could be [a] decrease.' *(Patient 2)*

> 'I am not asking for a whole year of stats. But if you are going to start giving statistics, give a reference point... Otherwise, that [figure] actually means nothing.' *(Patient 3)*

Some of the performance metrics on display also assumed a familiarity with medical terms, or the methodology used to calculate different values for performance, concepts that were quite foreign to patients and carers:

> 'That is not very clear to a layperson, what it means... I wouldn't have looked at it twice because it's not clear enough for me to understand.' *(Carer 5)*

> 'It's not very clear, that one [hygiene figure]. It doesn't give you a lot of information [about how they calculated it].' *(Patient 4)*

9.4.5.3.2 Legibility and access Font size, colour contrast, and heading clarity, could all be improved, given the patient and carer age group. Information was only available in English, limiting engagement from those speaking it as a second or third language.

Performance displays were not visible from patient bays. With patients spending the majority of time in bed or in their own bay, there were missed opportunities to help them engage with information that some would have found quite helpful:

> 'You are bound to watch them day by day or hour by hour [from your bed space], and that's the best publicity, I can assure you (laughs).' *(Patient 2)*

9.4.5.4 Unintended consequences of quality reporting

9.4.5.4.1 Quality measures were not a neutral reflection of previous care, but affected judgements about ongoing care The display (or absence) of quality data could actively influence how ongoing care was perceived, despite the preference for experiential judgements above reliance on quality metrics [see 9.4.5.1]. The absence of baseline data, trends or comparators in quality displays [see 9.4.5.3.1] raised suspicions that poor performance was being concealed, and that only positive data were aired:

> 'Let us face it... you have got your 100% figure there. Would you put up a 20% figure? ... What would you be doing? You would be ruining the confidence of the patients...' *(Patient 3)*

Seeing negative information could negatively influence perceptions of the quality of care - even if, up until that point, the patient had been quite satisfied. Interestingly, this was a concern raised by patient and carer representatives during the development of the interview topic guide. Indeed, the very effort to improve quality could portray current performance in a more negative light than was intended:

> 'To give you an analogy. When I went in 1991 to Jamaica, [I saw] a full size billboard at the side of the road saying, "Be nice to tourists!" The mere fact that the board is there (laughs)... It is a bit of an impossible situation.' *(Patient 3)*

9.4.5.4.2 Without an element of choice, the display of performance reports emphasised the lack of control in this setting Comparisons were drawn with other settings in which quality data were often provided: restaurant ratings, car purchases, and so on. In each case, the ability to act on those data seemed crucial. Although there would be a preference to be admitted to an area with good performance evaluations, in reality, patients and carers had no choice over the matter. Nor did they see other ways to act on the information provided:

> 'It would be great if I'm admitted and I'm given a choice of five wards, and I would say, "Well, how do I know which one's which, which one's best?" My next question would be, "Can you give me the audits of those wards to show which has the highest rating?" and I would go to that... If there's no choice, then it's all academic.' *(Patient 8)*

> 'Once you are here, you are here. It is not as if I could say, "Look, excuse me, you guys are crap, take me down to the Wellington [private hospital] (laughs)."' *(Patient 3)*

> 'There's not much you can do. You know there is a rating which is not really good, but what can we do? ... There's not much in your hand to do anything about it.' *(Carer 2)*

> 'Well, other than clean the wards, there's not a lot we can do is there? What else can you do?' *(Carer 4)*

The 'Friends and Family Test' question was found to be particularly challenging, given that these patients had no choice in arriving on the ward in the first place, nor could subsequent patients exercise a preference to get there:

> 'It is not as if I could say, "Okay look, you are going to be ill next week. Make sure you get into that ward." ... It does not relate. I am only in this ward because they

have not moved me to another one... To recommend someone go to a hospital, it is almost wishing them ill. I find the concept bizarre...' *(Patient 3)*

'Would I recommend a ward? How can you recommend a ward?... I mean, that's a daft question, because... they put you in the place you need to be, don't they?' *(Patient 5)*

Indeed, such were the service pressures on hospital admissions that the idea of choosing a ward seemed faintly ridiculous:

'[Recommend the ward?] They'd be lucky to get in.' *(Carer 5)*

9.5 Discussion

This study described the display of ward-based performance metrics in NHS medical wards, and the degree to which those metrics correlated with patients' and carers' perceptions of what it meant to provide high quality ward care. Patient and carer expressions of quality on the medical ward were not well aligned with existing ward metrics, which specifically assessed hygiene (and infections) but not staff communication, attitudes or competence. Whilst there may be clear practical opportunities to improve the delivery and presentation of ward metrics, there are more fundamental challenges. Patients' and carers' reliance on their individual experiences above other measures, the poor correlation between individual care quality and aggregated metrics, and the lack of opportunities to use data, all pose significant hurdles to patient and carer engagement with quality data.

9.5.1 Summary of findings

9.5.1.1 Current practice in performance metric display

Metrics for hand hygiene, cleaning, infections, pressure ulcers, falls, nurse staffing and patient feedback were all displayed in the ward environment. Whether these metrics were intended for

patients or staff was not necessarily clear. The metrics used a variety of methodologies, but each one provided only an isolated measure of performance, with no reference to previous results or an acceptable target. No patient-actionable information was provided, and ward improvement priorities were not linked clearly to performance evaluations.

9.5.1.2 Perceptions of high quality in this setting

The key elements of high quality ward care for patients and carers appeared to be:

1. Communication - between patients/carers and staff, *and* sharing information effectively within the multidisciplinary team;

2. Staff competence and staff attitudes, which were seen not so much as personal attributes as they were features of ward performance; and

3. Ward hygiene, both as an essential part of good care and as a marker of staff pride and attention to detail.

Only one of these elements (hygiene) was clearly reflected in the publicly-displayed ward metrics. Although 'Friends and Family Test' responses may reflect perceptions of staff attitudes and communication to some degree, it is too generic a question to specifically measure these priorities. With respect to these two elements of care, marked differences in performance between one shift and the next challenged the idea of a 'good ward': rather, patients and carers might describe good or bad shifts.

9.5.1.3 Patient & carer perceptions of existing metrics

Staffing metrics could help patients and carers place staff performance in context. There were some opportunities to act on quality data, particularly in terms of handwashing, in which all those present on the ward could participate. Limited quality reporting - particularly where no previous performance or trend was given - had the potential to negatively influence perceptions

of care. Quality data were not particularly accessible for patients and carers, nor was their presentation entirely comprehensible or legible for them. Quality metrics were perceived to be for staff use, patients and carers having no way to meaningfully use them, which emphasised their lack of control in this setting.

The greatest barriers to inpatient and carer use of quality metrics were conceptual ones. These groups were more likely to rely on their ongoing individual experiences to make judgements about care quality, rather than any objective or aggregate measure of previous performance. A deliberately inward view of care quality (downplaying others' experiences, and average standards, in comparison with one's own experience) also limited patient and carer engagement with these measures.

Overall, there was support for the hypotheses that personally-relevant information would be prized above averaged measures of ward quality, and that actionable and contextualised metrics would be held more valuable. However, with major barriers to patient and carer engagement with ward quality metrics, a greater focus on how staff use the data might be more appropriate. Many of the patients and carers in this study held that opinion.

9.5.2 Novel contributions of this work

To my knowledge, this is the first study to compare the display of performance metrics in the medical ward with patient and carer perceptions of high quality care. Previous studies have explored the provision of individualised information to engage inpatients in decisions about their own care (Wilkinson et al., 2008), or the use of service-level performance data to inform outpatients' selection of providers. This study combined the two concepts, describing the use of service-level data available to current inpatients and their carers. This group were neither able to exercise informed choices, nor to themselves use the data to directly improve performance - the two pathways by which Berwick et al. (2003) proposed that measurement might lead to QI. A sustained staff focus on patients' and carers' expressed priorities, as defined here, might be a more meaningful outlet for their engagement in this setting. Other authors have described

how patients may not be willing to participate *directly* in improving the safety and quality of their care, even for the elements of it that they prioritise (Berger et al., 2014).

These findings build on a body of work exploring patient priorities and involvement in the acute hospital setting. Boyd (2007) surveyed patients recently discharged from English hospitals, similarly finding that communication, patient-professional interactions, hygiene and the technical delivery of care were their main priorities. The proportion of medical patients in Boyd's sample was not reported; nonetheless, this study suggests that Boyd's findings - which excluded current inpatients - were not unduly affected by recall bias. My findings also suggest an explanation for patients' relative indifference to service-level performance and change, as described in patients' experiences of handovers between consultants (Wray et al., 2016). Efforts to engage and involve patients in service changes should respect, and even capitalise on, the self-centred lens of care quality which was identified here.

Patient and family engagement have been a focus for new healthcare models internationally (Flott et al., 2016). In the United States, for example, around half of surveyed institutions already routinely include patients and families in root cause analysis, facilitate their unrestricted access, and allow patients to directly activate a rapid response team (Herrin et al., 2016). In the UK, progress has been more limited: patient experience improvements have been seen as a nursing responsibility, with more limited involvement from medical directors and other executives (Raleigh et al., 2015). The manifest differences between patients' and carers' expressed priorities, and what was publicly measured and displayed, may have reflected this detachment (Jones et al., 2016).

9.5.3 Study limitations

With a relatively small sample from a single site, the findings of this study may not be widely generalisable. This is a common criticism of detailed interview studies, which necessarily involve deeper investigation at the expense of a larger sample. Nonetheless, there is evidence that the group of interviewees was a representative one. The patient and carer priorities elicited

here matched the aspects of care previously rated as most important (Boyd, 2007). Participants included older, frail patients, and a significant proportion did not speak English as a first language. Patients in these demographics are most at risk for experiencing adverse events, but are often excluded from safety and quality research (O'Hara and Lawton, 2016).

Hospital practice at this one study site may not have reflected standard UK practice. However, my personal working experience (of six London hospitals, not including the one assessed in this study) is that limited attention to the delivery and presentation of ward quality metrics has been the rule, rather than the exception. The inappropriate use of 'static' performance measures, showing only the last data point, is widespread (Anhøj and Hellesoe, 2016). Although specific hospitals have made well-publicised investments in public displays of ward quality metrics, this has been the result of a unique local commitment rather than any coordinated systemic effort (Brooks, 2016). Nonetheless, alignment of patient priorities and measurement efforts may well be different in those exceptional centres.

One possible limitation of this study is the uncertainty of the intended audience for the quality metrics. Arguably, the inaccessibility and misalignment of the metrics would be rendered moot, if they were in fact intended for staff (rather than patient) consumption. There are three reasons why this is not the case:

1. *Any* confusion over the intended audience for the data is likely to lessen the impact of those data. Indeed, displays of performance data should explicitly state their target audience - and their intended purpose (Association of American Medical Colleges, 2014). With staff- and patient-directed information closely intermingled, neither group was likely to derive much benefit.

2. Many of the presentation flaws highlighted by patients (lack of context, baseline data, targets or trends) would render the data equally invalid for staff consumption.

3. There is a central government mandate to display shift-by-shift ward staffing metrics in clinical areas, with the specific requirement that those metrics are accessible to patients,

families and carers (NHS England, 2014c). It is likely that this policy has spillover effects, in that staff will assume that other performance data should also be displayed to patients.

It is also possible that relevant data (better aligned with patient priorities) were collected, but not displayed on the study wards. Clearly, there are other repositories for quality metrics. However, there is little evidence that this was the case: the measures promoted at a national level are more often aggregates of the measures identified here, rather than entirely different tools. For example, the NHS 'Safety Thermometer' is a composite measure of pressure ulcers, falls, catheter-associated urine infections and venous thromboembolism (NHS Quality Observatory, 2016). In terms of patient experience data, the annual NHS inpatient survey *does* specifically ask about communication and staff attitudes - but results are aggregated at the hospital level, with no ward-level feedback (Care Quality Commission, 2016). Similarly, the American Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) patient experience survey generates hospital-level scores: attempts to derive ward-level data are specifically discouraged (Centers for Medicare and Medicaid Services, 2016[b]).

9.5.4 Implications for future work

This study has several implications for future work. Firstly, existing displays of ward quality metrics serve an uncertain purpose. Their intended audience should be made clear, and they should be optimised for that audience. Patients and carers should be involved in the design of those displays to ensure that they meet their needs. Given that these groups may not be familiar with the various aspects of care quality, providing a plain language framework to describe quality indicators is likely to be helpful (Hibbard et al., 2010).

Second, metrics for communication quality would complement existing measures, and better align patient and staff priorities. The quality of information transfer between patients and staff, and of information dissemination within interdisciplinary teams, could both be assessed formally. The results of this study suggest that these assessments might best be done on a shift-by-shift basis (rather than as a repeated point prevalence study, which is the methodology for

the NHS Safety Thermometer). Changes in care quality occur from one shift to another. Ideally, these results would be displayed in near-real time with health information technology, for maximum impact (Dowding et al., 2015). The impact on the overall burden of quality reporting should be borne in mind [see 1.5.3.4].

9.5.5 Conclusion

Current quality metrics on display in the medical ward setting (primarily relating to adverse events and complications) may not be prioritised by patients and their carers. This is an important counterpoint to the findings of the systematic review [chapter 3], where preventable complications of care were suggested as possible performance measures for the ward. Metrics for ward-level communication and attitudes would better reflect patients' and carers' dominant concerns. As quality measures proliferate, patient- and carer-prioritised metrics should inform the creation of data sets that are meaningful to all stakeholders on the ward.

Part V

Discussion

Chapter 10

Discussion

In this concluding chapter, I reiterate the aims of the thesis and how they were addressed. The key findings of the studies are summarised, before a discussion of their relative methodological strengths and weaknesses. I propose some implications of this work for clinical practice, policy and future research, and some concluding remarks on this doctoral research.

10.1 Thesis aims

The overarching goal of the thesis was to better understand the implementation of interdisciplinary care on the general medical ward, and identify effective strategies for its improvement. Part I [chapter 1] set out the arguments for why this research was warranted. The three principal aims of the thesis were:

1. To categorise and evaluate existing improvement strategies for the medical ward;
2. To design, implement and evaluate a novel interdisciplinary intervention to improve medical ward care; and
3. To identify new medical ward improvement targets from organisational and patient perspectives.

The thesis was structured to answer each of these aims in turn. In part II [chapters 2-3], a narrative review established the major categories of improvement intervention for medical ward care; a systematic review then evaluated the higher quality evidence for ITC interventions. Part III drew on this literature to design [chapter 4] and evaluate [chapters 5-7] a novel ITC intervention, with a mixed methods analysis incorporating quantitative and qualitative methods. Finally, part IV [chapters 8-9] described an organisational model of clinician-manager alignment derived from a focused literature review, and an interview study to elicit patient and carer priorities on medical wards.

10.2 Summary of findings

The key findings will be discussed as they relate to each thesis aim.

10.2.1 Aim 1: Categorise and evaluate existing improvement strategies for the medical ward

10.2.1.1 Five dominant categories of improvement intervention

The narrative review identified five common targets for medical ward care improvement interventions: staffing levels and team composition; interdisciplinary communication and collaboration; standardisation of care; early recognition and treatment of the deteriorating patient; and local safety climate [chapter 2]. Many of these interventions had good face validity, and were reported to successfully improve ward care. However, the methodology used to assess them was often inadequate.

10.2.1.2 Limited effects of interdisciplinary team care interventions

The systematic review focused on a subset of higher quality studies that described the effects of ITC interventions [chapter 3]. Most ITC interventions did not affect length of stay, readmissions or mortality rates. There was some evidence that improved ITC might reduce preventable

complications of care and adverse events. Appropriately nuanced metrics (such as *unnecessary* length of stay) might be the most appropriate outcomes for this type of intervention. The review's findings were tempered by the high risk of bias in the majority of the included studies.

10.2.2 Aim 2: Design, implement and evaluate a novel interdisciplinary intervention in this setting

10.2.2.1 Prospective clinical team surveillance - intervention design, pilot period and study power

PCTS consisted of a structured interdisciplinary briefing (HEADS-UP), identifying clinical and administrative problems, allied with cycles of facilitated ward-level and organisational feedback [chapter 4]. PCTS was piloted on a single medical ward in late 2013, with sufficient evidence that a formal trial of the intervention would be feasible. PCTS was then evaluated in a prospective, stepped wedge, cluster controlled trial. The intervention was progressively introduced to seven medical wards, from two hospitals, between December 2013 and February 2015. The study had between 75%-100% power to detect a 2%-14% absolute risk reduction in its primary outcome, eLOS.

10.2.2.2 PCTS effects on patient outcomes and processes of care

Using generalised linear mixed-effects models, high fidelity PCTS implementation significantly reduced eLOS [chapter 5]. An apparent increase in eLOS in the intention-to-treat model did not account for the significant impact of implementation fidelity, and may have reflected unmeasured confounding factors later in the study period. There was some evidence that PCTS increased the rate of escalation events (emergency calls, referrals to the ICU outreach service, and ICU transfers). PCTS had no effect on complications of care, 30-day readmissions, or a composite of in-hospital death/death or readmission within 30 days of discharge.

10.2.2.3 Mechanisms of PCTS action, qualitative impact, and effects on incident reporting

PCTS established a non-judgemental interdisciplinary forum, with a hard edge of accountability, as identified in focus groups and ethnography [chapter 6]. This reversed normalised deviance, and encouraged teams to proactively address problems. Briefings' effectiveness were hampered by a lack of continuity from one day to the next, and failures to explicitly agree a plan of action. Cost-neutral service changes, internal service investments, and training were all generated by the intervention, but not necessarily targeted to areas that had flagged concerns. Sustained PCTS implementation significantly improved safety and teamwork climate scores on the Safety Attitudes Questionnaire. In further mixed-effects models, high fidelity implementation significantly increased total incident reports, non-falls reporting and reports submitted by non-nursing staff.

10.2.2.4 PCTS implementation

PCTS implementation was analysed against the Health Foundation's *'Constructive comfort'* framework, and with inductive analysis of fieldnotes and focus groups [chapter 7]. Local enthusiasm for change was hampered by extreme service pressures, few specific resources for improvement, senior staff turnover, and concentration of improvement skills away from the frontline. Programmatic fidelity was constrained by the context, challenges inherent to the intervention itself, and tensions between implementation and evaluation. PCTS facilitators included training, expectation setting, facilitation, feedback and organisational support. Implementation barriers included inadequate staffing, service pressures, nomadic shift patterns, and maladaptive interdisciplinary norms. PCTS was a largely acceptable, appropriate and feasible intervention in this setting, although coverage and sustainability were more challenging.

10.2.3 Aim 3: Identify novel improvement targets from organisational and patient perspectives

10.2.3.1 Clinician-manager alignment

A focused narrative review identified clinician-manager alignment as a novel focus for organisational QI efforts [chapter 8]. A longstanding emphasis on clinical engagement, rather than the managerial structures needed to support it, may have limited the ability to replicate QI successes: reciprocal organisational change is also needed. A model of staff alignment at different organisational levels described specific incentives and actions at each level, to optimise the implementation of strategically-selected QI interventions.

10.2.3.2 Patient and carer perspectives on medical ward quality metrics

Metrics for hand hygiene, cleaning, infections, pressure ulcers, falls, nurse staffing, and patient feedback were on display in the ward environment [chapter 9]. Their intended audience was not clear, and there was no reference to previous results or targets. These metrics addressed one of the priorities identified in interviews with patients and carers (hygiene), but not their other concerns (high quality communication and staff attitudes). Patients and carers acknowledged some opportunities to act on quality data themselves, but there were considerable limitations with the data as currently displayed. Conceptual barriers to the use of the data by patients and carers included: (i) their reliance on their own judgements, above objective or aggregate measures; and (ii) a deliberately inward view of care quality, downplaying others' experiences in comparison with their own.

10.3 Thesis limitations: a critique of methodological choices

The limitations of each individual study were discussed in the relevant chapter. Here, I will present a broad critique of the thesis' key methodologies. I discuss the literature reviews, the

evaluation format for PCTS, and the choice of outcome measures.

10.3.1 Literature reviews: narrative, systematic, or realist?

The findings of the narrative and systematic reviews [chapters 2 and 3, respectively] were somewhat discordant. Much of the evidence identified in the narrative review suggested that medical ward ITC interventions were likely to be effective, whereas the systematic review found otherwise. There are a number of possible explanations for this discrepancy.

The systematic review restricted its analysis to higher quality study designs. Given that publication bias and inadequate study designs plague the QI literature [see 1.2.6], it is possible that ITC interventions are truly less effective than lower quality study designs would suggest. The systematic review results would therefore supersede the results of the narrative review, and those of other reviews which similarly included low quality study designs (Bhamidipati et al., 2016).

An alternative interpretation is that, given the complexity of these interventions - alterations to team composition, team practice, or both - a systematic review methodology was inappropriate. Would a 'realist' evaluation have served the review's goals better? Realist evaluations, seeking complex interactions between context, mechanisms, and outcomes, might bring more nuanced conclusions: what works, for whom, and in what settings. Editorials have commented that few ITC interventions have proposed a causal pathway (or mechanism) for their effect, that there is significant variation in the design of these interventions, and that their goals will necessarily differ with local priorities (O'Leary et al., 2016). A realist approach (identifying specifically where, and how, interdisciplinary interventions might be effective) could be appropriate.

Yet there are clear arguments that it would be a problematic to forsake quantitative evaluations in QI (Dixon-Woods, 2014a). Systematic reviews have an important role here, not least because the over-estimation of ITC interventions' benefits hampers their implementation. Setting unattainable targets, based on exaggerated estimates of effectiveness, can damage and

undermine team commitment to what are still valuable interventions (Mosher and Kaboli, 2015). Administrators may also be dissuaded from making necessary service investments, because of a mistaken belief that ITC adjustments alone may achieve their organisational goals. Whether those goals are sufficiently sensitive to changes in ward team care can now be questioned. Other authors have now taken up the call for more careful attention to outcome selection in teamwork interventions (Mosher and Kaboli, 2015; Madigosky and Schaik, 2016). These conclusions would not have been as apparent from a realist review.

10.3.2 PCTS evaluation: evaluability assessment, formative evaluation, or summative evaluation?

Interventions like PCTS are complex, both in terms of their multiple interdependent components, and their deployment in systems that adapt to context and produce non-linear responses (Shiell et al., 2008). In view of this complexity, appropriate evaluation methodologies should reflect the maturity of the intervention, and the evaluation resources available. Around 95% of complex interventions are not fully developed at the time of evaluation (Leviton, 2013). This hinders the interpretation of evaluation results; prematurely labelling interventions as ineffective can stifle innovation, and evaluation itself is not without costs. Evaluation experts therefore advise holding off from summative evaluation (evaluation of the merit of an intervention) until it is most likely to be helpful. Instead, one might first choose a formative evaluation (using evaluation to improve the intervention) or even an evaluability assessment (Leviton, 2013).

In this section, I will describe the trade-offs inherent in the PCTS evaluations [chapters 5-7]. I will discuss whether PCTS was ready to be evaluated, and whether an evaluability assessment would have been beneficial; how the PCTS evaluation ultimately contained both summative and formative components; and how the evaluation should be seen in the light of a deliberately pragmatic trial.

10.3.2.1 Evaluability assessment

Evaluability assessment is a *pre*-evaluation activity, designed to maximise the chances that any subsequent evaluation will produce useful information (Leviton et al., 2010):

> '[Evaluability assessment judges] whether the programme is ready to be managed for results, what changes are needed to do so, and whether the evaluation would contribute to improved programme performance.' (Shadish et al., 1991)

Evaluability assessment protects programmes from being evaluated against unrealistic goals, emphasising the need for a well-established, realistic programme logic, backed by sufficient resources (Leviton et al., 2010). The key judgement of an evaluability assessment is plausibility analysis: the likelihood that outcomes are achievable given the programme's timeframe, resources, activities and context. Evaluability assessment can indicate that full evaluation is inappropriate; that there is a need to adjust activities and resources (a programme development function); that formative evaluation would be merited; or suggest summative evaluation questions and methods.

Should PCTS have been subjected to an evaluability assessment rather than the more definitive quantitative and qualitative evaluations of chapters 5 and 6? There were opportunities to optimise PCTS further: a Delphi consensus process (Jones and Hunter, 1995) might have established the most important items for a PCTS briefing, before the local pilot. Facilitation expertise too could be improved: not all facilitators are equal, with a proposed facilitation pathway from novice to expert (Harvey and Kitson, 2016). Feedback did largely adhere to the recommended principles of timeliness, solution-finding, signposting to relevant resources, and non-judgemental delivery [see 6.5.1.2] (Reynolds et al., 2016). Nonetheless, with inconsistent data dissemination by ward sisters and service managers [see 7.4.2.3.1], a more reliable mechanism to deliver feedback would have been preferable. An evaluability assessment may have identified these opportunities earlier, and could have signposted how a formal stakeholder agreement on the purpose of the programme would be helpful.

However, there were good reasons too why an evaluability assessment was *not* required. The study protocol outlined a robust programme logic, supported by the experience of the pilot period. It was therefore plausible that PCTS might affect the study's outcomes - and with an auto-ethnographic study design, there was no separate need to 'scout the programme reality' to see if it matched payers' expectations (Leviton et al., 2010). A key driver for evaluability assessment is the cost-effectiveness of a full evaluation: this was not a concern here, evaluation requirements having already been established within an approved programme budget. Ultimately, evaluability assessments sift out interventions with little chance of improving outcomes, and those delivered in insufficient doses to do so. The results of chapters 5 and 6 suggest that the decisions to proceed with implementation and evaluation were - on balance - reasonable ones.

10.3.2.2 An evaluation with formative and summative components

In practice, most evaluations have both summative and formative elements, with measures of effectiveness as well as knowledge to inform future iterations of an intervention (Health Foundation, 2015). It is important to reiterate that the quantitative and qualitative evaluations here had important formative components. The results indicate the effects of *this* iteration of PCTS, with significant learning too about how it might be improved. Although a controlled trial is more often seen as a summative evaluation, HEADS-UP briefings, facilitation and feedback would warrant further development before a wider roll-out. The quantitative, qualitative and implementation evaluations showed how PCTS proved effective when implemented at high fidelity. Future PCTS development would focus on increasing the likelihood of high fidelity implementation, particularly in those teams with less favourable histories of QI engagement, or maladaptive interdisciplinary norms.

10.3.2.3 Pragmatism *vs* efficacy

The HEADS-UP study was a deliberately pragmatic one, designed to produce results that would be broadly applicable to wider practice. The published narrative review and study protocol both

highlighted the need for a rigorously-tested interdisciplinary ward intervention that hospitals could implement without major service upheaval (Pannick et al., 2014; Pannick et al., 2015). The PRECIS-2 tool helps position a trial on the spectrum from explanatory to pragmatic[1] (Loudon et al., 2015): HEADS-UP scores highly on pragmatism.

This pragmatism comes at a cost: an imperfect intervention. Greater investments in the organisational resources and expertise needed to deliver the intervention (e.g., establishing dedicated feedback sessions rather than adding to existing meetings) may have increased the effect of PCTS. Similarly, there was no attempt to assess organisational readiness for change, nor to create a more receptive environment in which QI efforts could take hold[2]. Had PCTS been evaluated in a setting better prepared for it, efficacy may have increased - but that result would be less applicable to routine practice, where implementation efforts are commonly hindered by resource constraints. Other authors have voiced concerns about the ability to replicate explanatory reports of high intensity, multifaceted team care interventions in selected centres (Colligan et al., 2015; Wong et al., 2015). This trial prioritised results that would predict real world effectiveness, above an idealised demonstration of the intervention. Any flaws in the programme's design and implementation should be reviewed in that light.

10.3.3 Study outcomes: validity and accuracy

There was a clear rationale for the selection of each of the PCTS study outcomes [see table 4.1]. However, these outcome measures were potentially constrained in their validity and accuracy.

10.3.3.1 Valid measures of interdisciplinary team care

Study outcomes had good face validity as measures of the quality of care, or as balancing measures to monitor unintended consequences of PCTS [see 4.5.4]. Few had been specifically validated as measures of ITC, however, which the intervention was designed to improve. The

[1] PRECIS-2 domains are (participant) eligibility; recruitment; setting; organisation; flexibility (delivery); flexibility (adherence); follow-up; primary outcome and primary analysis (Loudon et al., 2015).

[2] Tools to evaluate organisational context in the NHS include the validated Context Assessment Index (CAI), and the more recent Health Foundation organisational QI checklist [see 8.5.4].

systematic review [chapter 3], which ran concurrently with the HEADS-UP study, raised concerns about the limited sensitivity of some of these outcome measures to changes in ITC. The imperfect relationship between quality and quality metrics is a longstanding concern in health services research [see 1.5.3], and it was not feasible for a study of this scale to collect other patient outcome measures outside of routine datasets.

In mitigation, patient outcomes were supplemented with the detailed qualitative (survey and auto-ethnographic) analysis to provide deeper insight into the ITC changes that were taking place. The primary outcome (eLOS) also had supportive evidence for its as use as an ITC outcome, both from the systematic review and from work demonstrating the relationship between ward team behaviours and unnecessary length of stay (Leykum et al., 2015). Further research into medical ward outcomes is discussed later in the chapter [see 10.4.1.1].

10.3.3.2 Accuracy of administrative datasets

Patient outcomes and risk adjustment were established from routinely collected administrative datasets. These datasets are generated by local administrators (coders), who peruse clinical notes to identify relevant diagnostic and procedural labels. Few clinicians are regularly involved in this process, and the accuracy of the national hospital episodes statistics (HES) database has been questioned (Spencer and Davies, 2012). Again, alternative methods for the identification of patient outcomes, diagnoses and adverse events (such as medical record review) were not feasible here, given the resources available.

However, discharge coding accuracy has improved in recent years, and routinely collected data have been deemed sufficiently robust to support both research and managerial decision-making (Burns et al., 2012). Further, in establishing the local ward-level datasets, raw data underwent an additional level of verification and cleaning (beyond that normally expected prior to national database submission). I also considered enhanced measures to optimise administrative data for risk adjustment, such as a 'lookback' period to identify pre-existing diagnoses from previous admissions. These were ultimately discounted, because of the risk of introdu-

cing an unknown bias against patients with comorbidities who had not had a previous local admission (A. Bottle, Senior Lecturer in Medical Statistics, Imperial College London - personal communication, 27th May 2014).

In summary, the important methodological choices in the thesis included the use of a systematic review rather than a realist review; summative and formative evaluations instead of an evaluability assessment; and patient outcomes drawn from administrative datasets. These choices were informed by resource limitations and the prioritisation of results that would usefully inform real world practice. Whilst the strengths of alternative methodologies are acknowledged, different choices would have involved other trade-offs (Portela et al., 2015). These tensions are difficult to resolve in improvement research.

10.4 Implications for research, clinical practice and health policy

Having recognised its methodological limitations, the work described in this thesis has a number of implications for future research, clinical practice and health policy.

10.4.1 Avenues for further research in interdisciplinary ward care

There are two distinct streams of academic work which this thesis could underpin in the future. First, there are candidate patient outcomes for the measurement of interdisciplinary care quality on the medical ward. Second, a more specific model of ITC effectiveness can now be constructed for this setting.

10.4.1.1 Outcome sets for interdisciplinary ward teams

Chapter 5 reported the need for a better-defined set of quality metrics for ward-level ITC. There are strong arguments why these quality metrics should be outcomes, not process measures: process measures receive little attention from patients, do not distinguish between providers, and consume organisational resources (Porter et al., 2016). Preventable complications of care,

adverse events, and unnecessary length of stay are currently the best candidates for validation as objective ITC outcomes on medical wards. These outcomes should be complemented by a validated patient-reported measure of ITC quality, which evaluates staff attitudes and information provision - the priorities highlighted in chapter 9. Inpatients are able to flag these concerns: indeed, these are the safety domains they cite most frequently when asked (Giles et al., 2013). Structured collection and feedback of these ITC outcomes would provide ward teams with valuable, patient-centred measures for improvement. This is an important topic for further study, as previous efforts to identify teamwork-sensitive quality measures in medicine found only process measures, not outcomes (Sorbero et al., 2008).

10.4.1.2 Building a comprehensive model of interdisciplinary ward team effectiveness

ITC, as I have envisaged it, is not *only* teamwork. Systems theory suggests that team effectiveness depends on complex interactions between team inputs, team processes, and their outputs, 'teamwork' occurring in the process phase[3] (Agency for Healthcare Research and Quality, 2005). Expanding on the definition used in chapter 3 [box 3.1], a model of ITC effectiveness should therefore incorporate (i) the practices (structures) that establish the ward team and define the opportunities for its members to interact; (ii) the team processes that 'teamwork' is usually thought to entail; and (iii) the patient-facing behaviours that arise from them to affect outcomes.

Other groups have started to explore specific relationships between the ITC model's different components, without testing or validating the entire construct. For example, using organisational theory, McIntosh et al. (2014) identified various organisational predictors (structures) of staff-rated inpatient medical care coordination (team process), without analysing outcomes. Some of these organisational and unit structures have been adopted as routine practice (Kim

[3] Although there is no consensus on what constitutes 'teamwork' (Valentine et al., 2015), the term generally refers to 'behavioural processes that people use to accomplish interdependent work, and/or the affective, cognitive, and motivation states that emerge during the course of that work' (Ilgen et al., 2005). Behavioural processes include communication and coordination; examples of emergent states are mutual respect and psychological safety (Valentine et al., 2015).

et al., 2014).

The generic Integrated Health Care Team Effectiveness Model (ITEM), drawing on a separate team effectiveness literature, considered a variety of team processes in more detail (Lemieux-Charles and McGuire, 2006). Like McIntosh, Lemieux-Charles and McGuire called for further research in specific settings to validate the constructs they had described. Deneckere et al. (2012) used the ITEM model to describe multidisciplinary teamwork quality indicators, in the context of inpatient care pathways for the standardised treatment of low complexity conditions. Deneckere's group only considered measures that could be self-reported (Deneckere et al., 2011; Deneckere et al., 2012; Deneckere et al., 2013), and left patient outcomes unexplored. In contrast, many empirical tests of interdisciplinary medical ward interventions have described structures and patient outcomes, but - in the absence of a conceptual model - have neglected processes (O'Leary et al., 2016).

A full model of ITC could marry the conceptual and empirical strengths of these previous reports [figure 10.1]. Like the ITEM model, this model would distinguish between team structures such as interdisciplinary meetings, which are modifiable by unit teams, and shared organisational structures (e.g., payment incentives and information systems), which are not. These organisational structures help establish and influence team processes; the models proposed by Deneckere and McIntosh do not make this distinction. A novel addition to this model is the construct of measurable team behaviours, through which self-reported team processes and outcomes are then linked. Team behaviours like 'sensemaking' (making sense of the competing tasks required by a group of patients) correlate with ITC outcomes such as unnecessary length of stay and complications of care (Leykum et al., 2015). Incorporating team behaviours into the model in this way mitigates the limitations of relying only on *self-reported* team characteristics, with a trade-off between perceived team effectiveness and actual improvements in patient and organisational outcomes (Lemieux-Charles and McGuire, 2006). Measuring team behaviours might identify why, in some settings, there are improvements in self-reported team processes, yet no change in patient outcomes (O'Leary et al., 2015a).

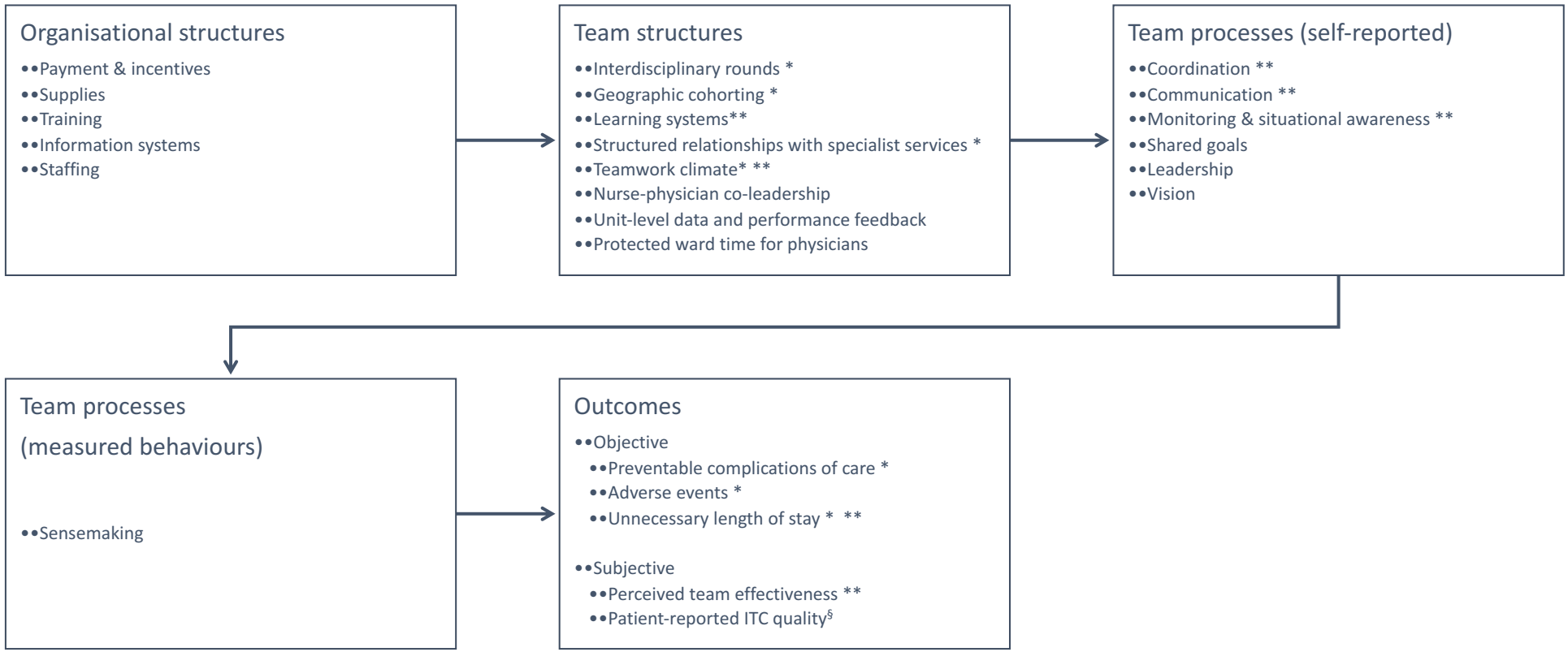


Figure 10.1: A model of interdisciplinary team care effectiveness

---

This model marries the theoretical and empirical strengths of previous work, and the understanding of ward-level QI implementation gained through this thesis. It builds on the similar, but generic, ITEM model (Lemieux-Charles and McGuire, 2006), supplemented by empirical data specific to medical ward teams. A Delphi process led by Deneckere et al. (2011) identified indicators of interdisciplinary team performance (in structure, context and self-reported outcomes), emphasising shared goals and knowledge, leadership, teamwork climate, perceived coordination of care, and vision. These were not operationalised. McIntosh et al. (2014) demonstrated a correlation between specific organisational and team factors and self-reported ward care coordination (a single team process). Further work is investigating the effects of sensemaking (Leykum et al., 2015). This thesis has focused mostly on the ward microsystem's ITC structures and specific outcomes, through the narrative and systematic reviews (*), evaluation of PCTS (**), and investigation of patient/carer priorities (§).

When should ward team behaviours be assessed? Observers can be assigned to dedicated interdisciplinary meetings, to ward rounds, or to specific ward locations (such as nursing stations) to observe work throughout a shift (Farley et al., 2010; O'Leary et al., 2012a; McAllister et al., 2014; Leykum et al., 2015). A whole-shift assessment, though arduous, has some appeal: different social rules and norms may apply during dedicated interdisciplinary meetings, which may inflate any team assessment. Alternatively, behaviour in interdisciplinary team meetings may be forced, because of the tension between the acknowledged need for communication, and a general resistance to time spent away from direct patient care (Baxter and Brumfitt, 2008; Nancarrow et al., 2012). This issue - at what points does team care matter most, and when should it be measured - is worthy of further research. The model of ITC effectiveness could then be tested more formally, subjecting its novel constructs to expert consensus, and empirically testing the links between the adjacent components.

10.4.2 Implications for clinical practice and policy

This thesis also suggests practical ways in which organisations and teams might immediately improve their medical ward care delivery. Boards, managers and ward leaders could use the intervention categories established in chapter 2 to review existing organisational safety efforts, and identify new opportunities for development. QI training for junior clinicians is already increasing (Ahmed et al., 2014; Gamble and Vaux, 2014). Local leaders will need to proactively develop frameworks to support and sustain these juniors' QI efforts, if they are to be effective. For organisations to see meaningful improvement, managerial job plans (both for pure managers and hybrid managers) should explicitly mandate and incentivise QI coordination. Without an organisational mandate and backing from each institution, the improvement efforts of junior staff are unlikely to be sustained in the workplace; many organisations have no strategy to best use this influx of QI-trained staff. There should be real concern that their efforts might go to waste.

At a ward level, there are opportunities to implement the team structures that represent cur-

rent best practice [figure 10.1] - although improvements in patient outcomes might depend very much on their local implementation. Existing interdisciplinary meetings and discharge-focused board rounds can be optimised with a more structured approach (e.g., PCTS), with organisational attention to staff support and learning. A particular challenge for the UK is securing physician engagement to drive forward ward improvement efforts, when few doctors (outside of the AAU) have unfettered commitments to these areas. To achieve real ward excellence, physicians will need to be released from their other commitments (outpatient clinics, procedures, etc.), at least for specific periods - or inpatient work will remain only a 'marginal activity' (Wachter and Goldman, 1996). Implementing protected 'ward blocks' for physicians should be a local priority.

Deliberate operational redesign should accompany these staffing changes: PCTS showed how clinical staff can identify the services and systems that no longer provide adequate support for them or their patients. 'Liberating' staff from these outdated services is an increasingly important commitment for organisations to make (Moore and Buchanan, 2013; University College London Hospitals, 2016). This commitment may improve patient outcomes as well as interdisciplinary team satisfaction and effectiveness, and is likely to be cost-effective (Zheng et al., 2016).

Finally, this thesis has highlighted the need for policymakers to invest in ward-level, rather than hospital-aggregated, data sets. Comprehensive ward-level data would be directly useful for local improvement work, and would foster a better national understanding of the resources and staff mix required for high quality care. These calculations currently rely on poor quality data (Griffiths et al., 2016b). In the context of a national debate on the adequacy of healthcare funding, policymakers might productively focus their efforts on national implementation of local QI structures that support unit-level improvements. This work suggests that statistical constructs of organisational performance are of less value, and proposals for additional organisational 'league tables' (Monitor, NHS Trust Development Authority, 2016) are unlikely to be the most efficient way of bringing about improvement.

10.5 Personal reflections

As a medical doctor, this period of postgraduate research has certainly made me reflect on my clinical practice. I began this thesis as a believer in the hard benefits of interdisciplinary team care: indeed, I was surprised to find so little objective evidence of them. I remain convinced that high performing interdisciplinary teams hold the key to timely, safe, cost-effective medical care. For them to fulfil their potential, teams must be empowered by work systems that make sense, not predispose them to error; adequate staffing, not skeleton staffing; smart equipment that nudges, not overwhelms; and learning systems that build better practice, before a serious incident. Although many of these challenges are international ones, shared by the worldwide medical community, there are particular challenges ahead for the United Kingdom in healthcare funding and staffing. It is important that these basic tenets of a safe service are not put aside: teams are important, but they are not everything. As I return to full-time clinical work, I take with me the belief that we need to be much harder on our systems, and much kinder to our people.

10.6 Concluding remarks

Defining and improving the performance of medical wards is a priority, not only because of the scale of the care they deliver, but because their stock in trade – dealing with complex, increasingly frail patients – is a standard bearer for the challenges facing health systems more broadly. In this thesis, I discussed a limited selection of opportunities for medical ward improvement: ward-level organisation of interdisciplinary team care, PCTS, alignment between QI efforts and organisational priorities, and a setting-specific understanding of what patients and carers hold most important. I am hopeful that these improvement opportunities will be taken up.

To my knowledge, this thesis represents the first specific compilation of academic medical ward research from the UK. Much of the ongoing work in this field is done under the auspices of

academic hospitalists in the US. I have learned a great deal from their experiences of delivering quality improvement, and their systems for conducting health services research in inpatient medicine. We in the UK must build our own clinical research capacity in this field.

The enthusiasm and dedication of the many teams who embraced HEADS-UP were inspiring, and one must be phlegmatic about the structural and historical barriers to making QI converts of every department. Frontline interdisciplinary teams can change their own practice and bring about meaningful service development, improving patients' outcomes. The challenge for the future is to instil in each organisation the infrastructure required for them to do so.

Bibliography

Academy of Medical Royal Colleges (2012). The benefits of consultant-delivered care (see p. 56).

Advisory Committee for the Safety of Nuclear Installations (1993). Human Factors Study Group Third Report: Organising for safety. London: HSE Books (see p. 22).

Agency for Healthcare Research and Quality (2005). Medical Teamwork and Patient Safety: The Evidence-based Relation. http://archive.ahrq.gov/research/findings/final-reports/medteam/index.html (accessed 07/07/2016) (see p. 293).

Agency for Healthcare Research and Quality (2016). Saving Lives and Saving Money: Hospital-Acquired Conditions Update. http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2014.html (accessed 24/06/2016) (see p. 7).

Agency for Healthcare Research and Quality Patient Safety Network (2014). The Collection. http://psnet.ahrq.gov (accessed 20/08/2014) (see p. 37).

Ahmad A, Purewal TS, Sharma D, Weston PJ (2011). The impact of twice-daily consultant ward rounds on the length of stay in two general medical wards. *Clin Med* **11**; 6:524–8. doi: 10.7861/clinmedicine.11-6-524 (see p. 57).

Ahmed M, Arora S, Tiew S, Hayden J et al. (2014). Building a safer foundation: the Lessons Learnt patient safety training programme. *BMJ Qual Saf* **23**; 1:78–86. doi: 10.1136/bmjqs-2012-001740 (see p. 296).

Allcock C, Dormon F, Taunt R, Dixon J (2015). Constructive comfort: accelerating change in the NHS. London: Health Foundation (see pp. 186, 187, 217).

American Hospital Association (2013). Reconfiguring the bedside care team of the future. (see p. 21).

Amin Y, Grewcock D, Andrews S, Halligan A (2012). Why patients need leaders: introducing a ward safety checklist. *J R Soc Med* **105**; 9:377–83. doi: 10.1258/jrsm.2012.120098 (see pp. 62, 63).

Andrews LB, Stocking C, Krizek T, Gottlieb L et al. (1997). An alternative strategy for studying adverse events in medical care. *Lancet* **349**; 9048:309–13. doi: 10.1016/S0140-6736(96)08268-2 (see p. 112).

Andrews T, Waterman H (2005). Packaging: a grounded theory of how to report physiological deterioration effectively. *J Adv Nurs* **52**; 5:473–81. doi: 10.1111/j.1365-2648.2005.03615.x (see p. 21).

Anhøj J, Hellesoe AB (2016). The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2015-004951 (see p. 275).

Anthony T, Murray BW, Sum-Ping JT, al. et (2011). Evaluating an evidence-based bundle for preventing surgical site infection: A randomized trial. *Arch Surg* **146**; 3:263–9. doi: 10.1001/archsurg.2010.249 (see p. 152).

Arora VM, Greenstein EA, Woodruff JN, Staisiunas PG et al. (2013). Implementing peer evaluation of handoffs: associations with experience and workload. *J Hosp Med* **8**; 3:132–6. doi: 10.1002/jhm.2002 (see p. 61).

Arora VM, Manjarrez E, Dressler DD, Basaviah P et al. (2009). Hospitalist handoffs: a systematic review and task force recommendations. *J Hosp Med* **4**; 7:433–40. doi: 10.1002/jhm.573 (see p. 62).

Arora V, Johnson J, Lovinger D, Humphrey HJ et al. (2005). Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis. *Qual Saf Health Care* **14**; 6:401–7. doi: 10.1136/qshc.2005.015107 (see p. 61).

Association of American Medical Colleges (2014). Guiding principles for public reporting of provider performance. (see p. 275).

Auerbach AD, Sehgal NL, Blegen MA, Maselli J et al. (2012). Effects of a multicentre teamwork and communication programme on patient outcomes: results from the Triad for Optimal Patient Safety (TOPS) project. *BMJ Qual Saf* **21**; 2:118–26. doi: 10.1136/bmjqs-2011-000311 (see pp. 60, 65, 81, 92, 94–96, 98).

Auerbach AD, Wachter RM, Cheng HQ, Maselli J et al. (2010). Comanagement of surgical patients between neurosurgeons and hospitalists. *Arch Intern Med* **170**; 22:2004–10. doi: 10.1001/archinternmed.2010.432 (see p. 16).

Averbukh Y, Southern W (2012). The impact of the number of admissions to the inpatient medical teaching team on patient safety outcomes. *J Grad Med Educ* **4**; 3:307–11. doi: 10.4300/JGME-D-11-00190.1 (see p. 56).

Bailey TC, Chen Y, Mao Y, Lu C et al. (2013). A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med* **8**; 5:236–42. doi: 10.1002/jhm.2009 (see p. 65).

Baines R, Langelaan M, Bruijne M de, Spreeuwenberg P et al. (2015). How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time. *BMJ Qual Saf* **24**;561–71. doi: 10.1136/bmjqs-2014-003702 (see pp. 7, 10, 22).

Baron J, Ritov I (2004). Omission bias, individual differences, and normality. *Organ Behav Hum Dec* **94**; 2:74–85. doi: 10.1016/j.obhdp.2004.03.003 (see p. 19).

Bate P (2014). Context is Everything. Perspectives on Context. London: Health Foundation (see p. 184).

Baxter SK, Brumfitt SM (2008). Benefits and losses: a qualitative study exploring healthcare staff perceptions of teamworking. *Qual Saf Health Care* **17**; 2:127–30. doi: 10.1136/qshc.2007.022277 (see p. 296).

Beckett DJ, Inglis M, Oswald S, Thomson E et al. (2013). Reducing cardiac arrests in the acute admissions unit: a quality improvement journey. *BMJ Qual Saf* **22**; 12:1025–31. doi: 10.1136/bmjqs-2012-001404 (see pp. 62, 66, 109).

Bell D, Lambourne A, Percival F, Laverty AA et al. (2013). Consultant input in acute medical admissions and patient outcomes in hospitals in England: a multivariate analysis. *PLoS One* **8**; 4:e61476. doi: 10.1371/journal.pone.0061476 (see p. 57).

Benning A, Dixon-Woods M, Nwulu U, Ghaleb M et al. (2011). Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* **342**; 7793:d199. doi: 10.1136/bmj.d199 (see p. 12).

Berger Z, Flickinger TE, Pfoh E, Martinez KA et al. (2014). Promoting engagement by patients and families to reduce adverse events in acute care settings: a systematic review. *BMJ Qual Saf* **23**; 7:548–55. doi: 10.1136/bmjqs-2012-001769 (see p. 274).

Bergs J, Hellings J, Cleemput I, Zurel O et al. (2014). Systematic review and meta-analysis of the effect of the World Health Organization surgical safety checklist on post-operative complications. *Br J Surg* **101**; 3:150–8. doi: 10.1002/bjs.9381 (see p. 74).

Berwick DM (2016). Era 3 for Medicine and Health Care. *JAMA* **315**; 13:1329–30. doi: 10.1001/jama.2016.1509 (see p. 29).

Berwick DM, James B, Coye MJ (2003). Connections between quality measurement and improvement. *Med Care* **41 (1 Suppl)**;I30–8 (see p. 273).

Berwick DM, Nolan TW, Whittington J (2008). The Triple Aim: Care, Health, And Cost. *Health Aff (Millwood)* **27**; 3:759–69. doi: 10.1377/hlthaff.27.3.759 (see p. 7).

Bhamidipati VS, Elliott DJ, Justice EM, Belleh E et al. (2016). Structure and outcomes of interdisciplinary rounds in hospitalized medicine patients: A systematic review and suggested taxonomy. *J Hosp Med* **11**; 7:513–23. doi: 10.1002/jhm.2575 (see pp. 105, 286).

Bharwani AM, Harris GC, Southwick FS (2012). Perspective: a business school view of medical interprofessional rounds: transforming rounding groups into rounding teams. *Acad Med* **87**; 12:1768–71. doi: 10.1097/ACM.0b013e318271f8da (see p. 18).

Bhattacharyya O, Reeves S, Garfinkel S, Zwarenstein M (2006). Designing theoretically-informed implementation interventions: fine in theory, but evidence of effectiveness in practice is needed. *Implement Sci* **1**;5. doi: 10.1186/1748-5908-1-5 (see p. 238).

Bilimoria KY, Chung JW, Hedges LV, Dahlke AR et al. (2016). National Cluster-Randomized Trial of Duty-Hour Flexibility in Surgical Training. *N Engl J Med* **374**; 8:713–27. doi: doi:10.1056/NEJMoa1515724 (see p. 17).

Bion J, Richardson A, Hibbert P, Beer J et al. (2013). '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* **22**; 2:110–23. doi: 10.1136/bmjqs-2012-001325 (see pp. 12, 139, 232, 233).

Birken SA, Lee SY, Weiner BJ (2012). Uncovering middle managers' role in healthcare innovation implementation. *Implement Sci* **7**;28. doi: 10.1186/1748-5908-7-28 (see pp. 157, 228–230).

Birken SA, Lee SY, Weiner BJ, Chin MH et al. (2013). Improving the effectiveness of health care innovation implementation: middle managers as change agents. *Med Care Res Rev* **70**; 1:29–45. doi: 10.1177/1077558712457427 (see p. 230).

Birken SA, Lee SY, Weiner BJ, Chin MH et al. (2015). From strategy to action: how top managers' support increases middle managers' commitment to innovation implementation in health care organizations. *Health Care Manage Rev* **40**; 2:159–68. doi: 10.1097/HMR.0000000000000018 (see pp. 230, 231).

Blackman J, Walsgrove J, Gunawardena I (2013). A study examining rates of medical staff recognition of pressure ulceration in hospital inpatients. *Postgrad Med J* **89**; 1051:258–61. doi: 10.1136/postgradmedj-2012-131411 (see p. 124).

Blackmore CC, Kaplan GS (2016). Lean and the perfect patient experience. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2016-005273 (see p. 10).

Blegen MA, Goode CJ, Park SH, Vaughn T et al. (2013). Baccalaureate education in nursing and patient outcomes. *J Nurs Adm* **43**; 2:89–94. doi: 10.1097/NNA.0b013e31827f2028 (see p. 56).

Blegen MA, Sehgal NL, Alldredge BK, Gearhart S et al. (2010). Republished paper: Improving safety culture on adult medical units through multidisciplinary teamwork and communication interventions: the TOPS Project. *Postgrad Med J* **86**; 1022:729–33. doi: 10.1136/qshc.2008.031252rep (see pp. 65, 66).

Bodenheimer T, Sinsky C (2014). From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* **12**; 6:573–6. doi: 10.1370/afm.1713 (see p. 7).

Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ (2009). Reality check for checklists. *Lancet* **374**; 9688:444–5. doi: 10.1016/S0140-6736(09)61440-9 (see pp. 74, 117).

Boyd CM, Darer J, Boult C, Fried LP et al. (2005). Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA* **294**; 6:716–24. doi: 10.1001/jama.294.6.716 (see p. 63).

Boyd J (2007). The 2006 inpatients importantance study. Oxford: Picker Institute (see pp. 274, 275).

Brady PW, Muething S, Kotagal U, Ashby M et al. (2013). Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics* **131**; 1:e298–308. doi: 10.1542/peds.2012-1364 (see pp. 112, 124).

Braithwaite J, Runciman WB, Merry AF (2009). Towards safer, better healthcare: harnessing the natural properties of complex sociotechnical systems. *Qual Saf Health Care* **18**; 1:37–41. doi: 10.1136/qshc.2007.023317 (see p. 230).

Braun V, Clarke V (2006). Using thematic analysis in psychology. *Qual Res Psychol* **3**; 2:77–101. doi: 10.1191/1478088706qp063oa (see pp. 158, 187, 246).

Bristol Royal Infirmary Inquiry (2001). Learning from Bristol. The report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995. London: The Stationery Office (see p. 22).

Broek J van den, Boselie P, Paauwe J (2014). Multiple Institutional Logics in Health Care: 'Productive Ward: Releasing Time to Care'. *Public Management Review* **16**; 1:1–20. doi: 10.1080/14719037.2013.770059 (see p. 9).

Brooks L (30th Jan. 2016). Sharing is caring: how one hospital is leading the way on transparency. *The Guardian*: https://www.theguardian.com/society/2016/jan/30/golden-jubilee-hospital-clydebank-leading-way-transparency-data-share (see p. 275).

Buchanan D (2013). Pure plays and hybrids: Acute trust management profile and capacity. *J Health Serv Res Policy* **18**; 2:90–7. doi: 10.1177/1355819613480350 (see pp. 157, 229).

Bump GM, Bost JE, Buranosky R, Elnicki M (2012). Faculty member review and feedback using a sign-out checklist: improving intern written sign-out. *Acad Med* **87**; 8:1125–31. doi: 10.1097/ACM.0b013e31825d1215 (see p. 61).

Burnett S, Benn J, Pinto A, Parand A et al. (2010). Organisational readiness: exploring the preconditions for success in organisation-wide patient safety improvement programmes. *Qual Saf Health Care* **19**; 4:313–7. doi: 10.1136/qshc.2008.030759 (see p. 227).

Burns E, Rigby E, Mamidanna R, Bottle A et al. (2012). Systematic review of discharge coding accuracy. *J Public Health* **34**; 1:138–48. doi: 10.1093/pubmed/fdr054 (see p. 291).

Byham M, Madathil KC (2015). Developing Consumer-Centered Healthcare Public Reports: Understanding the Healthcare Consumers' Decision Making Process. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **59**; 1:1501–5. doi: 10.1177/1541931215591325 (see p. 242).

Caminiti C, Meschi T, Braglia L, Diodati F et al. (2013). Reducing unnecessary hospital days to improve quality of care through physician accountability: a cluster randomised trial. *BMC Health Serv Res* **13**;14. doi: 10.1186/1472-6963-13-14 (see p. 64).

Camins BC, King MD, Wells JB, Googe HL et al. (2009). Impact of an antimicrobial utilization program on antimicrobial use at a large teaching hospital: A randomized controlled trial. *Infect Control Hosp Epidemiol* **30**; 10:931–8. doi: 10.1086/605924 (see pp. 81, 83, 87, 94–96).

Campbell MJ (2013). Doing clinical trials large enough to achieve adequate reductions in uncertainties about treatment effects. *J R Soc Med* **106**; 2:68–71. doi: 10.1177/0141076813477570 (see p. 134).

Care Quality Commission (2015). Intelligent monitoring - NHS acute hospitals: indicators and methodology guidance. http://www.cqc.org.uk/sites/default/files/20150526_acute_im_v5_indicators_methodology_guidance.pdf (accessed 06/04/2016) (see p. 25).

Care Quality Commission (2016). NHS Surveys: focused on patients' experience. http://www.nhssurveys.org (accessed 23/06/2016) (see p. 276).

Carey MR, Sheth H, Braithwaite RS (2005). A prospective study of reasons for prolonged hospitalizations on a general medicine teaching service. *J Gen Intern Med* **20**; 2:108–15. doi: 10.1111/j.1525-1497.2005.40269.x (see p. 57).

Casalino LP, Gans D, Weber R, Cea M et al. (2016). US Physician Practices Spend More Than $15.4 Billion Annually To Report Quality Measures. *Health Aff (Millwood)* **35**; 3:401–6. doi: 10.1377/hlthaff.2015.1258 (see p. 29).

Center for Healthcare Quality and Safety (2015). Scale computation instructions. https://med.uth.edu/chqs/surveys/safety-attitudes-and-safety-climate-questionnaire/ (accessed 25/05/2016) (see p. 159).

Centers for Medicare and Medicaid Services (2016a). Medicare.gov - Hospital Compare. https://www.medicare.gov/hospitalcompare/search.html (accessed 01/06/2016) (see p. 28).

Centers for Medicare and Medicaid Services (2016[b]). HCAHPS Fact Sheet. June 2015. http://www.hcahpsonline.org/facts.aspx (accessed 23/06/2016) (see p. 276).

Chang Y, Mark B (2011). Effects of learning climate and registered nurse staffing on medication errors. *J Nurs Adm* **41**; 7-8 Suppl:S6–13. doi: 10.1097/NNA.0b013e318221c213 (see p. 55).

Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* **40**; 5:373–83 (see p. 132).

Checkland K, Snow S, McDermott I, Harrison S et al. (2012). 'Animateurs' and animation: what makes a good commissioning manager? *J Health Serv Res Policy* **17**; 1:11–17. doi: 10.1258/jhsrp.2011.011010 (see p. 121).

Chen J, Rathore SS, Wang Y, Radford MJ et al. (2006). Physician board certification and the care and outcomes of elderly patients with acute myocardial infarction. *J Gen Intern Med* **21**; 3:238–44. doi: 10.1111/j.1525-1497.2006.00326.x (see p. 56).

Chen YF, Hemming K, Stevens AJ, Lilford RJ (2016). Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf* **25**; 5:303–10. doi: 10.1136/bmjqs-2015-004372 (see p. 12).

Chu ES, Reid M, Schulz T, Burden M et al. (2009). A structured handoff program for interns. *Acad Med* **84**; 3:347–52. doi: 10.1097/ACM.0b013e3181970829 (see p. 61).

Chung GS, Lawrence RE, Curlin FA, Arora V et al. (2012). Predictors of hospitalised patients' preferences for physician-directed medical decision-making. *J Med Ethics* **38**; 2:77–82. doi: 10.1136/jme.2010.040618 (see p. 244).

Clark J (2012). Medical engagement: too important to be left to chance. The King's Fund (see pp. 225, 226).

Coit MH, Katz JT, McMahon GT (2011). The effect of workload reduction on the quality of residents' discharge summaries. *J Gen Intern Med* **26**; 1:28–32. doi: 10.1007/s11606-010-1465-z (see p. 57).

Cole MG, McCusker J, Bellavance F, Primeau FJ et al. (2002). Systematic detection and multidisciplinary care of delirium in older medical inpatients: a randomized trial. *CMAJ* **167**; 7:753–9 (see pp. 81, 83, 84, 94–96).

Collier J (1957). Photography in Anthropology: A Report on Two Experiments. *Am Anthropol* **59**; 5:843–59. doi: 10.1525/aa.1957.59.5.02a00100 (see p. 245).

Colligan L, Brick D, Patterson E (2015). Changes in Medical Errors with a Handoff Program [letter to the Editor]. *N Engl J Med* **372**;490–1. doi: 10.1056/NEJMc1414788 (see p. 290).

Collins SA, Stein DM, Vawdrey DK, Stetson PD et al. (2011). Content overlap in nurse and physician handoff artifacts and the potential role of electronic health records: a systematic review. *J Biomed Inform* **44**; 4:704–12. doi: 10.1016/j.jbi.2011.01.013 (see p. 61).

Cowan MJ, Shapiro M, Hays RD, Afifi A et al. (2006). The effect of a multidisciplinary hospitalist/physician and advanced practice nurse collaboration on hospital costs. *J Nurs Adm* **36**; 2:79–85 (see p. 59).

Cowen S, Ellison SLR (2006). Reporting measurement uncertainty and coverage intervals near natural limits. *Analyst* **131**; 6:710–17. doi: 10.1039/B518084H (see p. 176).

Cox H, James J, Hunt J (2006). The experiences of trained nurses caring for critically ill patients within a general ward setting. *Intensive Crit Care Nurs* **22**; 5:283–93. doi: 10.1016/j.iccn.2006.02.003 (see p. 21).

Craig P, Dieppe P, Macintyre S, Michie S et al. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* **337**; doi: 10.1136/bmj.a1655 (see pp. 67, 105, 137).

Cumming RG, Sherrington C, Lord SR, Simpson JM et al. (2008). Cluster randomised trial of a targeted multifactorial intervention to prevent falls among older people in hospital. *BMJ* **336**; 7647:758–60. doi: 10.1136/bmj.39499.546030.BE (see p. 132).

Curley C, McEachern JE, Speroff T (1998). A firm trial of interdisciplinary rounds on the inpatient medical wards: an intervention designed using continuous quality improvement. *Med Care* **36**; 8 Suppl:AS4–12 (see pp. 59, 81, 83, 90, 95, 96).

Curry LA, Spatz E, Cherlin E, Thompson JW et al. (2011). What distinguishes top-performing hospitals in acute myocardial infarction mortality rates? A qualitative study. *Ann Intern Med* **154**; 6:384–90. doi: 10.7326/0003-4819-154-6-201103150-00003 (see p. 133).

Damman OC, De Jong A, Hibbard JH, Timmermans DR (2015). Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2015-004120 (see p. 242).

Damman OC, Hendriks M, Rademakers J, Spreeuwenberg P et al. (2012). Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. *Health Expect* **15**; 2:197–211. doi: 10.1111/j.1369-7625.2011.00671.x (see p. 242).

Damschroder LJ, Aron DC, Keith RE, Kirsh SR et al. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* **4**;50. doi: 10.1186/1748-5908-4-50 (see p. 218).

Davey P, Brown E, Charani E, Fenelon L et al. (2013). Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev* **4**;CD003543. doi: 10.1002/14651858.CD003543.pub3 (see p. 80).

Davidoff F, Dixon-Woods M, Leviton L, Michie S (2015). Demystifying theory and its use in improvement. *BMJ Qual Saf* **24**; 3:228–38. doi: 10.1136/bmjqs-2014-003627 (see pp. 113, 114, 238, 239).

Davies L, Ogrinc G (2015). New SQUIRE publication guidelines: supporting nuanced reporting and reflection on complex interventions. *BMJ Qual Saf* **24**; 3:184–5. doi: 10.1136/bmjqs-2015-003949 (see p. 240).

Davis K, Drey N, Gould D (2009). What are scoping studies? A review of the nursing literature. *Int J Nurs Stud* **46**; 10:1386–1400. doi: http://dx.doi.org/10.1016/j.ijnurstu.2009.02.010 (see p. 36).

Deilkas E, Hofoss D (2010). Patient safety culture lives in departments and wards: multilevel partitioning of variance in patient safety culture. *BMC Health Serv Res* **10**;85. doi: 10.1186/1472-6963-10-85 (see pp. 22, 23).

Deneckere S, Euwema M, Lodewijckx C, Panella M et al. (2013). Better interprofessional teamwork, higher level of organized care, and lower risk of burnout in acute health care teams using care pathways: a cluster randomized controlled trial. *Med Care* **51**; 1:99–107. doi: 10.1097/MLR.0b013e3182763312 (see p. 294).

Deneckere S, Euwema M, Van Herck P, Lodewijckx C et al. (2012). Care pathways lead to better teamwork: results of a systematic review. *Soc Sci Med* **75**; 2:264–8. doi: 10.1016/j.socscimed.2012.02.060 (see p. 294).

Deneckere S, Robyns N, Vanhaecht K, Euwema M et al. (2011). Indicators for Follow-Up of Multidisciplinary Teamwork in Care Processes: Results of an International Expert Panel. *Eval Health Prof* **34**; 3:258–277. doi: 10.1177/0163278710393736 (see pp. 294, 295).

Department of Health (2000). An organisation with a memory: report of an expert group on learning from adverse events in the NHS. London: The Stationery Office (see p. 6).

Department of Health (2008). High Quality Care For All: NHS Next Stage Review final report. London: The Stationery Office (see p. 7).

Department of Health (2011). Mixed-sex breaches hit a record low (press release 18th August 2011). https://www.gov.uk/government/news/mixed-sex-breaches-hit-a-record-low (accessed 24/06/2016) (see p. 13).

Desan PH, Zimbrean PC, Weinstein AJ, Bozzo JE et al. (2011). Proactive psychiatric consultation services reduce length of stay for admissions to an inpatient medical team. *Psychosomatics* **52**; 6:513–20. doi: 10.1016/j.psym.2011.06.002 (see pp. 81, 83, 90, 95, 96).

Detry MA, Ma Y (2016). Analyzing Repeated Measurements Using Mixed Models. *JAMA* **315**; 4:407–8. doi: 10.1001/jama.2015.19394 (see p. 141).

DeVore AD, Hammill BG, Hardy NC, Eapen ZJ et al. (2016). Has Public Reporting of Hospital Readmission Rates Affected Patient Outcomes? Analysis of Medicare Claims Data. *J Am Coll Cardiol* **67**; 8:963–72. doi: 10.1016/j.jacc.2015.12.037 (see p. 28).

Dey P, Woodman M, Gibbs A, Steele R et al. (2005). Early assessment by a mobile stroke team: a randomised controlled trial. *Age Ageing* **34**; 4:331–8 (see pp. 81, 83, 90, 94, 95).

Dijk T van (2009). Critical discourse studies: a sociocognitive approach. In: *Methods of critical discourse analysis*. Ed. by Wodak R, Meyer M. London: Sage (see p. 184).

Dijkema LM, Keus F, Dieperink W, Horst IC van der et al. (2015). Real-time information on preventable death provided by email from frontline intensivists: results in high response rates with useful information. *BMJ Qual Saf* **24**; 4:288. doi: 10.1136/bmjqs-2014-003899 (see p. 227).

Dixon-Woods M (2014a). The problem of context in quality improvement. Perspectives on context. London: Health Foundation (see pp. 12, 232, 286).

Dixon-Woods M (2014b). Understanding the challenges of improving safety in clinical systems. http://www.health.org.uk/slideshow-understanding-challenges-improving-safety-clinical-systems (accessed 13/11/2015) (see p. 11).

Dixon-Woods M, Agarwal S, Jones D, Young B et al. (2005). Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* **10**; 1:45–53 (see p. 36).

Dixon-Woods M, Baker R, Charles K, Dawson J et al. (2014a). Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study. *BMJ Qual Saf* **23**; 2:106–15. doi: 10.1136/bmjqs-2013-001947 (see p. 111).

Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA et al. (2011). Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q* **89**; 2:167–205. doi: 10.1111/j.1468-0009.2011.00625.x (see pp. 8, 223, 226).

Dixon-Woods M, Leslie M, Tarrant C, Bion J (2013). Explaining Matching Michigan: an ethnographic study of a patient safety program. *Implement Sci* **8**;70. doi: 10.1186/1748-5908-8-70 (see pp. 11, 223, 226, 229).

Dixon-Woods M, Suokas A, Pitchforth E, Tarrant C (2009). An ethnographic study of classifying and accounting for risk at the sharp end of medical wards. *Soc Sci Med* **69**; 3:362–9. doi: 10.1016/j.socscimed.2009.05.025 (see p. 4).

Dixon-Woods M, Martin G, Tarrant C, Bion J et al. (2014b). Safer Clinical Systems: evaluation findings. London: Health Foundation (see pp. 232, 233, 235).

Doherty C, Gatenby M, Hales C (2010). Role of the ward sister: tensions, pressures and opportunities. *Nurs Stand* **24**; 51:35–40. doi: 10.7748/ns2010.08.24.51.35.c7941 (see p. 229).

Donaldson MS (2008). An Overview of To Err is Human: Re-emphasizing the Message of Patient Safety. In: *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Ed. by Hughes RG. Rockville, MD (see p. 6).

Donze J, Aujesky D, Williams D, Schnipper JL (2013). Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med* **173**; 8:632–8. doi: 10.1001/jamainternmed.2013.3023 (see p. 133).

Dopson S, Bennett C, Fitzgerald L, Ferlie E et al. (2012). Health care managers' access and use of management research. Final report. NIHR Service Delivery and Organisation programme (see p. 229).

Doran GT (1981). There's a S.M.A.R.T. way to write managements's goals and objectives. *Manag Rev* **70**; 11:35 (see p. 10).

Dowding D, Randell R, Gardner P, Fitzpatrick G et al. (2015). Dashboards for improving patient care: Review of the literature. *Int J Med Inform* **84**; 2:87–100. doi: http://dx.doi.org/10.1016/j.ijmedinf.2014.10.001 (see p. 277).

Edwards C, Fox R, Gillard S, Gourlay S et al. (2013). Explaining health managers' information seeking behaviour and use. Final report. NIHR Service Delivery and Organisation programme (see p. 229).

Effective Practice and Organisation of Care (2013). Suggested risk of bias criteria for EPOC reviews. http://epocoslo.cochrane.org/epoc-specific-resources-review-authors (accessed 13/08/2014) (see p. 80).

Effective Practice and Organisation of Care (2014). What study designs should be included in an EPOC review and what should they be called? http://epocoslo.cochrane.org/epoc-specific-resources-review-authors (accessed 03/08/2014) (see pp. 12, 76, 103).

Elbel B, Gillespie C, Raven MC (2014). Presenting quality data to vulnerable groups: charts, summaries or behavioral economic nudges? *J Health Serv Res Policy* **19**; 3:161–8. doi: 10.1177/1355819614524186 (see p. 242).

Eldridge S (2011). Sample size calculations for cluster randomised trials. http://www.newton.ac.uk/seminar/20110815152516051 (accessed 11/12/2014) (see p. 132).

Eldridge SM, Ashby D, Kerry S (2006). Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* **35**; 5:1292–300. doi: 10.1093/ije/dyl129 (see p. 134).

Eldridge S, Kerry SM (2012). A practical guide to cluster randomised trials in health services research. Chichester: Wiley (see pp. 141, 142).

Elwyn G, Corrigan JM (2005). The patient safety story. *BMJ* **331**; 7512:302–4. doi: 10.1136/bmj.38562.690104.43 (see p. 5).

Ende J, Kazis L, Ash A, Moskowitz MA (1989). Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. *J Gen Intern Med* **4**; 1:23–30 (see p. 244).

Esparza JM, Shi R, McLarty J, Comegys M et al. (2013). The effect of a clinical medical librarian on in-patient care outcomes. *J Med Libr Assoc* **101**; 3:185–91. doi: 10.3163/1536-5050.101.3.007 (see pp. 81, 83, 88, 95, 96, 98).

Estabrooks CA, Midodzi WK, Cummings GG, Ricker KL et al. (2005). The impact of hospital nursing characteristics on 30-day mortality. *Nurs Res* **54**; 2:74–84 (see pp. 55, 56).

Etchells E, Ho M, Shojania KG (2016). Value of small sample sizes in rapid-cycle quality improvement projects. *BMJ Qual Saf* **25**; 3:202–6. doi: 10.1136/bmjqs-2015-005094 (see p. 26).

Evans SM, Berry JG, Smith BJ, Esterman A et al. (2006). Attitudes and barriers to incident reporting: a collaborative hospital study. *Qual Saf Health Care* **15**; 1:39–43. doi: 10.1136/qshc.2004.012559 (see p. 226).

Evans SM, Smith BJ, Esterman A, Runciman WB et al. (2007). Evaluation of an intervention aimed at improving voluntary incident reporting in hospitals. *Qual Saf Health Care* **16**; 3:169–75. doi: 10.1136/qshc.2006.019349 (see p. 227).

Farley DO, Haviland A, Champagne S, Jain AK et al. (2008). Adverse-event-reporting practices by US hospitals: results of a national survey. *Qual Saf Health Care* **17**; 6:416–23. doi: 10.1136/qshc.2007.024638 (see pp. 111, 226).

Farley D, Sorbero ME, Lovejoy S, Salibury M (2010). Achieving strong teamwork practices in hospital labor and delivery units. Santa Monica, CA: RAND Corporation (see p. 296).

Federico F, Bonacum D (2010). Strengthening the core. Middle managers play a vital role in improving safety. *Healthc Exec* **25**; 1:68–70 (see p. 229).

Fielding R, Kause J, Arnell-Cullen V, Sandeman D (2013). The impact of consultant-delivered multidisciplinary inpatient medical care on patient outcomes. *Clin Med* **13**; 4:344–8. doi: 10.7861/clinmedicine.13-4-344 (see p. 57).

Fine MJ, Stone RA, Lave JR, Hough LJ et al. (2003). Implementation of an evidence-based guideline to reduce duration of intravenous antibiotic therapy and length of stay for patients hospitalized with community-acquired pneumonia: a randomized controlled trial. *Am J Med* **115**; 5:343–51. doi: 10.1016/s0002-9343(03)00395-4 (see pp. 81, 83, 86, 94–96, 98).

Fletcher KE, Reed DA, Arora VM (2011). Patient safety, resident education and resident well-being following implementation of the 2003 ACGME duty hour rules. *J Gen Intern Med* **26**; 8:907–19. doi: 10.1007/s11606-011-1657-1 (see p. 58).

Flin R, Burns C, Mearns K, Yule S et al. (2006). Measuring safety climate in health care. *Qual Saf Health Care* **15**; 2:109–15. doi: 10.1136/qshc.2005.014761 (see p. 22).

Flin R, Mearns K, O'Connor P, Bryden R (2000). Measuring safety climate: identifying the common features. *Safety Science* **34**; 1–3:177–92. doi: http://dx.doi.org/10.1016/S0925-7535(00)00012-6 (see p. 22).

Flott KM, Graham C, Darzi A, Mayer E (2016). Can we use patient-reported feedback to drive change? The challenges of using patient-reported feedback and how they might be addressed. *BMJ Quality & Safety*. doi: 10.1136/bmjqs-2016-005223 (see p. 274).

Floyd SW, Wooldridge B (1997). Middle management's strategic influence and organizational performance. *J Manage Stud* **34**; 3:465–85. doi: 10.1111/1467-6486.00059 (see p. 230).

Forster AJ, Fung I, Caughey S, Oppenheimer L et al. (2006). Adverse events detected by clinical surveillance on an obstetric service. *Obstet Gynecol* **108**; 5:1073–83. doi: 10.1097/01.AOG.0000242565.28432.7c (see p. 112).

Forster AJ, Kyeremanteng K, Hooper J, Shojania KG et al. (2008). The impact of adverse events in the intensive care unit on hospital mortality and length of stay. *BMC Health Serv Res* **8**;259. doi: 10.1186/1472-6963-8-259 (see p. 112).

Forster AJ, Worthington JR, Hawken S, Bourke M et al. (2011). Using prospective clinical surveillance to identify adverse events in hospital. *BMJ Qual Saf* **20**; 9:756–63. doi: 10.1136/bmjqs.2010.048694 (see pp. 17, 112, 113).

Fourcade A, Blache JL, Grenier C, Bourgain JL et al. (2012). Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf* **21**; 3:191–7. doi: 10.1136/bmjqs-2011-000094 (see p. 226).

Francis R (2010). Independent inquiry into care provided by Mid Staffordshire NHS Foundation Trust January 2005 - March 2009. London (see p. 22).

Francis R (2013). Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. The Stationery Office. London: The Stationery Office (see pp. 23, 55).

Frankel A, Grillo SP, Pittman M, Thomas EJ et al. (2008). Revealing and resolving patient safety defects: the impact of leadership WalkRounds on frontline caregiver assessments of patient safety. *Health Serv Res* **43**; 6:2050–66. doi: 10.1111/j.1475-6773.2008.00878.x (see p. 128).

Fretheim A, Tomic O (2015). Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement. *BMJ Quality & Safety* **24**;748–52. doi: 10.1136/bmjqs-2014-003756 (see p. 26).

Fuller C, Michie S, Savage J, McAteer J et al. (2012). The Feedback Intervention Trial (FIT)–improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. *PLoS One* **7**; 10:e41617. doi: 10.1371/journal.pone.0041617 (see p. 126).

Future Hospital Commission (2013). Future hospital: Caring for medical patients. A report from the Future Hospital Commission to the Royal College of Physicians. Royal College of Physicians (see p. 21).

Gamble J, Vaux E (2014). Learning leadership skills in practice through quality improvement. *Clin Med* **14**; 1:12–5. doi: 10.7861/clinmedicine.14-1-12 (see p. 296).

Gardiner JC, Luo Z, Roman LA (2009). Fixed effects, random effects and GEE: what are the differences? *Stat Med* **28**; 2:221–39. doi: 10.1002/sim.3478 (see p. 142).

Gestel YR van, Lemmens VE, Lingsma HF, Hingh IH de et al. (2012). The hospital standardized mortality ratio fallacy: a narrative review. *Med Care* **50**; 8:662–7. doi: 10.1097/MLR.0b013e31824ebd9f (see p. 25).

Giles SJ, Lawton RJ, Din I, McEachan RRC (2013). Developing a patient measure of safety (PMOS). *BMJ Qual Saf* **22**; 7:554–562. doi: 10.1136/bmjqs-2012-000843 (see p. 293).

Gillespie BM, Chaboyer W, Thalib L, John M et al. (2014). Effect of using a safety checklist on patient complications after surgery: a systematic review and meta-analysis. *Anesthesiology* **120**; 6:1380–9. doi: 10.1097/ALN.0000000000000232 (see p. 74).

Girling AJ, Hofer TP, Wu J, Chilton PJ et al. (2012). Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. *BMJ Qual Saf* **21**; 12:1052–6. doi: 10.1136/bmjqs-2012-001202 (see p. 25).

Goddard AF, Hodgson H, Newbery N (2010). Impact of EWTD on patient:doctor ratios and working practices for junior doctors in England and Wales 2009. *Clin Med* **10**; 4:330–5. doi: 10.7861/clinmedicine.10-4-330 (see p. 57).

Gold RB, Singh S, Frost J (1993). The Medicaid eligibility expansions for pregnant women: evaluating the strength of state implementation efforts. *Fam Plann Perspect* **25**; 5:196–207 (see p. 79).

Goldenhar LM, Brady PW, Sutcliffe KM, Muething SE (2013). Huddling for high reliability and situation awareness. *BMJ Qual Saf* **22**; 11:899–906. doi: 10.1136/bmjqs-2012-001467 (see pp. 112, 180).

Goulding L, Adamson J, Watt I, Wright J (2012). Patient safety in patients who occupy beds on clinically inappropriate wards: a qualitative interview study with NHS staff. *BMJ Qual Saf* **21**; 3:218–24. doi: 10.1136/bmjqs-2011-000280 (see p. 60).

Graham KL, Marcantonio ER, Huang GC, Yang J et al. (2013). Effect of a systems intervention on the quality and safety of patient handoffs in an internal medicine residency program. *J Gen Intern Med* **28**; 8:986–93. doi: 10.1007/s11606-013-2391-7 (see p. 61).

Greenhalgh T, Robert G, Macfarlane F, Bate P et al. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* **82**; 4:581–629. doi: 10.1111/j.0887-378X.2004.00325.x (see p. 186).

Greenstein EA, Arora VM, Staisiunas PG, Banerjee SS et al. (2013). Characterising physician listening behaviour during hospitalist handoffs using the HEAR checklist. *BMJ Qual Saf* **22**; 3:203–9. doi: 10.1136/bmjqs-2012-001138 (see p. 61).

Griffiths A, Beaussier AL, Demeritt D, Rothstein H (2016a). Intelligent Monitoring? Assessing the ability of the Care Quality Commission's statistical surveillance

tool to predict quality and prioritise NHS hospital inspections. *BMJ Quality & Safety*. DOI: 10.1136/bmjqs-2015-004687 (see p. 25).

GRIFFITHS P, BALL J, MURRELLS T, JONES S et al. (2016b). Registered nurse, healthcare support worker, medical staffing levels and mortality in English hospital trusts: a cross-sectional study. *BMJ Open* **6**; 2:e008751. DOI: 10.1136/bmjopen-2015-008751 (see p. 297).

GRIFFITHS R (1983). NHS Management Inquiry. London: HMSO (see p. 224).

GUEST G, BUNCE A, JOHNSON L (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods* **18**; 1:59–82. DOI: 10.1177/1525822x05279903 (see p. 246).

HALES B, TERBLANCHE M, FOWLER R, SIBBALD W (2008). Development of medical checklists for improved quality of patient care. *Int J Qual Health Care* **20**; 1:22–30. DOI: 10.1093/intqhc/mzm062 (see p. 63).

HAM C, KIPPING R, MCLEOD H (2002). Capacity, culture and leadership : lessons from experience of improving access to hospital services: final report from the evaluation of the National Booked Admissions Programme first wave pilots. University of Birmingham, Health Services Management Centre (see p. 128).

HAMILTON J, VERRALL T, MABEN J, GRIFFITHS P et al. (2014). One size does not fit all: a qualitative content analysis of the importance of existing quality improvement capacity in the implementation of Releasing Time to Care: the Productive Ward in Saskatchewan, Canada. *BMC Health Serv Res* **14**;642. DOI: 10.1186/s12913-014-0642-x (see pp. 9, 153).

HANDS C, REID E, MEREDITH P, SMITH GB et al. (2013). Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol. *BMJ Qual Saf* **22**; 9:719–26. DOI: 10.1136/bmjqs-2013-001954 (see p. 20).

HARGREAVES JR, GOODMAN C, DAVEY C, WILLEY BA et al. (2016). Measuring implementation strength: lessons from the evaluation of public health strategies in low- and middle-income settings. *Health Policy Plann*. DOI: 10.1093/heapol/czw001 (see p. 79).

Harper D (2002). Talking about pictures: A case for photo elicitation. *Visual Studies* **17**; 1:13–26. doi: 10.1080/14725860220137345 (see p. 245).

Hart S (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. HFES:904–8 (see pp. 132, 159).

Harvey G, Jas P, Walshe K (2015). Analysing organisational context: case studies on the contribution of absorptive capacity theory to understanding inter-organisational variation in performance improvement. *BMJ Qual Saf* **24**; 1:48–55. doi: 10.1136/bmjqs-2014-002928 (see pp. 223, 227, 239).

Harvey G, Kitson A (2016). PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. *Implement Sci* **11**;33. doi: 10.1186/s13012-016-0398-2 (see pp. 117, 288).

Haugen AS, Softeland E, Almeland SK, Sevdalis N et al. (2015). Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. *Ann Surg* **261**; 5:821–8. doi: 10.1097/SLA.0000000000000716 (see pp. 74, 232, 233).

Haugen AS, Softeland E, Eide GE, Sevdalis N et al. (2013). Impact of the World Health Organization's Surgical Safety Checklist on safety culture in the operating theatre: a controlled intervention study. *Br J Anaesth* **110**; 5:807–15. doi: 10.1093/bja/aet005 (see p. 232).

Havyer RD, Wingo MT, Comfere NI, Nelson DR et al. (2014). Teamwork assessment in internal medicine: a systematic review of validity evidence and outcomes. *J Gen Intern Med* **29**; 6:894–910. doi: 10.1007/s11606-013-2686-8 (see p. 133).

Haynes AB, Weiser TG, Berry WR, Lipsitz SR et al. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* **360**; 5:491–9. doi: 10.1056/NEJMsa0810119 (see pp. 62, 232, 233).

Hayward RA, Asch SM, Hogan MM, Hofer TP et al. (2005). Sins of omission: getting too little medical care may be the greatest threat to patient safety. *J Gen Intern Med* **20**; 8:686–91. doi: 10.1111/j.1525-1497.2005.0152.x (see pp. 18, 19).

Health and Social Care Information Centre (2014). Indicator Specification: Summary Hospital-level Mortality Indicator methodology (Appendix D, version 1.17). http://www.hscic.gov.uk/media/11151/Indicator-Specification-Summary-Hospital-level-Mortality-Indicator-methodology/pdf/SHMI_Specification.pdf (accessed 09/10/2014) (see p. 132).

Health and Social Care Information Centre (2016a). Introduction to Healthcare Resource Groups. http://www.hscic.gov.uk/hrg (accessed 04/02/2016) (see pp. 131, 139).

Health and Social Care Information Centre (2016b). Summary Hospital-level Mortality Indicator. http://www.hscic.gov.uk/SHMI (accessed 14/07/2016) (see p. 24).

Health Foundation (2015). Evaluation: what to consider. London: Health Foundation (see p. 289).

Health Foundation (2016). Fit for purpose? Workforce policy in the English NHS. Health Foundation (see p. 21).

Helms AS, Perez TE, Baltz J, Donowitz G et al. (2012). Use of an appreciative inquiry approach to improve resident sign-out in an era of multiple shift changes. *J Gen Intern Med* **27**; 3:287–91. doi: 10.1007/s11606-011-1885-4 (see pp. 61, 62).

Hemming K, Haines TP, Chilton PJ, Girling AJ et al. (2015a). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* **350**;h391. doi: 10.1136/bmj.h391 (see p. 105).

Hemming K, Lilford R, Girling AJ (2015b). Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Stat Med* **34**; 2:181–96. doi: 10.1002/sim.6325 (see p. 134).

Hemming K, Taljaard M (2016). Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. *J Clin Epidemiol* **69**;137–46. doi: 10.1016/j.jclinepi.2015.08.015 (see p. 135).

Herrin J, Harris KG, Kenward K, Hines S et al. (2016). Patient and family engagement: a survey of US hospital practices. *BMJ Quality & Safety* **25**; 3:182–9. doi: 10.1136/bmjqs-2015-004006 (see p. 274).

Hibbard JH, Greene J, Daniel D (2010). What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Med Care Res Rev* **67**; 3:275–93. doi: 10.1177/1077558709356300 (see p. 276).

Hibbert E, St Arnaud S, Dharampaul S (1994). Nurses' satisfaction with the patient care team. *Can J Rehabil* **8**;87–95 (see p. 14).

Hogan H, Healey F, Neale G, Thomson R et al. (2012). Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf* **21**; 9:737–45. doi: 10.1136/bmjqs-2011-001159 (see p. 19).

Hogan H, Healey F, Neale G, Thomson R et al. (2014). Relationship between preventable hospital deaths and other measures of safety: an exploratory study. *Int J Qual Health Care* **26**; 3:298–307. doi: 10.1093/intqhc/mzu049 (see p. 25).

Hogan H, Zipfel R, Neuburger J, Hutchings A et al. (2015). Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ* **351**;h3239. doi: 10.1136/bmj.h3239 (see p. 25).

Horwitz LI, Moin T, Krumholz HM, Wang L et al. (2008). Consequences of inadequate sign-out for patient care. *Arch Intern Med* **168**; 16:1755–60. doi: 10.1001/archinte.168.16.1755 (see p. 61).

Horwitz LI, Moin T, Krumholz HM, Wang L et al. (2009). What are covering doctors told about their patients? Analysis of sign-out among internal medicine house staff. *Qual Saf Health Care* **18**; 4:248–55. doi: 10.1136/qshc.2008.028654 (see p. 61).

Huang C, Loewen P, Pelletier T, Slater J et al. (2008). Implementation of proven interventions in general medical inpatients: development and evaluation of a new quality indicator for drug therapy. *Qual Saf Health Care* **17**; 4:269–74. doi: 10.1136/qshc.2007.023416 (see pp. 18, 62).

Huang X, Lin J, Demner-Fushman D (2006). Evaluation of PICO as a Knowledge Representation for Clinical Questions. *AMIA Annual Symposium Proceedings* **2006**;359–363 (see p. 75).

Hurst K (2008). UK ward design: patient dependency, nursing workload, staffing and quality-an observational study. *Int J Nurs Stud* **45**; 3:370–81. doi: 10.1016/j.ijnurstu.2006.09.007 (see pp. 3, 13).

Hussey MA, Hughes JP (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* **28**; 2:182–91. doi: 10.1016/j.cct.2006.05.007 (see p. 134).

Hutchinson A, Young TA, Cooper KL, McIntosh A et al. (2009). Trends in healthcare incident reporting and relationship to safety and quality data in acute hospitals: results from the National Reporting and Learning System. *Qual Saf Health Care* **18**; 1:5–10. doi: 10.1136/qshc.2007.022400 (see p. 133).

Huy QN (2001). In praise of middle managers. *Harv Bus Rev* **79**; 8:72–9, 160 (see p. 230).

Ilgen DR, Hollenbeck JR, Johnson M, Jundt D (2005). Teams in organizations: from input-process-output models to IMOI models. *Annu Rev Psychol* **56**;517–43. doi: 10.1146/annurev.psych.56.091103.070250 (see p. 293).

Institute of Medicine (2001). Crossing the quality chasm: a new health system for the 21st century. Washington, D.C.: National Academy Press (see p. 7).

Jenq GY, Doyle MM, Belton BM, Herrin J et al. (2016). Quasi-experimental evaluation of the effectiveness of a large-scale readmission reduction program. *JAMA Intern Med* **176**; 5:681–90. doi: 10.1001/jamainternmed.2016.0833 (see p. 139).

Jha AK, Epstein AM (2013). A survey of board chairs of English hospitals shows greater attention to quality of care than among their US counterparts. *Health Aff (Millwood)* **32**; 4:677–85. doi: 10.1377/hlthaff.2012.1060 (see p. 228).

Jha AK, Orav EJ, Li Z, Epstein AM (2007). The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood)* **26**; 4:1104–10. doi: 10.1377/hlthaff.26.4.1104 (see p. 133).

Jha A, Epstein A (2010). Hospital governance and the quality of care. *Health Aff (Millwood)* **29**; 1:182–7. doi: 10.1377/hlthaff.2009.0297 (see p. 228).

Jiang HJ, Lockee C, Bass K, Fraser I (2009). Board Oversight of Quality: Any Differences in Process of Care and Mortality? *J Healthc Manag* **54**; 1:15–30 (see p. 228).

Jones A, Lankshear A, Kelly D (2016). Giving voice to quality and safety matters at board level: A qualitative study of the experiences of executive nurses working in England and Wales. *Int J Nurs Stud* **59**;169–76. doi: http://dx.doi.org/10.1016/j.ijnurstu.2016.04.007 (see p. 274).

Jones B, Woodhead T (2015). Building the foundations for improvement. How five UK trusts built quality improvement capability at scale within their organisations. London: Health Foundation (see p. 228).

Jones J, Hunter D (1995). Qualitative Research: Consensus methods for medical and health services research. *BMJ* **311**; 7001:376–80. doi: 10.1136/bmj.311.7001.376 (see p. 288).

Jong JD de, Westert GP, Lagoe R, Groenewegen PP (2006). Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work? *Health Serv Res* **41**; 2:374–94. doi: 10.1111/j.1475-6773.2005.00486.x (see p. 64).

Joynt KE, Jha AK (2012). Thirty-day readmissions–truth and consequences. *N Engl J Med* **366**; 15:1366–9. doi: 10.1056/NEJMp1201598 (see p. 26).

Joynt KE, Orav EJ, Zheng J, Jha AK (2016). Public Reporting of Mortality Rates for Hospitalized Medicare Patients and Trends in Mortality for Reported Conditions. *Ann Intern Med* **165**; 3:153–60. doi: 10.7326/M15-1462 (see p. 28).

Kane RL, Shamliyan TA, Mueller C, Duval S et al. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Med Care* **45**; 12:1195–204. doi: 10.1097/MLR.0b013e3181468ca3 (see p. 55).

Kaplan HC, Provost LP, Froehle CM, Margolis PA (2012). The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf* **21**; 1:13–20. doi: 10.1136/bmjqs-2011-000010 (see pp. 223, 239).

Kastner M, Tricco AC, Soobiah C, Lillie E et al. (2012). What is the most appropriate knowledge synthesis method to conduct a review? Protocol for a scoping review. *BMC Medical Research Methodology* **12**; 1:1–10. doi: 10.1186/1471-2288-12-114 (see p. 36).

Kennedy TJ, Regehr G, Baker GR, Lingard L (2009). Preserving professional credibility: grounded theory study of medical trainees' requests for clinical support. *BMJ* **338**;b128. doi: 10.1136/bmj.b128 (see pp. 58, 124, 133).

Keogh B (2013). Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report (see pp. 23, 25).

Kessner DM, Kalk CE, Singer J (1973). Assessing Health Quality-the Case for Tracers. *New England Journal of Medicine* **288**; 4:189–94. doi: doi:10.1056/NEJM197301252880406 (see p. 27).

Kim CS, King E, Stein J, Robinson E et al. (2014). Unit-based interprofessional leadership models in six US hospitals. *J Hosp Med* **9**; 8:545–50. doi: 10.1002/jhm.2200 (see p. 293).

King ES, Moyer DV, Couturie MJ, Gaughan JP et al. (2006). Getting doctors to report medical errors: project DISCLOSE. *Jt Comm J Qual Patient Saf* **32**; 7:382–92 (see p. 227).

Kings College London (2009). The Productive Ward: Releasing time to care. Learning and impact review. London: NHS Institute for Innovation and Improvement (see pp. 9, 26).

Kings College London (2011). The Productive Ward: Releasing time to care. Rapid impact assessment. London: NHS Institute for Innovation and Improvement (see p. 10).

Kircher TT, Wormstall H, Muller PH, Schwarzler F et al. (2007). A randomised trial of a geriatric evaluation and management consultation services in frail hospitalised patients. *Age Ageing* **36**; 1:36–42. doi: 10.1093/ageing/afl102 (see pp. 81, 83, 85, 94, 95, 98, 100).

Kitson AL, Schultz TJ, Long L, Shanks A et al. (2013). The prevention and reduction of weight loss in an acute tertiary care setting: protocol for a pragmatic stepped wedge randomised cluster trial (the PRoWL project). *BMC Health Serv Res* **13**;299. doi: 10.1186/1472-6963-13-299 (see p. 126).

Kitson A, Harvey G (2015). Getting started with facilitation: the facilitator's role. In: *Implementing evidence-based practice in healthcare: a facilitation guide*. Ed. by Kitson A, Harvey G. Abingdon, Oxon: Routledge:70–84 (see p. 121).

Knox R (23rd Mar. 1995). Doctor's orders killed cancer patient: Dana-Farber admits drug overdose caused death of Globe columnist, damage to second woman. *The Boston Globe*: Metro/Region section:p1 (see p. 5).

Ko HC, Turner TJ, Finnigan MA (2011). Systematic review of safety checklists for use by medical care teams in acute hospital settings - limited evidence of effectiveness. *BMC Health Serv Res* **11**;211. doi: 10.1186/1472-6963-11-211 (see pp. 63, 66, 67).

Kohn LT, Corrigan J, Donaldson MS (2000). To err is human: building a safer health system. Washington, D.C.: National Academy Press (see p. 5).

Korbkitjaroen M, Vaithayapichet S, Kachintorn K, Jintanothaitavorn D et al. (2011). Effectiveness of comprehensive implementation of individualized bundling infection control measures for prevention of health care-associated infections in general medical wards. *American Journal of Infection Control* **39**; 6:471–76. doi: 10.1016/j.ajic.2010.09.017 (see pp. 81, 83, 87, 95, 96).

Koroshetz W (2015). A core set of trial outcomes for every medical discipline? *BMJ* **350**;h85. doi: 10.1136/bmj.h85 (see pp. 28, 104).

Krein SL, Damschroder LJ, Kowalski CP, Forman J et al. (2010). The influence of organizational context on quality improvement and patient safety efforts in infection prevention: a multi-center qualitative study. *Soc Sci Med* **71**; 9:1692–701. doi: 10.1016/j.socscimed.2010.07.041 (see p. 197).

Kucukarslan SN, Peters M, Mlynarek M, Nafziger DA (2003). Pharmacists on rounding teams reduce preventable adverse drug events in hospital general medicine units. *Arch Intern Med* **163**; 17:2014–8. doi: 10.1001/archinte.163.17.2014 (see pp. 81, 83, 88, 95).

Kul S, Vanhaecht K, Panella M (2014). Intraclass correlation coefficients for cluster randomized trials in care pathways and usual care: hospital treatment for heart failure. *BMC Health Serv Res* **14**;84. doi: 10.1186/1472-6963-14-84 (see p. 132).

Lagoe RJ, Westert GP, Kendrick K, Morreale G et al. (2005). Managing hospital length of stay reduction: a multihospital approach. *Health Care Manage Rev* **30**; 2:82–92 (see pp. 64, 66).

Lai JC, Montero A, Lebwohl B, Brown Jr RS (2009). A novel housestaff educational model for quaternary-care patients at an academic health center. *Acad Med* **84**; 2:206–11. doi: 10.1097/acm.0b013e31819382d3 (see pp. 81, 83, 85, 95, 96).

Lamba AR, Linn K, Fletcher KE (2014). Identifying patient safety problems during team rounds: an ethnographic study. *BMJ Qual Saf* **23**; 8:667–9. doi: 10.1136/bmjqs-2013-002324 (see p. 112).

Lawton R, O'Hara JK, Sheard L, Reynolds C et al. (2015). Can staff and patient perspectives on hospital safety predict harm-free care? An analysis of staff and patient survey data and routinely collected outcomes. *BMJ Qual Saf* **24**; 6:369–76. doi: 10.1136/bmjqs-2014-003691 (see p. 111).

Leape LL (2014). The checklist conundrum. *N Engl J Med* **370**; 11:1063–4. doi: 10.1056/NEJMe1315851 (see p. 227).

Lear R, Vincent C, Van Herzeele I, Cheshire N et al. (2013). Structured team self-report of intraoperative error can identify obstacles to safe surgery. *Jt Comm J Qual Patient Saf* **39**; 10:480 (see pp. 112, 227).

Lemieux-Charles L, McGuire WL (2006). What Do We Know about Health Care Team Effectiveness? A Review of the Literature. *Med Care Res Rev* **63**; 3:263–300. doi: 10.1177/1077558706287003 (see pp. 294, 295).

Lesprit P, Landelle C, Brun-Buisson C (2013). Clinical impact of unsolicited post-prescription antibiotic review in surgical and medical wards: A randomized controlled trial. *Clin Microbiol Infect* **19**; 2:E91–7. doi: 10.1111/1469-0691.12062 (see pp. 81, 83, 88, 94, 95, 98).

Levac D, Colquhoun H, O'Brien KK (2010). Scoping studies: advancing the methodology. *Implement Sci* **5**; 1:1–9. doi: 10.1186/1748-5908-5-69 (see p. 36).

Leviton L (2013). Webinar: The benefits and challenges of evaluating improvements (see p. 287).

Leviton LC, Khan LK, Rog D, Dawkins N et al. (2010). Evaluability assessment to improve public health policies, programs, and practices. *Annu Rev Public Health* **31**;213–33. doi: 10.1146/annurev.publhealth.012809.103625 (see pp. 288, 289).

Lewis C, Fischer S, Weiner B, Stanick C et al. (2015). Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implement Sci* **10**; 1:155. doi: 10.1186/s13012-015-0342-x (see p. 217).

Leykum LK, Chesser H, Lanham HJ, Carla P et al. (2015). The Association Between Sensemaking During Physician Team Rounds and Hospitalized Patients' Outcomes. *J Gen Intern Med* **30**; 12:1821–7. doi: 10.1007/s11606-015-3377-4 (see pp. 291, 294–296).

Li JW, Morway L, Velasquez A, Weingart SN et al. (2015). Perceptions of medical errors in cancer care: an analysis of how the news media describe sentinel events. *J Patient Saf* **11**; 1:42–51. doi: 10.1097/pts.0000000000000039 (see p. 5).

Lilford R, Mohammed MA, Spiegelhalter D, Thomson R (2004). Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* **363**; 9415:1147–54. doi: 10.1016/S0140-6736(04)15901-1 (see p. 25).

Lilford R, Pronovost P (2010). Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* **340**;c2016. doi: 10.1136/bmj.c2016 (see p. 25).

Lingard L, Regehr G, Orser B, Reznick R et al. (2008). Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Arch Surg* **143**; 1:12–7, discussion 18. doi: 10.1001/archsurg.2007.21 (see p. 226).

Lisby M, Thomsen A, Nielsen LP, Lyhne NM et al. (2010). The effect of systematic medication review in elderly patients admitted to an acute ward of internal medicine. *Basic Clin Pharmacol Toxicol* **106**; 5:422–7. doi: 10.1111/j.1742-7843.2009.00511.x (see pp. 81, 83, 89, 94–96, 98, 100).

Litvak E, Pronovost PJ (2010). Rethinking rapid response teams. *JAMA* **304**; 12:1375–6. doi: 10.1001/jama.2010.1385 (see p. 65).

Loudon K, Treweek S, Sullivan F, Donnan P et al. (2015). The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* **350**;h2147. doi: 10.1136/bmj.h2147 (see pp. 219, 290).

Lubberding S, Zwaan L, Timmermans DR, Wagner C (2011). The nature and causes of unintended events reported at 10 internal medicine departments. *J Patient Saf* **7**; 4:224–31. doi: 10.1097/PTS.0b013e3182388f97 (see pp. 18, 115, 118).

Lucas BP, Trick WE, Evans AT, Mba B et al. (2012). Effects of 2- vs 4-week attending physician inpatient rotations on unplanned patient revisits, evaluations by trainees, and attending physician burnout: a randomized trial. *JAMA* **308**; 21:2199–207. doi: 10.1001/jama.2012.36522 (see p. 58).

Lugtenberg M, Burgers JS, Clancy C, Westert GP et al. (2011). Current guidelines have limited applicability to patients with comorbid conditions: a systematic analysis of evidence-based guidelines. *PLoS One* **6**; 10:e25987. doi: 10.1371/journal.pone.0025987 (see p. 63).

Ma Y, Mazumdar M, Memtsoudis SG (2012). Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Reg Anesth Pain Med* **37**; 1:99–105. doi: 10.1097/AAP.0b013e31823ebc74 (see pp. 141, 142).

Maben J, Griffiths P, Penfold C, Simon M et al. (2015). One size fits all? Mixed methods evaluation of the impact of 100% single-room accommodation on staff and patient experience, safety and costs. *BMJ Qual Saf* **25**; 4:241–56. doi: 10.1136/bmjqs-2015-004265 (see pp. 13, 16).

MacLeod D, Clarke N (2011). Engaging for success: enhancing performance through employee engagement. Department for Business, Innovation & Skills (see p. 225).

Macrae C (2016). The problem with incident reporting. *BMJ Qual Saf* **25**; 2:71–5. doi: 10.1136/bmjqs-2015-004732 (see p. 159).

Madigosky W, Schaik S van (2016). Context matters: groupthink and outcomes of health care teams. *Med Educ* **50**; 4:387–9. doi: 10.1111/medu.12989 (see p. 287).

Makary MA, Sexton JB, Freischlag JA, Holzmueller CG et al. (2006). Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *J Am Coll Surg* **202**; 5:746–52. doi: 10.1016/j.jamcollsurg.2006.01.017 (see p. 66).

Makowsky MJ, Koshman SL, Midodzi WK, Tsuyuki RT (2009). Capturing outcomes of clinical activities performed by a rounding pharmacist practicing in a team environment: the COLLABORATE study [NCT00351676]. *Med Care* **47**; 6:642–50. doi: 10.1097/mlr.0b013e3181926032 (see pp. 81, 83, 89, 95, 96, 98).

Malterud K, Siersma VD, Guassora AD (2015). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qual Health Res*. doi: 10.1177/1049732315617444 (see p. 246).

Mannheimer B, Ulfvarson J, Eklof S, Bergqvist M et al. (2006). Drug-related problems and pharmacotherapeutic advisory intervention at a medicine clinic. *Eur J Clin Pharmacol* **62**; 12:1075–81 (see pp. 81, 83, 89, 94, 95, 98).

Manuel O, Burnand B, Bady P, Kammerlander R et al. (2010). Impact of standardised review of intravenous antibiotic therapy 72 hours after prescription in two internal medicine wards. *J Hosp Infect* **74**; 4:326–31. doi: 10.1016/j.jhin.2009.07.011 (see pp. 81, 83, 87, 95, 96).

Mardon RE, Khanna K, Sorra J, Dyer N et al. (2010). Exploring relationships between hospital patient safety culture and adverse events. *J Patient Saf* **6**; 4:226–32. doi: 10.1097/PTS.0b013e3181fd1a00 (see pp. 65, 133).

Marshall M, Pagel C, French C, Utley M et al. (2014). Moving improvement research closer to practice: the Researcher-in-Residence model. *BMJ Qual Saf* **23**; 10:801–5. doi: 10.1136/bmjqs-2013-002779 (see p. 217).

Marsteller JA, Sexton JB, Hsu YJ, Hsiao CJ et al. (2012). A multicenter, phased, cluster-randomized controlled trial to reduce central line-associated bloodstream infections in intensive care units*. *Crit Care Med* **40**; 11:2933–9. doi: 10.1097/CCM.0b013e31825fd4d8 (see p. 8).

Martin GP, McKee L, Dixon-Woods M (2015). Beyond metrics? Utilizing 'soft intelligence' for healthcare quality and safety. *Soc Sci Med* **142**;19–26. doi: 10.1016/j.socscimed.2015.07.027 (see p. 112).

Mason SL, Kuruvilla S, Riga CV, Gohel MS et al. (2013). Design and validation of an error capture tool for quality evaluation in the vascular and endovascular surgical theatre. *Eur J Vasc Endovasc Surg* **45**; 3:248–54. doi: 10.1016/j.ejvs.2012.11.028 (see pp. 113, 114).

Massey D, Aitken LM, Chaboyer W (2009). What factors influence suboptimal ward care in the acutely ill ward patient? *Intensive Crit Care Nurs* **25**; 4:169–80. doi: 10.1016/j.iccn.2009.03.005 (see pp. 20, 21).

Mayer EK, Sevdalis N, Rout S, Caris J et al. (2016). Surgical Checklist Implementation Project: The Impact of Variable WHO Checklist Compliance on Risk-adjusted Clinical Outcomes After National Implementation: A Longitudinal Study. *Ann Surg* **263**; 1:58–63. doi: 10.1097/SLA.0000000000001185 (see p. 153).

McAlister FA, Bakal JA, Majumdar SR, Dean S et al. (2014). Safely and effectively reducing inpatient length of stay: a controlled study of the General Internal Medicine Care Transformation Initiative. *BMJ Qual Saf* **23**; 6:446–56. doi: 10.1136/bmjqs-2013-002289 (see pp. 66, 109).

McAllister C, Leykum LK, Lanham H, Reisinger HS et al. (2014). Relationships within inpatient physician housestaff teams and their association with hospitalized patient outcomes. *J Hosp Med* **9**; 12:764–71. doi: 10.1002/jhm.2274 (see p. 296).

McCormack B, McCarthy G, Wright J, Slater P et al. (2009). Development and testing of the Context Assessment Index (CAI). *Worldviews Evid Based Nurs* **6**; 1:27–35. doi: 10.1111/j.1741-6787.2008.00130.x (see p. 218).

McElroy LM, Khorzad R, Nannicelli AP, Brown AR et al. (2016). Failure mode and effects analysis: a comparison of two common risk prioritisation methods. *BMJ Qual Saf* **25**; 5:329–36. doi: 10.1136/bmjqs-2015-004130 (see p. 235).

McGinity R, Salokangas M (2014). Introduction: 'embedded research' as an approach into academia for emerging researchers. *Management in Education* **28**; 1:3–5. doi: 10.1177/0892020613508863 (see p. 122).

McHugh MD, Lake ET (2010). Understanding clinical expertise: nurse education, experience, and the hospital context. *Res Nurs Health* **33**; 4:276–87. doi: 10.1002/nur.20388 (see p. 56).

McHugh MD, Ma C (2013). Hospital nursing and 30-day readmissions among Medicare patients with heart failure, acute myocardial infarction, and pneumonia. *Med Care* **51**; 1:52–9. doi: 10.1097/MLR.0b013e3182763284 (see p. 56).

McIntosh N, Meterko M, Burgess J. F. J, Restuccia JD et al. (2014). Organizational predictors of coordination in inpatient medicine. *Health Care Manage Rev* **39**; 4:279–92. doi: 10.1097/HMR.0000000000000004 (see pp. 293, 295).

McMahon GT, Katz JT, Thorndike ME, Levy BD et al. (2010). Evaluation of a redesign initiative in an internal-medicine residency. *N Engl J Med* **362**; 14:1304–11. doi: 10.1056/NEJMsa0908136 (see p. 57).

McMullen H, Griffiths C, Leber W, Greenhalgh T (2015). Explaining high and low performers in complex intervention trials: a new model based on diffusion of innovations theory. *Trials* **16**;242. doi: 10.1186/s13063-015-0755-5 (see pp. 152, 156, 186, 218, 227).

McNeill G, Bryden D (2013). Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. *Resuscitation* **84**; 12:1652–67. doi: 10.1016/j.resuscitation.2013.08.006 (see pp. 64, 65, 133).

McQuillan P, Pilkington S, Allan A, Taylor B et al. (1998). Confidential inquiry into quality of care before admission to intensive care. *BMJ* **316**; 7148:1853–8 (see pp. 19, 67).

Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ (2011). Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* **64**; 9:936–48. doi: 10.1016/j.jclinepi.2010.12.003 (see p. 126).

Meltzer D, Manning WG, Morrison J, Shah MN et al. (2002). Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists. *Ann Intern Med* **137**; 11:866–74 (see p. 57).

Milliken AD (2014). Physician engagement: a necessary but reciprocal process. *CMAJ* **186**; 4:244–5. doi: 10.1503/cmaj.131178 (see pp. 225, 228).

Moloney ED, Bennett K, Silke B (2007). Effect of an acute medical admission unit on key quality indicators assessed by funnel plots. *Postgrad Med J* **83**; 984:659–63. doi: 10.1136/pgmj.2007.058511 (see p. 62).

Monitor (2014). Costing Webinar 2nd December 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/383127/Costing_and_PLICS.pdf (accessed 05/01/2016) (see p. 4).

Monitor, NHS Trust Development Authority (2016). Learning from mistakes league. https://www.gov.uk/government/publications/learning-from-mistakes-league (accessed 12/07/2016) (see p. 297).

Moore C, Buchanan DA (2013). Sweat the small stuff: A case study of small-scale change processes and consequences in acute care. *Health Services Management Research* **26**; 1:9–17. doi: 10.1177/0951484813488060 (see pp. 121, 297).

Moore GF, Audrey S, Barker M, Bond L et al. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* **350**;h1258. doi: 10.1136/bmj.h1258 (see pp. 67, 105, 137, 184).

Moraros J, Lemstra M, Nwankwo C (2016). Lean interventions in healthcare-do they actually work? A systematic literature review. *Int J Qual Health Care*. doi: 10.1093/intqhc/mzv123 (see p. 10).

Morello RT, Lowthian JA, Barker AL, McGinnes R et al. (2013). Strategies for improving patient safety culture in hospitals: a systematic review. *BMJ Qual Saf* **22**; 1:11–8. doi: 10.1136/bmjqs-2011-000582 (see p. 65).

Morgan L, New S, Robertson E, Collins G et al. (2015). Effectiveness of facilitated introduction of a standard operating procedure into routine processes in the operating theatre: a controlled interrupted time series. *BMJ Qual Saf* **24**; 2:120–7. doi: 10.1136/bmjqs-2014-003158 (see pp. 229–231).

Morris D, Karlson A (2011). Dynamic Accessibility Requirements for Hospital Patients. (see p. 242).

Mosher HJ, Kaboli PJ (2015). Inpatient Interdisciplinary Care: Can the Goose Lay Some Golden Eggs? *JAMA Intern Med* **175**; 8:1298–300. doi: 10.1001/jamainternmed.2015.2430 (see pp. 26, 104, 287).

Mudge A, Laracy S, Richter K, Denaro C (2006). Controlled trial of multidisciplinary care teams for acutely ill medical inpatients: enhanced multidisciplinary care. *Intern Med J* **36**; 9:558–63. doi: 10.1111/j.1445-5994.2006.01135.x (see pp. 81, 91, 94–96, 98, 100).

Muething SE, Goudie A, Schoettker PJ, Donnelly LF et al. (2012). Quality improvement initiative to reduce serious safety events and improve patient safety culture. *Pediatrics* **130**; 2:e423–31. doi: 10.1542/peds.2011-3566 (see p. 180).

Muller-Juge V, Cullati S, Blondon KS, Hudelson P et al. (2013). Interprofessional collaboration on an internal medicine ward: role perceptions and expectations among nurses and residents. *PLoS One* **8**; 2:e57570. doi: 10.1371/journal.pone.0057570 (see p. 59).

Nabors C, Peterson SJ, Aronow WS, Sule S et al. (2011). Physician reporting of clinically significant events through a computerized patient sign-out system. *J Patient Saf* **7**; 3:155–61. doi: 10.1097/PTS.0b013e31822d7a66 (see p. 227).

Nabors C, Peterson SJ, Aronow WS, Sule S et al. (2014). Mobile physician reporting of clinically significant events-a novel way to improve handoff communication and supervision of resident on call activities. *J Patient Saf* **10**; 4:211–7. doi: 10.1097/PTS.0b013e31829952ff (see p. 227).

Nancarrow S, Enderby P, Ariss S, Smith T et al. (2012). Enhancing the effectiveness of interprofessional working: cost and outcomes. Final report. NIHR Service Delivery and Organisation Programme (see pp. 14, 15, 296).

National Institute for Health Research (2014). Stepping Up: A Phased Evaluation of the Impact of High-Intensity Specialist-Led Acute Care (HiSLAC) of Emergency Medical Admissions to NHS Hospitals. http://www.nets.nihr.ac.uk/projects/hsdr/1212817 (accessed 13/07/2016) (see p. 16).

National Institute for Health Research (2016). Evaluating the ten year impact of the Productive Ward at the clinical microsystem level in English acute trusts. http://www.nets.nihr.ac.uk/projects/hsdr/1315744 (accessed 13/07/2016) (see p. 10).

National Outreach Forum (2012). National Outreach Forum Operational Standards and Competencies for Critical Care Outreach Services. (see p. 139).

National Patient Safety Foundation (2015). Free from Harm: Accelerating Patient Safety Improvement Fifteen Years after To Err Is Human (see p. 6).

Needleman J, Buerhaus PI, Stewart M, Zelevinsky K et al. (2006). Nurse staffing in hospitals: is there a business case for quality? *Health Aff (Millwood)* **25**; 1:204–11. doi: 10.1377/hlthaff.25.1.204 (see p. 55).

Needleman J, Buerhaus P, Mattke S, Stewart M et al. (2002). Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med* **346**; 22:1715–22. doi: 10.1056/NEJMsa012247 (see p. 55).

Needleman J, Buerhaus P, Pankratz VS, Leibson CL et al. (2011). Nurse staffing and inpatient hospital mortality. *N Engl J Med* **364**; 11:1037–45. doi: 10.1056/NEJMsa1001025 (see pp. 55, 152).

Neily J, Mills PD, Young-Xu Y, Carney BT et al. (2010). Association between implementation of a medical team training program and surgical mortality. *JAMA* **304**; 15:1693–700. doi: 10.1001/jama.2010.1506 (see p. 19).

Nelson EC, Batalden PB, Huber TP, Mohr JJ et al. (2002). Microsystems in health care: Part 1. Learning from high-performing front-line clinical units. *Jt Comm J Qual Improv* **28**; 9:472–93 (see p. 14).

Nelson EC, Eftimovska E, Lind C, Hager A et al. (2015). Patient reported outcome measures in practice. *BMJ* **350**;g7818. doi: 10.1136/bmj.g7818 (see p. 242).

Neuhauser D (2004). Assessing health quality: the case for tracers. *J Health Serv Res Policy* **9**; 4:246–247. doi: 10.1258/1355819042250168 (see p. 27).

NHS England (2013). NHS Services, Seven Days a Week Forum: Levers and Incentives. https://www.england.nhs.uk/wp-content/uploads/2013/12/levers-and-incentives1.pdf (accessed 01/06/2016) (see p. 28).

NHS England (2014a). 2014/15 Clostridium difficile infection objectives. http://www.england.nhs.uk/ourwork/patientsafety/associated-infections/clostridium-difficile/ (accessed 31/03/2014) (see p. 133).

NHS England (2014b). Friends and Family Test: Guidance. Publications Gateway Ref No.03548 (see p. 247).

NHS England (2014c). Hard Truths commitments regarding the publishing of staffing data - frequently asked questions (FAQs). https://www.england.nhs.uk/wp-content/uploads/2014/04/staff-data-faqs-0414.pdf (accessed 23/06/2016) (see p. 276).

NHS England (2014d). Zero tolerance - guidance on the post infection review. http://www.england.nhs.uk/ourwork/patientsafety/zero-tolerance/ (accessed 31/03/2014) (see p. 133).

NHS England (2015). Serious incident framework (see p. 165).

NHS England (2016a). Seven Day Services Clinical Standards. http://www.nhsiq.nhs.uk/media/2638611/clinical_standards.pdf (accessed 01/06/2016) (see p. 28).

NHS England (2016b). Special measures. http://www.nhs.uk/NHSEngland/specialmeasures/Pages/about-special-measures.aspx (accessed 06/04/2016) (see p. 25).

NHS Estates (2002). Enhancing privacy and dignity: achieving single sex accommodation. The Stationery Office (see p. 13).

NHS Institute for Innovation and Improvement (2013). Quality and Service Improvement Tools: Driver Diagrams. http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/driver_diagrams.html (accessed 18/07/2016) (see p. 114).

NHS Quality Observatory (2016). NHS Safety Thermometer. https://www.safetythermometer.nhs.uk/index.php?option=com_content%5C&view=article%5C&id=1%5C&Itemid=101 (accessed 24/06/2016) (see p. 276).

Nicholas LH, Dimick JB (2011). Too much of a good thing? Multi-tasking and distraction in quality improvement. *Arch Surg* **146**; 3:270. doi: 10.1001/archsurg.2011.7 (see p. 152).

Nielsen PE, Goldman MB, Mann S, Shapiro DE et al. (2007). Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol* **109**; 1:48–55. doi: 10.1097/01.AOG.0000250900.53126.c2 (see p. 104).

Nilsen P (2015). Making sense of implementation theories, models and frameworks. *Implement Sci* **10**; 1:53. doi: 10.1186/s13012-015-0242-0 (see pp. 223, 238).

Norcini JJ, Lipner RS, Kimball HR (2002). Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* **36**; 9:853–9 (see p. 56).

Ocloo J, Matthews R (2016). From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* **25**;626–32. doi: 10.1136/bmjqs-2015-004839 (see p. 225).

O'Connor P, Reddin C, O'Sullivan M, O'Duffy F et al. (2013). Surgical checklists: the human factor. *Patient Saf Surg* **7**; 1:14. doi: 10.1186/1754-9493-7-14 (see p. 226).

O'Hara JK, Lawton RJ (2016). At a crossroads? Key challenges and future opportunities for patient involvement in patient safety. *BMJ Qual Saf* **25**;565–8. doi: 10.1136/bmjqs-2016-005476 (see p. 275).

O'Leary KJ, Boudreau YN, Creden AJ, Slade ME et al. (2012a). Assessment of teamwork during structured interdisciplinary rounds on medical units. *J Hosp Med* **7**; 9:679–83. doi: 10.1002/jhm.1970 (see pp. 14, 296).

O'Leary KJ, Buck R, Fligiel HM, Haviley C et al. (2011a). Structured interdisciplinary rounds in a medical teaching unit: improving patient safety. *Arch Intern Med* **171**; 7:678–84. doi: 10.1001/archinternmed.2011.128 (see pp. 59, 81, 83, 91, 95).

O'Leary KJ, Creden AJ, Slade ME, Landler MP et al. (2015a). Implementation of unit-based interventions to improve teamwork and patient safety on a medical service. *Am J Med Qual* **30**; 5:409–16. doi: 10.1177/1062860614538093 (see p. 294).

O'Leary KJ, Haviley C, Slade ME, Shah HM et al. (2011b). Improving teamwork: impact of structured interdisciplinary rounds on a hospitalist unit. *J Hosp Med* **6**; 2:88–93. doi: 10.1002/jhm.714 (see pp. 65, 81, 83, 92, 95, 96).

O'Leary KJ, Killarney A, Hansen LO, Jones S et al. (2015b). Effect of patient-centred bedside rounds on hospitalised patients' decision control, activation and satisfaction with care. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2015-004561 (see pp. 17, 153).

O'Leary KJ, Liebovitz DM, Baker DW (2006). How hospitalists spend their time: insights on efficiency and safety. *J Hosp Med* **1**; 2:88–93. doi: 10.1002/jhm.88 (see p. 57).

O'Leary KJ, Sehgal NL, Terrell G, Williams MV et al. (2012b). Interdisciplinary teamwork in hospitals: a review and practical recommendations for improvement. *J Hosp Med* **7**; 1:48–54. doi: 10.1002/jhm.970 (see p. 60).

O'Leary KJ, Wayne DB, Haviley C, Slade ME et al. (2010). Improving teamwork: impact of structured interdisciplinary rounds on a medical teaching unit. *J Gen Intern Med* **25**; 8:826–32. doi: 10.1007/s11606-010-1345-6 (see pp. 58–60, 65, 81, 83, 91, 95, 96).

O'Leary KJ, Wayne DB, Landler MP, Kulkarni N et al. (2009). Impact of localizing physicians to hospital units on nurse-physician communication and agreement on the plan of care. *J Gen Intern Med* **24**; 11:1223–7. doi: 10.1007/s11606-009-1113-7 (see p. 60).

O'Leary KJ, Johnson JK, Auerbach AD (2016). Do interdisciplinary rounds improve patient outcomes? Only if they improve teamwork. *J Hosp Med* **11**; 7:524–5. doi: 10.1002/jhm.2587 (see pp. 105, 286, 294).

Olsen S, Neale G, Schwab K, Psaila B et al. (2007). Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. *Qual Saf Health Care* **16**; 1:40–4. doi: 10.1136/qshc.2005.017616 (see p. 111).

O'Mahony S, Mazur E, Charney P, Wang Y et al. (2007). Use of multidisciplinary rounds to simultaneously improve quality outcomes, enhance resident education, and shorten length of stay. *J Gen Intern Med* **22**; 8:1073–9. doi: 10.1007/s11606-007-0225-1 (see p. 59).

Ong M, Bostrom A, Vidyarthi A, McCulloch C et al. (2007). House staff team workload and organization effects on patient outcomes in an academic general internal medicine inpatient service. *Arch Intern Med* **167**; 1:47–52. doi: 10.1001/archinte.167.1.47 (see p. 56).

Overall JE, Tonidandel S (2004). Robustness of Generalized Estimating Equation (GEE) Tests of Significance against Misspecification of the Error Structure Model. *Biom J* **46**; 2:203–13. doi: 10.1002/bimj.200210017 (see p. 142).

Øvretveit J (2011). Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf* **20 Suppl 1**;i18–23. doi: 10.1136/bmjqs.2010.045955 (see p. 226).

Øvretveit J (2014). How does context affect quality improvement? Perspectives on context. London: Health Foundation (see p. 218).

Øvretveit JC, Shekelle PG, Dy SM, McDonald KM et al. (2011). 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* **20**; 7:604–10. doi: 10.1136/bmjqs.2010.047035 (see p. 226).

Oxman AD, Fretheim A, Flottorp S (2005). The OFF theory of research utilization. *J Clin Epidemiol* **58**; 2:113–6. doi: 10.1016/j.jclinepi.2004.10.002 (see p. 238).

Pannick S, Beveridge I, Ashrafian H, Long SJ et al. (2015). A stepped wedge, cluster controlled trial of an intervention to improve safety and quality on medical wards: the HEADS-UP study protocol. *BMJ Open* **5**; 6:e007510. doi: 10.1136/bmjopen-2014-007510 (see pp. 130, 235, 290).

Pannick S, Beveridge I, Wachter RM, Sevdalis N (2014). Improving the quality and safety of care on the medical ward: A review and synthesis of the evidence base. *Eur J Intern Med* **25**; 10:874–87. doi: 10.1016/j.ejim.2014.10.013 (see p. 290).

Parand A, Burnett S, Benn J, Pinto A et al. (2011). The disparity of frontline clinical staff and managers' perceptions of a quality and patient safety initiative. *J Eval Clin Pract* **17**; 6:1184–90. doi: 10.1111/j.1365-2753.2010.01506.x (see p. 229).

Parand A, Dopson S, Renz A, Vincent C (2014). The role of hospital managers in quality and patient safety: a systematic review. *BMJ Open* **4**; 9:e005055. doi: 10.1136/bmjopen-2014-005055 (see pp. 228, 229, 231, 239).

Parand A, Dopson S, Vincent C (2013). The role of chief executive officers in a quality improvement initiative: a qualitative study. *BMJ Open* **3**; 1. doi: 10.1136/bmjopen-2012-001731 (see p. 229).

Parekh VI, Roy CL (2010). Nonphysician providers in hospital medicine: not so fast. *J Hosp Med* **5**; 2:103–6. doi: 10.1002/jhm.584 (see p. 56).

Parry GJ, Carson-Stevens A, Luff DF, McPherson ME et al. (2013). Recommendations for Evaluation of Health Care Improvement Initiatives. *Acad Pediatr* **13**; 6, Supplement:S23–S30. doi: http://dx.doi.org/10.1016/j.acap.2013.04.007 (see p. 114).

Pearse R (2014). 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. http://www.epochtrial.org/docs/EPOCH%20Protocol%20revision%20V2%200%20280414.pdf (accessed 10/12/2014) (see pp. 126, 134).

Pendharkar SR, Woiceshyn J, Silveira GJ da, Bischak D et al. (2016). What happens when healthcare innovations collide? *BMJ Qual Saf* **25**;9–13. doi: 10.1136/bmjqs-2015-004441 (see p. 238).

Person SD, Allison JJ, Kiefe CI, Weaver MT et al. (2004). Nurse staffing and mortality for Medicare patients with acute myocardial infarction. *Med Care* **42**; 1:4–12. doi: 10.1097/01.mlr.0000102369.67404.b0 (see p. 55).

Peters D, Tran N, Adam T (2013). Implementation research in health: a practical guide. Alliance for Health Policy and Systems Research, World Health Organization (see pp. 183, 184, 188).

Petersen LA, Orav EJ, Teich JM, O'Neil AC et al. (1998). Using a computerized sign-out program to improve continuity of inpatient care and prevent adverse events. *Jt Comm J Qual Improv* **24**; 2:77–87 (see p. 61).

Pettigrew AM, Ferlie E, McKee L (1992). Shaping strategic change: the case of the National Health Service. London: Sage (see p. 185).

Pettigrew AM, Woodman RW, Cameron KS (2001). Studying Organizational Change and Development: Challenges for Future Research. *Acad Manage J* **44**; 4:697–713. DOI: 10.2307/3069411 (see p. 184).

Pinnock H, Epiphaniou E, Pearce G, Parke H et al. (2015). Implementing supported self-management for asthma: a systematic review and suggested hierarchy of evidence of implementation studies. *BMC Med* **13**;127. DOI: 10.1186/s12916-015-0361-0 (see p. 238).

Pitkälä KH, Laurila JV, Strandberg TE, Tilvis RS (2006). Multicomponent geriatric intervention for elderly inpatients with delirium: A randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* **61**; 2:176–81 (see pp. 81, 83, 84, 94–96).

Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE et al. (2013). The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. *BMC Cardiovasc Disord* **13**;77. DOI: 10.1186/1471-2261-13-77 (see p. 126).

Popay J, Roberts H, Sowden A, Petticrew M et al. (2006). Guidance on the conduct of narrative synthesis in systematic reviews: a product of the ESRC methods programme. (see p. 70).

Portela MC, Pronovost PJ, Woodcock T, Carter P et al. (2015). How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* **24**; 5:325–36. DOI: 10.1136/bmjqs-2014-003620 (see pp. 8, 12, 292).

Porter ME, Larsson S, Lee TH (2016). Standardizing Patient Outcomes Measurement. *N Engl J Med* **374**; 6:504–6. DOI: 10.1056/NEJMp1511701 (see pp. 28, 292).

Powell BJ, Beidas RS, Lewis CC, Aarons GA et al. (2015a). Methods to Improve the Selection and Tailoring of Implementation Strategies. *J Behav Health Serv Res*. DOI: 10.1007/s11414-015-9475-6 (see p. 219).

Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ et al. (2015b). A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci* **10**;21. doi: 10.1186/s13012-015-0209-1 (see p. 187).

Pronovost PJ, Cleeman JI, Wright D, Srinivasan A (2016). Fifteen years after To Err is Human: a success story to learn from. *BMJ Qual Saf* **25**;396–9. doi: 10.1136/bmjqs-2015-004720 (see p. 6).

Pronovost P, Berenholtz S, Dorman T, Lipsett PA et al. (2003). Improving communication in the ICU using daily goals. *J Crit Care* **18**; 2:71–5. doi: 10.1053/jcrc.2003.50008 (see p. 133).

Pronovost P, Needham D, Berenholtz S, Sinopoli D et al. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* **355**; 26:2725–32. doi: 10.1056/NEJMoa061115 (see pp. 8, 74, 232, 233).

Pronovost P, Sexton B (2005). Assessing safety culture: guidelines and recommendations. *Qual Saf Health Care* **14**; 4:231–3. doi: 10.1136/qshc.2005.015180 (see p. 22).

Proudlove N, Moxham C, Boaden R (2008). Lessons for Lean in healthcare from using six sigma in the NHS. *Public Money & Management* **28**; 1:27–34. doi: 10.1111/j.1467-9302.2008.00615.x (see p. 230).

Provenzano A, Rohan S, Trevejo E, Burdick E et al. (2015). Evaluating inpatient mortality: a new electronic review process that gathers information from front-line providers. *BMJ Qual Saf* **24**; 1:31–7. doi: 10.1136/bmjqs-2014-003120 (see p. 227).

Pucher PH, Aggarwal R, Qurashi M, Singh P et al. (2014). Randomized clinical trial of the impact of surgical ward-care checklists on postoperative care in a simulated environment. *Br J Surg* **101**; 13:1666–73. doi: 10.1002/bjs.9654 (see p. 15).

Quinn JB (1992). Intelligent enterprise : a knowledge and service based paradigm for industry. New York: Free Press (see p. 14).

Raduma-Tomas MA, Flin R, Yule S, Close S (2012). The importance of preparation for doctors' handovers in an acute medical assessment unit: a hierarchical task analysis. *BMJ Qual Saf* **21**; 3:211–7. doi: 10.1136/bmjqs-2011-000220 (see p. 62).

Raleigh V, Thompson J, Jabbal J, Graham C et al. (2015). Patients' experience of using hospital services. London: The King's Fund (see p. 274).

Reames BN, Scally CP, Thumma JR, Dimick JB (2015). Evaluation of the Effectiveness of a Surgical Checklist in Medicare Patients. *Med Care* **53**; 1:87–94. doi: 10.1097/MLR.0000000000000277 (see pp. 104, 229).

Reason JT (1997). Managing the risks of organizational accidents. Aldershot: Ashgate (see p. 6).

Reeves S, Perrier L, Goldman J, Freeth D et al. (2013). Interprofessional education: effects on professional practice and healthcare outcomes (update). *Cochrane Database Syst Rev* **3**;CD002213. doi: 10.1002/14651858.CD002213.pub3 (see p. 60).

Reeves S, Rice K, Conn LG, Miller KL et al. (2009). Interprofessional interaction, negotiation and non-negotiation on general internal medicine wards. *J Interprof Care* **23**; 6:633–45. doi: 10.3109/13561820902886295 (see p. 58).

Reynolds M, Jheeta S, Benn J, Sanghera I et al. (2016). Improving feedback on junior doctors' prescribing errors: mixed-methods evaluation of a quality improvement project. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2015-004717 (see pp. 121, 288).

Rice K, Zwarenstein M, Conn LG, Kenaszchuk C et al. (2010). An intervention to improve interprofessional collaboration and communications: a comparative qualitative study. *J Interprof Care* **24**; 4:350–61. doi: 10.3109/13561820903550713 (see p. 60).

Robb E, Jarman B, Suntharalingam G, Higgens C et al. (2010). Using care bundles to reduce in-hospital mortality: quantitative survey. *BMJ* **340**;c1234. doi: 10.1136/bmj.c1234 (see p. 63).

Robins NS (1995). The Girl Who Died Twice: Every Patient's Nightmare - The Libby Zion Case and the Hidden Hazards of Hospitals. New York, N.Y.: Delacorte Press (see p. 5).

Robinson D (2002). Using Photographs to Elicit Narrative Accounts. In: *Narrative, Memory and Life Transitions*. Ed. by Horrocks C, Milnes K, Roberts B, Robinson D. Huddersfield: Huddersfield University Press:179–87 (see p. 245).

Rooney T, Moloney ED, Bennett K, O'Riordan D et al. (2008). Impact of an acute medical admission unit on hospital mortality: a 5-year prospective study. *QJM* **101**; 6:457–65. doi: 10.1093/qjmed/hcn025 (see p. 62).

Rose G (2014). On the relation between 'visual research methods' and contemporary visual culture. *Sociol Rev* **62**; 1:24–46. doi: 10.1111/1467-954X.12109 (see p. 243).

Rotter T, Kinsman L, James E, Machotta A et al. (2010). Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database Syst Rev* 3:CD006632. doi: 10.1002/14651858.CD006632.pub2 (see pp. 63, 180).

Rotter T, Kinsman L, James E, Machotta A et al. (2012). Have we drawn the wrong conclusions about the value of care pathways? Is a Cochrane review appropriate? Response to the commentary article published by Kris Vanhaecht et al. *Eval Health Prof* **35**; 1:43–6. doi: 10.1177/0163278711409209 (see p. 63).

Royal College of Physicians (2013). The medical registrar: empowering the unsung heroes of patient care. London: RCP (see pp. 16, 17).

Russ SJ, Sevdalis N, Moorthy K, Mayer EK et al. (2015). A qualitative evaluation of the barriers and facilitators toward implementation of the WHO surgical safety checklist across hospitals in England: lessons from the "Surgical Checklist Implementation Project". *Ann Surg* **261**; 1:81–91. doi: 10.1097/SLA.0000000000000793 (see pp. 226, 227).

Rycroft-Malone J, Seers K, Chandler J, Hawkes CA et al. (2013). The role of evidence, context, and facilitation in an implementation trial: implications for the development of the PARIHS framework. *Implement Sci* **8**;28. doi: 10.1186/1748-5908-8-28 (see pp. 218, 230).

Saint S, Fowler KE, Krein SL, Flanders SA et al. (2013). An academic hospitalist model to improve healthcare worker communication and learner education: Results from

a quasi-experimental study at a veterans affairs medical center. *J Hosp Med* **8**; 12:702–10. DOI: dx.doi.org/10.1002/jhm.2105 (see pp. 81, 93–96, 98).

SALERNO SM, ARNETT MV, DOMANSKI JP (2009). Standardized sign-out reduces intern perception of medical errors on the general internal medicine ward. *Teach Learn Med* **21**; 2:121–6. DOI: 10.1080/10401330902791354 (see p. 61).

SAVITZ LA, JONES CB, BERNARD S (2005). Quality Indicators Sensitive to Nurse Staffing in Acute Care Settings. In: *Advances in Patient Safety: From Research to Implementation (Volume 4: Programs, Tools, and Products)*. Ed. by HENRIKSEN K, BATTLES JB, MARKS ES, LEWIN DI. Rockville, MD: Agency for Healthcare Research and Quality (see p. 133).

SCHAUFELI W, SALANOVA M, GONZÁLEZ-ROMÁ V, BAKKER A (2002). The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *J Happiness Stud* **3**; 1:71–92. DOI: 10.1023/A:1015630930326 (see p. 225).

SCHELLENBERG JA, BOBROVA N, AVAN BI (2012). Measuring implementation strength: literature review draft report. Working paper. London School of Hygiene and Tropical Medicine (see p. 79).

SCHILLIG J, KAATZ S, HUDSON M, KROL GD et al. (2011). Clinical and safety impact of an inpatient pharmacist-directed anticoagulation service. *J Hosp Med* **6**; 6:322–8. DOI: 10.1002/jhm.910 (see pp. 81, 83, 89, 95).

SCHMIDT E, GOLDHABER-FIEBERT SN, HO LA, MCDONALD KM (2013). Simulation exercises as a patient safety strategy: a systematic review. *Ann Intern Med* **158**; 5 Pt 2:426–32. DOI: 10.7326/0003-4819-158-5-201303051-00010 (see p. 61).

SCHUERER DJ, NAST PA, HARRIS CB, KRAUSS MJ et al. (2006). A new safety event reporting system improves physician reporting in the surgical intensive care unit. *J Am Coll Surg* **202**; 6:881–7. DOI: 10.1016/j.jamcollsurg.2006.02.035 (see p. 227).

SCHWARTZ D (1989). Visual ethnography: Using photography in qualitative research. *Qual Sociol* **12**; 2:119–54. DOI: 10.1007/bf00988995 (see p. 244).

Schwendimann R, Zimmermann N, Kung K, Ausserhofer D et al. (2013). Variation in safety culture dimensions within and between US and Swiss Hospital Units: an exploratory study. *BMJ Qual Saf* **22**; 1:32–41. doi: 10.1136/bmjqs-2011-000446 (see pp. 22, 23).

Scott IA, Brand CA, Phelps GE, Barker AL et al. (2011). Using hospital standardised mortality ratios to assess quality of care–proceed with extreme caution. *Med J Aust* **194**; 12:645–8 (see p. 25).

Scott I, Vaughan L, Bell D (2009). Effectiveness of acute medical units in hospitals: a systematic review. *Int J Qual Health Care* **21**; 6:397–407. doi: 10.1093/intqhc/mzp045 (see p. 62).

Seago JA, Williamson A, Atwood C (2006). Longitudinal analyses of nurse staffing and patient outcomes: more about failure to rescue. *J Nurs Adm* **36**; 1:13–21 (see p. 55).

Sexton JB, Helmreich RL, Neilands TB, Rowan K et al. (2006). The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* **6**;44. doi: 10.1186/1472-6963-6-44 (see pp. 131, 133, 159).

Shadish WR, Cook TD, Leviton LC (1991). Foundations of program evaluation: theories of practice. London: Sage (see p. 288).

Shah A, Carter T, Kuwani T, Sharpe R (2013). Simulation to develop tomorrow's medical registrar. *Clin Teach* **10**; 1:42–6. doi: 10.1111/j.1743-498X.2012.00598.x (see p. 61).

Shaller D, Kanouse DE, Schlesinger M (2014). Context-based strategies for engaging consumers with public reports about health care providers. *Med Care Res Rev* **71**; 5 Suppl:17S–37S. doi: 10.1177/1077558713493118 (see p. 242).

Shaw SE, Russell J, Parsons W, Greenhalgh T (2014). The view from nowhere? How think tanks work to shape health policy. *Critical Policy Studies* **9**; 1:58–77. doi: 10.1080/19460171.2014.964278 (see p. 156).

Sheard L, O'Hara J, Armitage G, Wright J et al. (2014). Evaluating the PRASE patient safety intervention - a multi-centre, cluster trial with a qualitative process evaluation: study

protocol for a randomised controlled trial. *Trials* **15**;420. DOI: 10.1186/1745-6215-15-420 (see p. 235).

SHEARER B, MARSHALL S, BUIST MD, FINNIGAN M et al. (2012). What stops hospital clinical staff from following protocols? An analysis of the incidence and factors behind the failure of bedside clinical staff to activate the rapid response system in a multi-campus Australian metropolitan healthcare service. *BMJ Qual Saf* **21**; 7:569–75. DOI: 10.1136/bmjqs-2011-000692 (see p. 21).

SHEKELLE PG (2013). Nurse-patient ratios as a patient safety strategy: a systematic review. *Ann Intern Med* **158**; 5 Pt 2:404–9. DOI: 10.7326/0003-4819-158-5-201303051-00007 (see p. 55).

SHEKELLE PG, PRONOVOST PJ, WACHTER RM, MCDONALD KM et al. (2013a). The top patient safety strategies that can be encouraged for adoption now. *Ann Intern Med* **158**; 5 Pt 2:365–8. DOI: 10.7326/0003-4819-158-5-201303051-00001 (see p. 227).

SHEKELLE PG, WACHTER RM, PRONOVOST PJ, SCHOELLES K et al. (2013b). Making health care safer II: an updated critical analysis of the evidence for patient safety practices. Rockville, MD: Agency for Healthcare Research and Quality (see p. 227).

SHIELL A, HAWE P, GOLD L (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ* **336**; 7656:1281–3. DOI: 10.1136/bmj.39569.510521.AD (see p. 287).

SHOJANIA KG (2008). The frustrating case of incident-reporting systems. *Qual Saf Health Care* **17**; 6:400–2. DOI: 10.1136/qshc.2008.029496 (see p. 111).

SHOJANIA KG (2013). Conventional evaluations of improvement interventions: more trials or just more tribulations? *BMJ Qual Saf* **22**; 11:881–4. DOI: 10.1136/bmjqs-2013-002377 (see pp. 59, 103).

SHOJANIA KG, THOMAS EJ (2013). Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* **22**; 4:273–7. DOI: 10.1136/bmjqs-2013-001935 (see p. 8).

Shojania KG, Forster AJ (2008). Hospital mortality: when failure is not a good measure of success. *CMAJ* **179**; 2:153–7. doi: 10.1503/cmaj.080010 (see p. 25).

Shojania KG, Ranji S, Shaw L, Charo L et al. (2004). Closing the quality gap: a critical analysis of quality improvement strategies (Vol. 2: Diabetes Care). Rockville, MD: Agency for Healthcare Research and Quality (see p. 11).

Shortell SM, Bennett CL, Byck GR (1998). Assessing the Impact of Continuous Quality Improvement on Clinical Practice: What It Will Take to Accelerate Progress. *Milbank Q* **76**; 4:593–624. doi: 10.1111/1468-0009.00107 (see p. 184).

Sikka R, Morath JM, Leape L (2015). The Quadruple Aim: care, health, cost and meaning in work. *BMJ Qual Saf* **24**; 10:608–10. doi: 10.1136/bmjqs-2015-004160 (see p. 7).

Simon D, Kriston L, Loh A, Spies C et al. (2010). Confirmatory factor analysis and recommendations for improvement of the Autonomy-Preference-Index (API). *Health Expect* **13**; 3:234–43. doi: 10.1111/j.1369-7625.2009.00584.x (see p. 244).

Sinaiko AD, Eastman D, Rosenthal MB (2012). How Report Cards On Physicians, Physician Groups, And Hospitals Can Have Greater Impact On Consumer Choices. *Health Aff (Millwood)* **31**; 3:602–11. doi: 10.1377/hlthaff.2011.1197 (see p. 242).

Singer SJ, Gaba DM, Falwell A, Lin S et al. (2009). Patient safety climate in 92 US hospitals: differences by work area and discipline. *Med Care* **47**; 1:23–31. doi: 10.1097/MLR.0b013e31817e189d (see pp. 9, 112).

Singh H, Thomas EJ, Petersen LA, Studdert DM (2007). Medical errors involving trainees: a study of closed malpractice claims from 5 insurers. *Arch Intern Med* **167**; 19:2030–6. doi: 10.1001/archinte.167.19.2030 (see p. 58).

Singh S, Tarima S, Rana V, Marks DS et al. (2012). Impact of localizing general medical teams to a single nursing unit. *J Hosp Med* **7**; 7:551–6. doi: 10.1002/jhm.1948 (see pp. 60, 81, 92, 94–96, 98).

Slaughter S, Hill J, Snelgrove-Clarke E (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: a scoping review. *Implement Sci* **10**; 1:129 (see p. 184).

Soifer NE, Borzak S, Edlin BR, Weinstein RA (1998). Prevention of peripheral venous catheter complications with an intravenous therapy team: a randomized controlled trial. *Arch Intern Med* **158**; 5:473–7 (see pp. 81, 83, 88, 95).

Solomon DH, Van Houten L, Glynn RJ, Baden L et al. (2001). Academic detailing to improve use of broad-spectrum antibiotics at an academic medical center. *Arch Intern Med* **161**; 15:1897–902 (see pp. 81, 83, 86, 94–96, 98).

Sorbero ME, Farley DO, Mattke S, Lovejoy S (2008). Outcome measures for effective teamwork in inpatient care: final report. Santa Monica, CA: RAND Corporation (see p. 293).

Spellberg B, Lewis RJ, Sue D, Chavoshan B et al. (2012). A controlled investigation of optimal internal medicine ward team structure at a teaching hospital. *PLoS One* **7**; 4:e35576. doi: 10.1371/journal.pone.0035576 (see p. 57).

Spencer SA, Davies MP (2012). Hospital episode statistics: improving the quality and value of hospital data: a national internet e-survey of hospital consultants. *BMJ Open* **2**; 6. doi: 10.1136/bmjopen-2012-001651 (see p. 291).

Spranca M, Minsk E, Baron J (1991). Omission and Commission in Judgment and Choice. *J Exp Soc Psychol* **27**; 1:76–105. doi: Doi10.1016/0022-1031(91)90011-T (see p. 19).

Spurgeon P, Mazelan PM, Barwell F (2011). Medical engagement: a crucial underpinning to organizational performance. *Health Serv Manage Res* **24**; 3:114–20. doi: 10.1258/hsmr.2011.011006 (see p. 225).

Starmer AJ, Spector ND, Srivastava R, West DC et al. (2014). Changes in medical errors after implementation of a handoff program. *N Engl J Med* **371**; 19:1803–12. doi: 10.1056/NEJMsa1405556 (see pp. 19, 218).

Stavropoulou C, Doherty C, Tosey P (2015). How Effective Are Incident-Reporting Systems for Improving Patient Safety? A Systematic Literature Review. *Milbank Q* **93**; 4:826–866. doi: 10.1111/1468-0009.12166 (see p. 111).

Stelfox HT, Palmisani S, Scurlock C, Orav EJ et al. (2006). The "To Err is Human" report and the patient safety literature. *Qual Saf Health Care* **15**; 3:174–8. doi: 10.1136/qshc.2006.017947 (see p. 6).

Stewart GL, Manges KA, Ward MM (2015). Empowering Sustained Patient Safety: The Benefits of Combining Top-down and Bottom-up Approaches. *J Nurs Care Qual* **30**; 3:240–6. doi: 10.1097/NCQ.0000000000000103 (see pp. 231–233).

Stickrath C, Noble M, Prochazka A, Anderson M et al. (2013). Attending rounds in the current era: what is and is not happening. *JAMA Intern Med* **173**; 12:1084–9. doi: 10.1001/jamainternmed.2013.6041 (see p. 57).

Stiefel M, Nolan K (2012). A Guide to Measuring the Triple Aim: Population Health, Experience of Care, and Per Capita Cost. IHI Innovation Series white paper. Institute for Healthcare Improvement (see p. 7).

Sujan MA, Ingram C, McConkey T, Cross S et al. (2011). Hassle in the dispensary: pilot study of a proactive risk monitoring tool for organisational learning based on narratives and staff perceptions. *BMJ Qual Saf* **20**; 6:549–56. doi: 10.1136/bmjqs.2010.048348 (see pp. 112, 227).

Svendsen ML, Ehlers LH, Andersen G, Johnsen SP (2009). Quality of care and length of hospital stay among patients with stroke. *Med Care* **47**; 5:575–82. doi: 10.1097/MLR.0b013e318195f852 (see p. 133).

Taitz JM, Lee TH, Sequist TD (2012). A framework for engaging physicians in quality and safety. *BMJ Qual Saf* **21**; 9:722–8. doi: 10.1136/bmjqs-2011-000167 (see p. 223).

Taylor SL, Dy S, Foy R, Hempel S et al. (2011). What context features might be important determinants of the effectiveness of patient safety practice interventions? *BMJ Qual Saf* **20**; 7:611–7. doi: 10.1136/bmjqs.2010.049379 (see pp. 223, 226, 239).

Thanarajasingam U, McDonald FS, Halvorsen AJ, Naessens JM et al. (2012). Service census caps and unit-based admissions: resident workload, conference attendance, duty hour compliance, and patient safety. *Mayo Clin Proc* **87**; 4:320–7. doi: 10.1016/j.mayocp.2011.12.012 (see pp. 81, 93–95, 98).

Thomas EJ (2015). The future of measuring patient safety: prospective clinical surveillance. *BMJ Qual Saf* **24**; 4:244–5. doi: 10.1136/bmjqs-2015-004078 (see p. 113).

Thylefors I, Persson O, Hellstrom D (2005). Team types, perceived efficiency and team climate in Swedish cross-professional teamwork. *J Interprof Care* **19**; 2:102–14. doi: 10.1080/13561820400024159 (see p. 14).

Timmermans MJ, Vught AJ van, Wensing M, Laurant MG (2014). The effectiveness of substitution of hospital ward care from medical doctors to physician assistants: a study protocol. *BMC Health Serv Res* **14**;43. doi: 10.1186/1472-6963-14-43 (see p. 131).

Tobiano G, Bucknall T, Marshall A, Guinane J et al. (2016). Patients' perceptions of participation in nursing care on medical wards. *Scand J Caring Sci* **30**; 2:260–70. doi: 10.1111/scs.12237 (see p. 245).

Tourangeau AE, Cranley LA, Jeffs L (2006). Impact of nursing on hospital patient mortality: a focused review and related policy implications. *Qual Saf Health Care* **15**; 1:4–8. doi: 10.1136/qshc.2005.014514 (see p. 56).

Tourangeau AE, Doran DM, McGillis Hall L, O'Brien Pallas L et al. (2007). Impact of hospital nursing care on 30-day mortality for acute medical patients. *J Adv Nurs* **57**; 1:32–44. doi: 10.1111/j.1365-2648.2006.04084.x (see pp. 55, 56).

Tourangeau AE, Giovannetti P, Tu JV, Wood M (2002). Nursing-related determinants of 30-day mortality for hospitalized patients. *Can J Nurs Res* **33**; 4:71–88 (see p. 56).

Treadwell JR, Lucas S, Tsou AY (2014). Surgical checklists: a systematic review of impacts and implementation. *BMJ Qual Saf* **23**; 4:299–318. doi: 10.1136/bmjqs-2012-001797 (see p. 226).

Trebble TM, Paul M, Hockey PM, Heyworth N et al. (2015). Clinically led performance management in secondary healthcare: evaluating the attitudes of medical and non-clinical managers. *BMJ Qual Saf* **24**; 3:212–20. doi: 10.1136/bmjqs-2014-003219 (see p. 231).

Tsai TC, Jha AK, Gawande AA, Huckman RS et al. (2015). Hospital Board And Management Practices Are Strongly Related To Hospital Performance On Clinical Quality Metrics. *Health Aff (Millwood)* **34**; 8:1304–11. doi: 10.1377/hlthaff.2014.1282 (see pp. 228, 231).

Tsai TC, Joynt KE, Orav EJ, Gawande AA et al. (2013). Variation in surgical-readmission rates and quality of hospital care. *N Engl J Med* **369**; 12:1134–42. doi: 10.1056/NEJMsa1303118 (see p. 26).

Tucker AL (2004). The impact of operational failures on hospital nurses and their patients. *Journal of Operations Management* **22**; 2:151–169. doi: 10.1016/j.jom.2003.12.006 (see pp. 8, 112, 125).

Tucker AL (2007). An empirical study of system improvement by frontline employees in hospital units. *Manufacturing & Service Operations Management* **9**; 4:492–505. doi: 10.1287/msom.1060.0156 (see p. 125).

Tucker AL, Edmondson AC (2003). Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *California Management Review* **45**; 2:55–72. doi: 10.2307/41166165 (see pp. 8, 111, 112, 125).

Tucker AL, Heisler WS, Janisse LD (2014). Designed for workarounds: a qualitative study of the causes of operational failures in hospitals. *Perm J* **18**; 3:33–41. doi: 10.7812/TPP/13-141 (see p. 112).

Tucker AL, Singer SJ, Hayes JE, Falwell A (2008). Front-line staff perspectives on opportunities for improving the safety and efficiency of hospital work systems. *Health Serv Res* **43**; 5 Pt 2:1807–29. doi: 10.1111/j.1475-6773.2008.00868.x (see pp. 9, 10, 112, 125, 153).

Tvedt C, Sjetne IS, Helgeland J, Bukholm G (2014). An observational study: associations between nurse-reported hospital characteristics and estimated 30-day survival probabilities. *BMJ Qual Saf* **23**; 9:757–64. doi: 10.1136/bmjqs-2013-002781 (see p. 111).

UK Clinical Research Collaboration (2015). UK Health Research Analysis 2014 (see p. 4).

Umscheid CA, Brennan PJ (2015). Incentivizing "structures" over "outcomes" to bridge the knowing-doing gap. *JAMA Intern Med* **175**; 3:354–5. doi: 10.1001/jamainternmed.2014.5293 (see p. 17).

University College London Hospitals (2016). Liberating sisters to lead: final report. London: Health Foundation (see p. 297).

University of Pennsylvania (2016). Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education. (iCOMPARE). https://clinicaltrials.gov/ct2/show/NCT02274818 (accessed 24/06/2016) (see p. 17).

Urbach DR, Govindarajan A, Saskin R, Wilton AS et al. (2014). Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med* **370**; 11:1029–38. doi: 10.1056/NEJMsa1308261 (see pp. 227, 232, 233).

Valentine MA, Nembhard IM, Edmondson AC (2015). Measuring teamwork in health care settings: a review of survey instruments. *Med Care* **53**; 4:e16–30. doi: 10.1097/MLR.0b013e31827feef6 (see p. 293).

Van Eaton EG, McDonough K, Lober WB, Johnson EA et al. (2010). Safety of using a computerized rounding and sign-out system to reduce resident duty hours. *Acad Med* **85**; 7:1189–95. doi: 10.1097/ACM.0b013e3181e0116f (see p. 61).

Vanhaecht K, Øvretveit J, Elliott MJ, Sermeus W et al. (2012). Have we drawn the wrong conclusions about the value of care pathways? Is a Cochrane review appropriate? *Eval Health Prof* **35**; 1:28–42. doi: 10.1177/0163278711408293 (see p. 63).

Vawdrey DK, Stein DM, Fred MR, Bostwick SB et al. (2013). Implementation of a computerized patient handoff application. *AMIA Annu Symp Proc* **2013**;1395–400 (see p. 61).

Vazirani S, Hays RD, Shapiro MF, Cowan M (2005). Effect of a multidisciplinary intervention on communication and collaboration among physicians and nurses. *Am J Crit Care* **14**; 1:71–7 (see p. 60).

Ven J van de (2012). 127: Effectiveness of multidisciplinary team training in obstetric emergencies: a randomized controlled trial. *American Journal of Obstetrics and Gynecology* **206**; 1:S68. doi: 10.1016/j.ajog.2011.10.145 (see p. 104).

Venkatesan C, Mishra A, Morgan A, Stepanova M et al. (2016). Outcomes trends for acute myocardial infarction, congestive heart failure, and pneumonia, 2005-2009. *Am J Manag Care* **22**; 1:e9–e17 (see p. 27).

Vidyarthi AR, Arora V, Schnipper JL, Wall SD et al. (2006). Managing discontinuity in academic medical centers: strategies for a safe and effective resident sign-out. *J Hosp Med* **1**; 4:257–66. doi: 10.1002/jhm.103 (see p. 61).

Vincent C, Amalberti R (2015). Safety in healthcare is a moving target. *BMJ Qual Saf* **24**; 9:539–40. doi: 10.1136/bmjqs-2015-004403 (see p. 8).

Vindrola-Padros C, Pape T, Utley M, Fulop NJ (2016). The role of embedded research in quality improvement: a narrative review. *BMJ Qual Saf*. doi: 10.1136/bmjqs-2015-004877 (see pp. 122, 157).

Vries EN de, Ramrattan MA, Smorenburg SM, Gouma DJ et al. (2008). The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* **17**; 3:216–23. doi: 10.1136/qshc.2007.023622 (see p. 5).

Wachter RM (2013). Personal accountability in healthcare: searching for the right balance. *BMJ Qual Saf* **22**; 2:176–80. doi: 10.1136/bmjqs-2012-001227 (see p. 6).

Wachter RM (17th Jan. 2016). How measurement fails doctors and teachers. *New York Times*:SR5 (see p. 29).

Wachter RM, Flanders SA, Fee C, Pronovost PJ (2008). Public reporting of antibiotic timing in patients with pneumonia: lessons from a flawed performance measure. *Ann Intern Med* **149**; 1:29–32 (see p. 28).

Wachter RM, Goldman L (1996). The emerging role of "hospitalists" in the American health care system. *N Engl J Med* **335**; 7:514–7. doi: 10.1056/NEJM199608153350713 (see p. 297).

Wachter R (2015). The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York: McGraw-Hill Education (see p. 13).

Wagner C, Merten H, Zwaan L, Lubberding S et al. (2016). Unit-based incident reporting and root cause analysis: variation at three hospital unit types. *BMJ Open* **6**; 6:e011277. doi: 10.1136/bmjopen-2016-011277 (see p. 18).

Wald HL, Glasheen JJ, Guerrasio J, Youngwerth JM et al. (2011). Evaluation of a hospitalist-run acute care for the elderly service. *J Hosp Med* **6**; 6:313–21. doi: 10.1002/jhm.906 (see pp. 81, 92, 94–96, 98).

Walshe K, Rundall TG (2001). Evidence-based management: from theory to practice in health care. *Milbank Q* **79**; 3:429–57, IV–V (see p. 229).

Wang Y, Eldridge N, Metersky ML, Verzier NR et al. (2014). National trends in patient safety for four common conditions, 2005-2011. *N Engl J Med* **370**; 4:341–51. doi: 10.1056/NEJMsa1300991 (see p. 27).

Watson TJ (2011). Ethnography, Reality, and Truth: The Vital Need for Studies of 'How Things Work' in Organizations and Management. *J Manage Stud* **48**; 1:202–17. doi: 10.1111/j.1467-6486.2010.00979.x (see p. 156).

Weaver SJ, Lubomksi LH, Wilson RF, Pfoh ER et al. (2013). Promoting a culture of safety as a patient safety strategy: a systematic review. *Ann Intern Med* **158**; 5 Pt 2:369–74. doi: 10.7326/0003-4819-158-5-201303051-00002 (see pp. 65, 71, 104).

Webster JR, Chew RB, Mailliard L, Moran MB (1999). Improving clinical and cost outcomes in delirium: Use of practice guidelines and a delirium care team. *Ann Longterm Care* **7**; 4:128–34 (see pp. 81, 83, 84, 94–96).

Weinberg DB, Miner DC, Rivlin L (2009). 'It depends': medical residents' perspectives on working with nurses. *Am J Nurs* **109**; 7:34–43, quiz 44. doi: 10.1097/01.NAJ.0000357167.63636.98 (see p. 59).

Weingart SN, Ship AN, Aronson MD (2000). Confidential clinician-reported surveillance of adverse events among medical inpatients. *J Gen Intern Med* **15**; 7:470–7 (see p. 226).

White M, Wells JSG, Butterworth T (2014). The impact of a large-scale quality improvement programme on work engagement: Preliminary results from a national cross-sectional-survey of the 'Productive Ward'. *Int J Nurs Stud* **51**; 12:1634–43. doi: http://dx.doi.org/10.1016/j.ijnurstu.2014.05.002 (see p. 9).

Wild D, Nawaz H, Chan W, Katz DL (2004). Effects of interdisciplinary rounds on length of stay in a telemetry unit. *J Public Health Manag Pract* **10**; 1:63–9 (see pp. 59, 139).

Wilkinson C, Khanji M, Cotter PE, Dunne O et al. (2008). Preferences of acutely ill patients for participation in medical decision-making. *Quality Saf Health Care* **17**; 2:97–100. doi: 10.1136/qshc.2006.021378 (see pp. 244, 273).

Wilkinson J, Powell A, Davies H (2011). Are clinicians engaged in quality improvement? A review of the literature on healthcare professionals' views on quality improvement initiatives. London: Health Foundation (see pp. 225–227, 239).

Wilson CB (2006). Adoption of new surgical technology. *BMJ* **332**; 7533:112–4. doi: 10.1136/bmj.332.7533.112 (see p. 128).

Wilson SJ, Wong D, Clifton D, Fleming S et al. (2013). Track and trigger in an emergency department: an observational evaluation study. *Emerg Med J* **30**; 3:186–91. doi: 10.1136/emermed-2011-200499 (see p. 64).

Winters BD, Weaver SJ, Pfoh ER, Yang T et al. (2013). Rapid-response systems as a patient safety strategy: a systematic review. *Ann Intern Med* **158**; 5 Pt 2:417–25. doi: 10.7326/0003-4819-158-5-201303051-00009 (see p. 65).

Womack JP, Jones DT, Roos D (1991). The machine that changed the world: how Japan's secret weapon in the global auto wars will revolutionize western industry. New York, NY: HarperPerennial (see p. 9).

Wong BM, Dyal S, Etchells EE, Knowles S et al. (2015). Application of a trigger tool in near real time to inform quality improvement activities: a prospective study in a general medicine ward. *BMJ Qual Saf* **24**; 4:272–81. doi: 10.1136/bmjqs-2014-003432 (see pp. 19, 112, 217, 290).

World Health Organisation (2014a). Patient safety. http://www.who.int/features/factfiles/patient_safety/en/ (accessed 27/03/2016) (see p. 5).

World Health Organisation (2014b). WHO Surgical Safety Checklist. http://www.who.int/patientsafety/safesurgery/checklist/en/ (accessed 21/01/2016) (see pp. 117, 128).

World Health Organisation (2016). WHO Safe Childbirth Checklist. http://www.who.int/patientsafety/implementation/checklists/childbirth/en/ (accessed 18/07/2016) (see p. 128).

Wray CM, Farnan JM, Arora VM, Meltzer DO (2016). A qualitative analysis of patients' experience with hospitalist service handovers. *J Hosp Med*. doi: 10.1002/jhm.2608 (see p. 274).

Wright S, McSherry W (2013). A systematic literature review of Releasing Time to Care: The Productive Ward. *J Clin Nurs* **22**; 9-10:1361–71. doi: 10.1111/jocn.12074 (see p. 9).

Wright S, McSherry W (2014). Evaluating the Productive Ward at an acute NHS trust: experiences and implications of releasing time to care. *J Clin Nurs* **23**; 13-14:1866–76. doi: 10.1111/jocn.12435 (see p. 9).

Zheng S, Tucker AL, Ren Z, Heineke J et al. (2016). The Impact of Internal Service Quality on Nurse Inefficiency and Medical Errors. http://ssrn.com/abstract=2784503 (accessed 11/07/2016) (see p. 297).

Zwarenstein M, Goldman J, Reeves S (2009). Interprofessional collaboration: effects of practice-based interventions on professional practice and healthcare outcomes. *Cochrane*

Database Syst Rev 3:CD000072. DOI: 10.1002/14651858.CD000072.pub2 (see pp. 59, 66, 71, 103, 104).

Appendices

Appendix A

Systematic review: search terms

A.1 Objective outcomes

(adverse event$ or adverse drug or avoidable delay$ or avoidable harm$ or bacteraemia or bacteremia or bleeding or cardiac arrest$ or care deficienc$ or catheter-associated urine tract infection$ or catheter-associated uti$ or clostridium difficile or critical incident$ or critical care admission$ or critical outcome$ or day$ in hospital or death$ or delay$ or deteriorat$ or diagnos$ error$ or disabilit$ on discharge or deep vein thromb$ or delirium or difficile or drug error$ or dvt$ or error$ of omission or error$ of commission or error prevention or escalation or failure to rescue or falls or fatalit$ or hand hygiene or handwashing or hospital-acquired infection$ or hospital-associated infection$ or healthcare-acquired infection$ or healthcare-associated infection$ or healthcare failure$ or healthcare safety or healthcare outcome$ or hospital-acquired pneumonia$ or hospital day$ or hospital failure$ or hospital outcome$ or hospital safety or HSMR or hypotension or hypoxia or iatrogenic or inappropriate care or infection control or injury or injuries or intensive care admission$ or lapse$ or length$ of stay or medical complication$ or medical error$ or medical failure$ or medical outcome$ or medication error$ or methicillin-resistant staphylococc$ or mrsa or misdiagnos$ or misidentification$ or mistake$ or morbidity or mortality or move$ or near miss$ or negligen$ or never event$ or nosocomial

or patient deterioration or patient outcome$ or patient safety or performance indicator$ or preventable error$ or pulmonary emboli$ or quality indicator$ or quality of care or readmission$ or resuscitation or safe care or safety attitude$ or safety climate$ or safety culture$ or safety practice$ or sentinel event$ or sepsis or serious untoward incident$ or service deficienc$ or SHMI or slip$ or standard$ of care or suboptimal care or substandard care or thromboemboli$ or thromboprophylaxis or transfer$ or undesirable consequence$ or undesirable event$ or unintended injur$ or unintended harm$ or unintended consequence$ or unanticipated harm$ or unanticipated injur$ or unexpected event$ or unexpected injur$ or unexpected consequence$ or unexpected harm$ or unnecessary event$ or unnecessary injur$ or unnecessary consequence$ or unnecessary harm$ or waiting time$ or ward safety or ward performance).ab,ti.

A.2 Program evaluations and trials

(bar cod$ or change or checklist$ or computer-assisted or decision making or decision support or endpoint or evaluation or expert system or feedback$ or impact or implement$ or initiative$ or intervention*1 or medical informatics or outcome measure$ or order entry or program$ or reminder$ or trial).ab,ti.

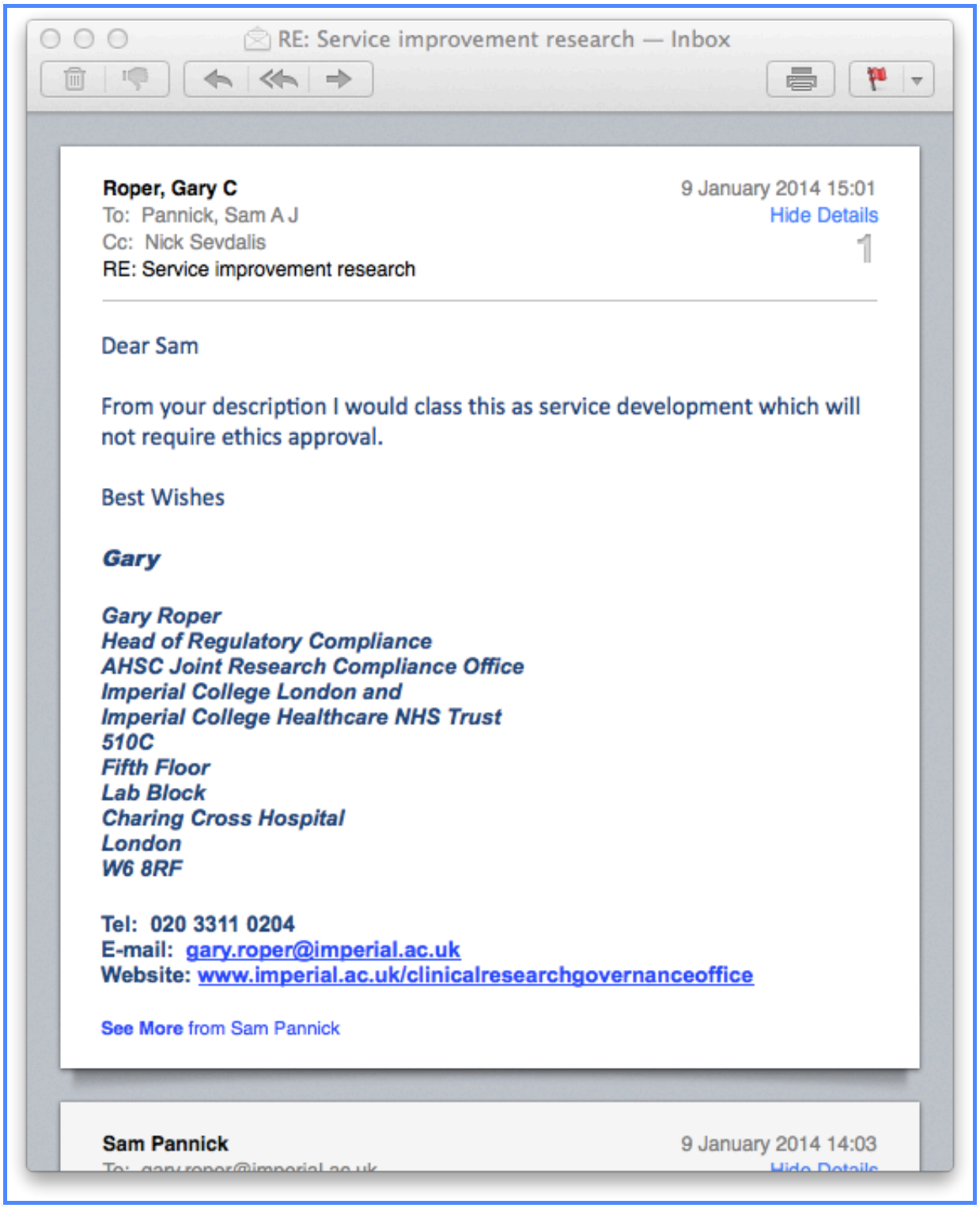
A.3 Medical ward descriptors

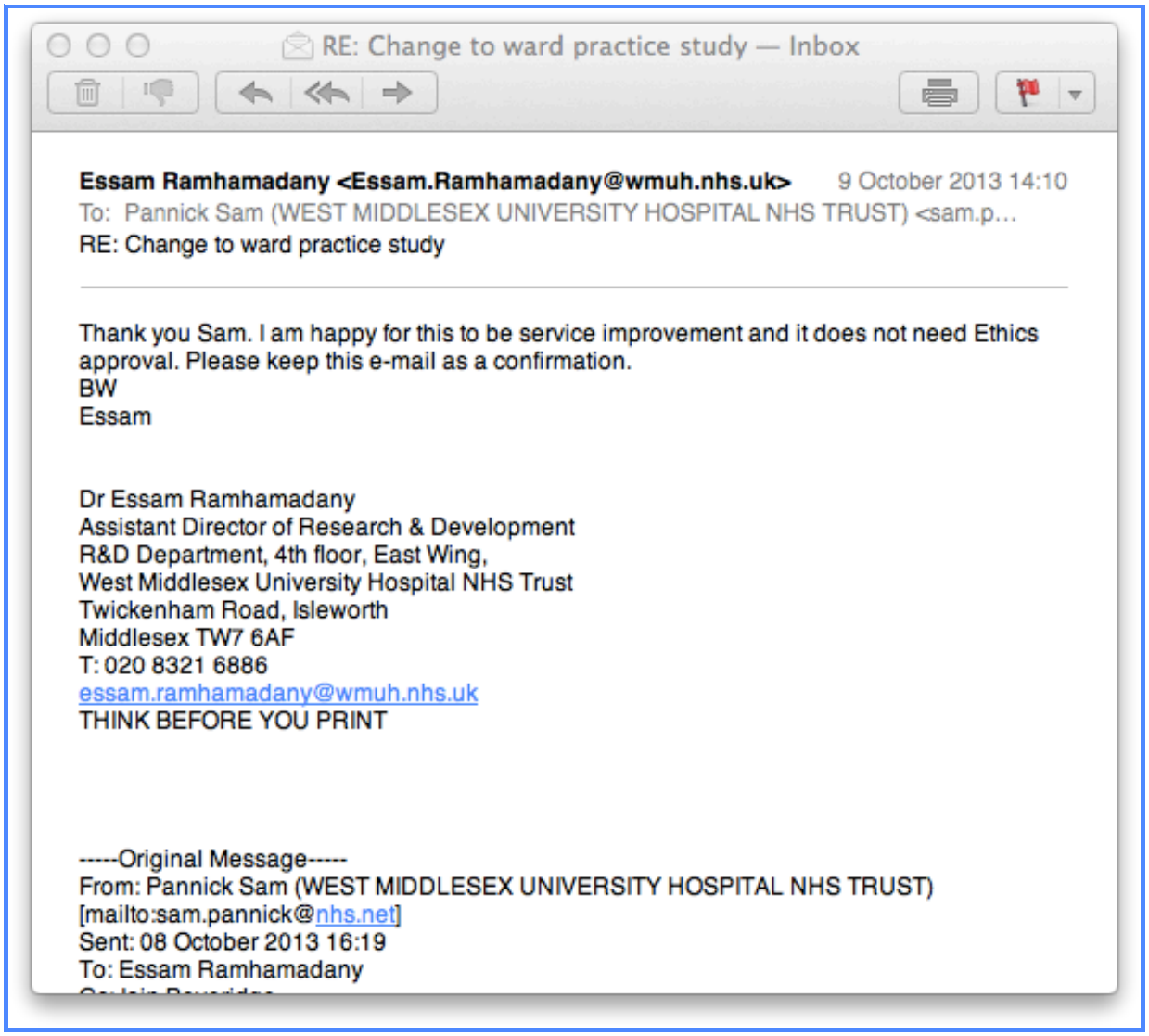
(academic medical unit$ or academic teaching unit$ or acute admission$ unit$ or acute assessment unit$ or acute disease management or acute medical unit$ or acute medicine or general medicine or general medical or general medical ward$ or general medical department$ or general medical service$ or general ward$ or hospital inpatient$ or hospital$ medicine or hospital$ unit$ or hospitalist$ or inpatient medical service$ or inpatient medical team$ or inpatient medical unit$ or inpatient medicine or inpatient resident service$ or inpatient teaching service$ or internal medical service$ or internal medicine or medical admission$ or medical assessment unit$ or medical environment$ or medical inpatient$ or medical patient$ or medical specialt$

or medical teaching unit$ or medical unit$ or medical ward$ or resident teaching service$ or telemetry ward$).ab,ti.

Appendix B

HEADS-UP: research and development authority approvals


RE: Service improvement research — Inbox
Roper, Gary C
9 January 2014 15:01
To: Pannick, Sam A J
Hide Details
Cc: Nick Sevdalis
1
RE: Service improvement research
Dear Sam
From your description I would class this as service development which will not require ethics approval.
Best Wishes
Gary
Gary Roper
Head of Regulatory Compliance
AHSC Joint Research Compliance Office
Imperial College London and
Imperial College Healthcare NHS Trust
510C
Fifth Floor
Lab Block
Charing Cross Hospital
London
W6 8RF
Tel: 020 3311 0204
E-mail: gary.roper@imperial.ac.uk
Website: www.imperial.ac.uk/clinicalresearchgovernanceoffice
See More from Sam Pannick
Sam Pannick
9 January 2014 14:03


RE: Change to ward practice study — Inbox

Essam Ramhamadany <Essam.Ramhamadany@wmuh.nhs.uk> 9 October 2013 14:10
To: Pannick Sam (WEST MIDDLESEX UNIVERSITY HOSPITAL NHS TRUST) <sam.p…
RE: Change to ward practice study

Thank you Sam. I am happy for this to be service improvement and it does not need Ethics approval. Please keep this e-mail as a confirmation.
BW
Essam

Dr Essam Ramhamadany
Assistant Director of Research & Development
R&D Department, 4th floor, East Wing,
West Middlesex University Hospital NHS Trust
Twickenham Road, Isleworth
Middlesex TW7 6AF
T: 020 8321 6886
essam.ramhamadany@wmuh.nhs.uk
THINK BEFORE YOU PRINT

-----Original Message-----
From: Pannick Sam (WEST MIDDLESEX UNIVERSITY HOSPITAL NHS TRUST) [mailto:sam.pannick@nhs.net]
Sent: 08 October 2013 16:19
To: Essam Ramhamadany

Appendix C

HEADS-UP: PRISMA diagram and supplementary results from generalised estimating equations models

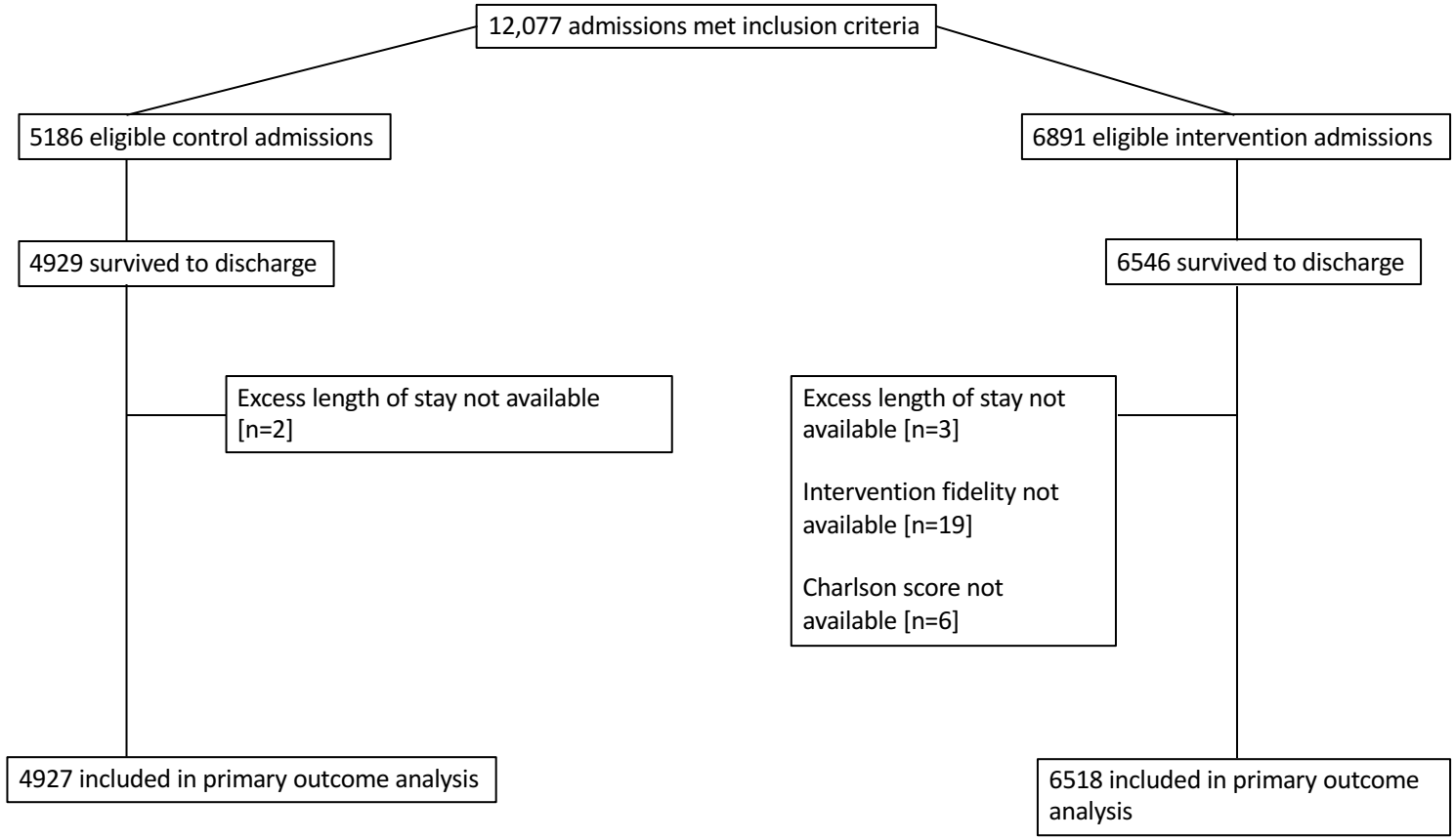


Figure C.1: Patient flow diagram for the HEADS-UP study

Table C.1: Intention-to-treat and per protocol analyses for patient outcomes and processes of care (generalised estimating equations models)

Outcome	HEADS-UP	Control	Adjusted outcome ratio*					
			Intention-to-treat		Moderate fidelity**		High fidelity***	
			Outcome ratio (95% CI)	P value	Outcome ratio (95% CI)	P value	Outcome ratio (95% CI)	P value
Primary outcome								
eLOS [n (%)]	1728 (26.5)	1279 (26.0)	**1.34 (1.11-1.62)**	**0.003**	0.92 (0.75-1.14)	0.464	**0.72 (0.62-0.84)**	**<0.001**
Secondary outcomes								
Readmission within 30 days [n (%)]	950 (14.5)	565 (11.5)	1.10 (0.87-1.38)	0.433	1.00 (0.79-1.27)	0.972	1.10 (0.92-1.32)	0.279
In-hospital death or death/readmission within 30 days [n (%)]	1662 (24.1)	1190 (22.9)	0.99 (0.82-1.19)	0.921	1.08 (0.89-1.30)	0.459	1.13 (0.97-1.31)	0.117
Escalation events/month**** [median (IQR)]	8 (3.5-16)	3 (1-6)	**0.78 (0.69-0.89)**	**<0.001**	**1.38 (1.21-1.57)**	**<0.001**	1.10 (0.997-1.20)	0.059
Complications of care/month [median (IQR)]	1 (0-2)	1 (0-3)	0.86 (0.55-1.36)	0.527	1.20 (0.67-2.12)	0.538	1.27 (0.82-1.96)	0.290

*Odds ratios for binary outcomes, and incidence rate ratios for continuous outcomes. **Medium frequency briefings with high engagement. ***High frequency briefings with high engagement. ****Emergency calls, referrals to the ICU outreach service, and ICU transfers.

Appendix D

HEADS-UP: focus group topic guide and survey instrument

Topic	Question number	Questions	Probes (if group not elaborating answers to main questions)
Usual systems for identifying team concerns and improving patient care	1.	There are different ways to flag up problems you see in your work (for example: informal conversations with senior staff, ward meetings, incident reports, morbidity and mortality meetings). How well do these different options identify the day-to-day issues that affect patient care?	Do you think these systems give an accurate picture of what it's like working on your ward? If not, why not?
	2.	Once an issue has been raised, what increases the chances of it getting resolved?	For example: - a patient was harmed - a near miss - senior staff reported the issue themselves or were quickly made aware of it - the issue is long-standing - staff can't find a work-around or temporary fix
The impact of HEADS-UP	3.	Did HEADS-UP change what you thought of your ward's safety and quality? How?	
	4.	What was it like discussing problems in the HEADS-UP briefing?	Was it easier to raise your concerns? Did you use it as a way of bringing issues to the attention of more senior members of the team? Was there more of an expectation that problems should be resolved? Did HEADS-UP make it easier to challenge other members of the team (for example, to discuss a decision or accept responsibility for resolving an issue)?
	5.	Did your team make any changes to how things were done on the ward as a result of HEADS-UP?	What changes did you make?
	6.	Can you think of a specific HEADS-UP briefing that you felt had been particularly useful for you?	Why was that? Did someone share information that you may not have received otherwise? Did it help you to get an answer from a senior colleague?

			Did it prompt the team to resolve an issue?
HEADS-UP barriers and facilitators	7.	How did your team react to being asked to use HEADS-UP?	Is your team open to trying new things in general? If not, why do you think that is? Was there anyone who was particularly enthusiastic, or particularly negative, about HEADS-UP on your ward? What was their role? Did that affect how other people saw HEADS-UP?
	8.	Who 'owned' HEADS-UP on your ward (nurses, doctors, or everyone)? Did that make a difference?	How did that affect how the rest of the team saw HEADS-UP, and how it worked on your ward? Were senior staff actively participating in HEADS-UP or did they "let it happen"?
	9.	How effectively did your team, or the department of medicine, address the issues raised in HEADS-UP?	
	10.	If there were changes you'd like to have made after raising issues in HEADS-UP: - What helped you to implement them? - What stopped the team going ahead with them?	Did you need additional time, or other resources, that weren't available?
	11.	Do you think your team needed HEADS-UP?	What do you think of your team's approach to quality and safety compared to other teams you've worked with? Does that affect how much value HEADS-UP could add?
	12.	Is there space and time to introduce new programmes like HEADS-UP?	
	14.	Do the most senior people in the organisation understand the challenges in your day to day work?	
	15.	Did you hear about any changes HEADS-UP brought about? How did you hear about them?	e.g. - feedback from ward managers - posters - hospital newsletter - emails - feedback from divisional quality & risk meetings - departmental presentation

	16.	What would you change in the structure or content of HEADS-UP?	
	17.	What do teams like yours need to introduce a programme like HEADS-UP successfully?	Protected time? Training? Better links between safety / governance committees and ward teams?

Optional questions, if time allows:	18.	What affected how well each HEADS-UP briefing went?	What made it easier? What made it harder? e.g. - time constraints - the job of the person leading it (doctor / nurse) - more than one team on the ward needing to be present - general attitudes to whether the problems could be fixed
	19.	What kind of issues did you discuss in HEADS-UP that you wouldn't normally raise as concerns?	Why do you think the team identified these issues in HEADS-UP when you wouldn't normally raise them as concerns? Are the issues raised in HEADS-UP important?
	20.	What typically happened at the end of the HEADS-UP briefings?	How did you decide at the end of the briefings what issues needed to be followed up? Did everyone on the team agree what needed to be followed up? How did you assign a person to follow up the issue? How successful were they in doing what the team had agreed?

UTHealth
The University of Texas
Health Science Center at Houston

Medical School
University of Texas at Houston-Memorial Hermann
Center for Healthcare Quality and Safety

October 22, 2015

Dear Sam Pannick,

You have our permission to use any of the following Safety Attitudes Questionnaires and the corresponding scoring keys:

Safety Attitudes Questionnaire – Short Form
Safety Attitudes Questionnaire – Teamwork and Safety Climate
Safety Attitudes Questionnaire – Ambulatory Version
Safety Attitudes Questionnaire – ICU Version
Safety Attitudes Questionnaire – Labor and Delivery Version
Safety Attitudes Questionnaire – Operating Room Version
Safety Attitudes Questionnaire – Pharmacy Version
Safety Climate Survey

Please note, we do not have editable versions for any of the SAQ surveys but feel free to modify the surveys to meet your research endeavors.

Respectfully,

University of Texas at Houston-Memorial Hermann
Center for Healthcare Quality and Safety Team

6410 Fannin Street
UTPB Suite 1100
Houston, TX 77030
https://med.uth.edu/chqs/

DOME safety survey 2015

YOUR WARD: .....

Imperial College Healthcare Charity

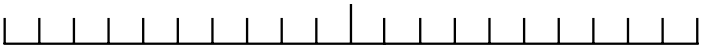
PLEASE CIRCLE YOUR ROLE: **F1 / F2 / ST 1-2 / ST 3 - 8 / CONSULTANT**
ADMIN / STAFF NURSE / SENIOR NURSE / PHYSIO / OT / PHARMACIST

SAFETY CLIMATE ON THIS WARD:	**Disagree strongly**	**Disagree slightly**	**Neutral**	**Agree slightly**	**Agree strongly**	**Not applicable**
The levels of staffing on this ward are sufficient to handle the number of patients	❑	❑	❑	❑	❑	❑
I would feel safe being treated here as a patient	❑	❑	❑	❑	❑	❑
I am encouraged by my colleagues to report any patient safety concerns I may have	❑	❑	❑	❑	❑	❑
Staff frequently disregard rules or guidelines (e.g. hand-washing, treatment protocols / clinical pathways, sterile field, etc.) established for the wards	❑	❑	❑	❑	❑	❑
The culture on this ward makes it easy to learn from the errors of others	❑	❑	❑	❑	❑	❑
I receive appropriate feedback about my performance	❑	❑	❑	❑	❑	❑
Medical errors are handled appropriately here	❑	❑	❑	❑	❑	❑
I know the proper channels to direct questions regarding patient safety on this ward	❑	❑	❑	❑	❑	❑
On this ward, it is difficult to discuss errors	❑	❑	❑	❑	❑	❑
Hospital management does not knowingly compromise the safety of patients	❑	❑	❑	❑	❑	❑
This institution is doing more for patient safety now than it did one year ago	❑	❑	❑	❑	❑	❑
Leadership is driving us to be a safety-centred institution	❑	❑	❑	❑	❑	❑
My suggestions about safety would be acted upon if I expressed them to management	❑	❑	❑	❑	❑	❑
TEAMWORK CLIMATE ON THIS WARD:	**Disagree strongly**	**Disagree slightly**	**Neutral**	**Agree slightly**	**Agree strongly**	**Not applicable**
Nurse input is well received on this ward	❑	❑	❑	❑	❑	❑
On this ward, it is difficult to speak up if I perceive a problem with patient care	❑	❑	❑	❑	❑	❑
Decision-making on this ward utilises input from relevant staff	❑	❑	❑	❑	❑	❑
The doctors and nurses here work together as a well-coordinated team	❑	❑	❑	❑	❑	❑
Disagreements on this ward are resolved appropriately (i.e. not *who* is right, but *what* is best for the patient)	❑	❑	❑	❑	❑	❑
I am frequently unable to express disagreement with senior staff here (e.g. consultants / senior nurses)	❑	❑	❑	❑	❑	❑
It is easy for staff here to ask questions when there is something that they do not understand	❑	❑	❑	❑	❑	❑
I have the support I need from other staff to care for patients	❑	❑	❑	❑	❑	❑
I know the first and last names of all the staff I worked with during my last shift	❑	❑	❑	❑	❑	❑
Important issues are well communicated at shift changes (i.e. handovers)	❑	❑	❑	❑	❑	❑
Briefing staff before the start of a shift (i.e. to plan for possible contingencies) is important for patient safety	❑	❑	❑	❑	❑	❑
Briefings are common on this ward	❑	❑	❑	❑	❑	❑
I am satisfied with the quality of collaboration that I experience with doctors on this ward	❑	❑	❑	❑	❑	❑
I am satisfied with the quality of collaboration that I experience with nurses on this ward	❑	❑	❑	❑	❑	❑

PLEASE TURN OVER

PLEASE NOW THINK ABOUT THE LAST DAY YOU WORKED ON THIS WARD:
PLACE AN **X** IN THE APPROPRIATE SPACE

How mentally demanding was it?



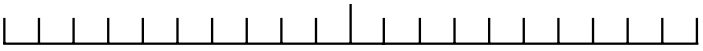
Not demanding at all — Very demanding

How physically demanding was it?



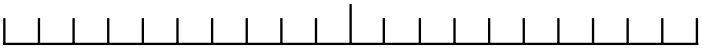
Not demanding at all — Very demanding

How hurried or rushed was it?



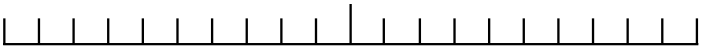
Not rushed at all — Very rushed

How successful were you in accomplishing what you needed to do?



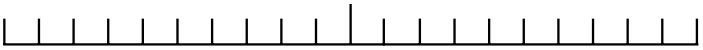
Total failure — Perfect performance

How hard did you have to work to accomplish what you needed to do?



Not hard at all — Very hard

How insecure, discouraged, irritated, stressed, and annoyed were you?



Not at all — Very much

- How long have you been working on this ward? **0-3 months** **4-6 months** **6-12 months** **> 12 months**
- How many times to date have you participated in HEADS-UP morning briefings? **0** **< 5** **5-10** **> 10**
- Have you reported any concerns differently as a result of participating in the HEADS-UP briefings? **Yes** **No** **Not sure** **Not applicable**

Any other comments?

……

……

THANK YOU FOR YOUR TIME. YOUR ANSWERS WILL BE TREATED IN THE STRICTEST CONFIDENCE.

Appendix E

HEADS-UP: a mobile application to improve automated data collection



What happened on the last shift?

Complete a new review

Complete a saved review

See a completed review

About

Help

Imperial College Healthcare Charity

Exit

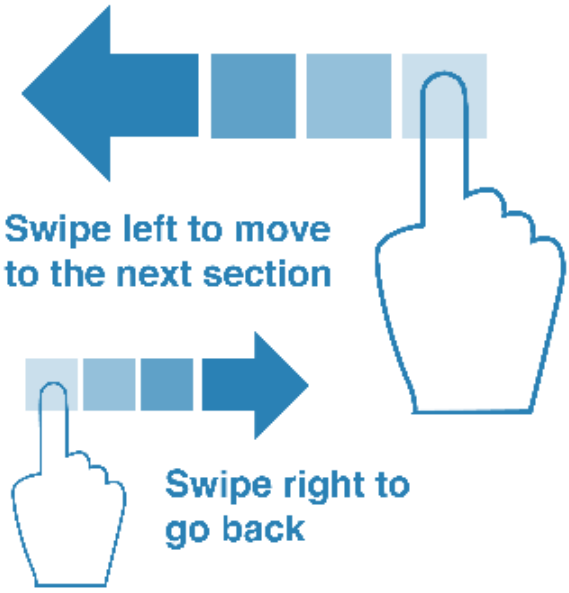
New review

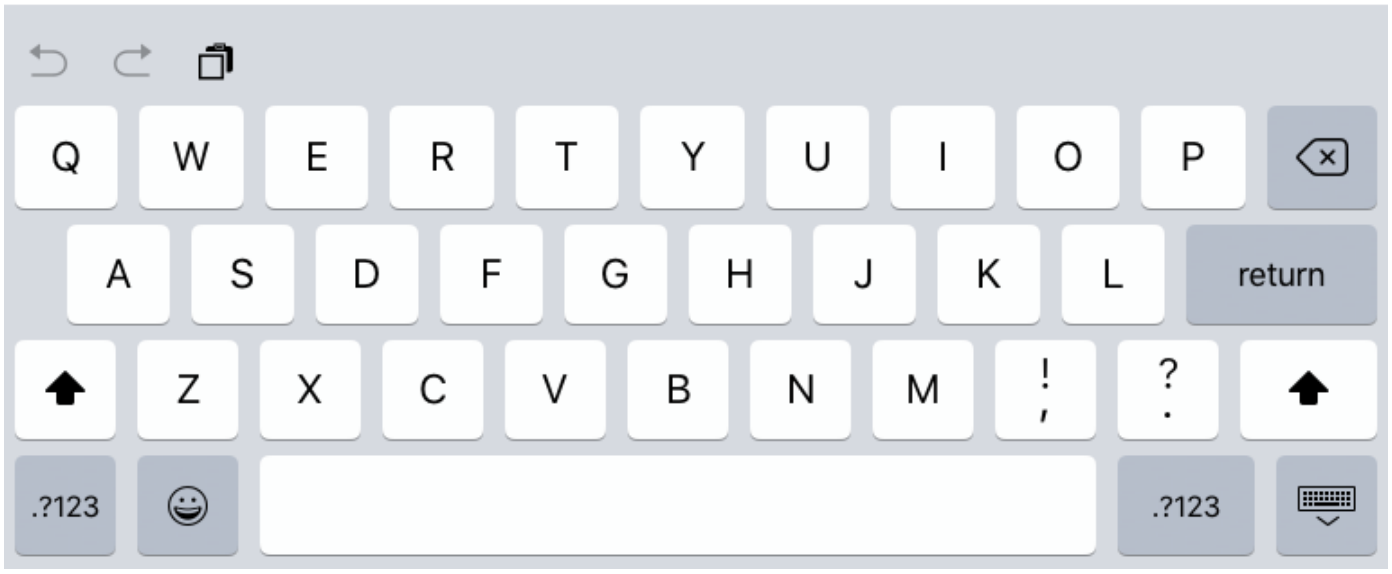
1. Review details
2. Team members present
3. Concerns
4. Events
5. Drug errors and other issues
6. What went really well
7. Who will...

8. Submit (server not active)

1. Review details

What is the name of the ward?





3. Concerns

Were there concerns with staffing levels?

Yes | No

3. Concerns

Were there concerns with services?

Yes | No

- [ ] Phlebotomy
- [ ] Echo
- [ ] Radiology
- [ ] Endoscopy

Please add details...

No SIM 15:32 94%

6. What went really well

Things that went well:

- Communication and teamwork ☑
- Positive feedback from patient or family ☑
- Managing a deteriorating patient ☐

Appendix F

Patient-prioritised quality metrics: interview topic guide, patient information sheets and ethical approval

TOPIC GUIDE: Current patient / carer for current patient

1. Let's find out about you.

Age
Employment
Education – school / university / postgraduate
Social support structures & marital status
Ethnicity

2. Why are you in hospital now?

Current diagnosis
Other conditions
Approximate length of stay to date
- < 1 day
- 1 – 5 days
- 5 – 10 days
- > 10 days

3. How many times have you been admitted to hospital in the last 6 months?

1-5
5-10
>10

4. Do you always come to this hospital or have you been admitted to other local hospitals?

5. How do you know if you're on a good ward? What is a 'good ward' to you?

Environment
- Clean
- Quiet
- Toilet and shower are available when required
- Meal timeliness, warmth
- Help available when requested
- Staff are responsiveness to my needs / my family's needs

Welcome
- My arrival is expected
- Staff introduce themselves
- Staff make me feel I will be well looked after; show a caring attitude; and don't rush me

Communication and use of personal information
- Accurate knowledge of previous medical history / current diagnosis / current investigations / discharge plan / medication reconciliation
- Quality of communication / teamwork

Discharge preparation

Friends' / families' recommendations

Ward information boards / quality and safety boards

Ward information leaflets / other printed materials.

Ward information displays / electronic screens

6. **If you had to decide whether a ward was good or not, what information would you need to make that decision?**

7. **Have you noticed any of the information the ward displays about itself? What do you think of the information you've seen?**

Friends and family test results
Safety cross
Shift-by-shift staffing
Falls
Pressure ulcers
Safety thermometer / harm-free care
Venous thromboembolism prophylaxis
Hand hygiene compliance
Hospital-acquired infections
Infection rates
Incident reporting

8. **What would you like to know about how your ward is performing?**

Hand hygiene compliance
Staffing levels
Friends and family results
Hospital-acquired infections
Pressure Ulcers
Falls
Venous thromboembolism prophylaxis
Complaints
Compliments
Length of stay
Mortality
Readmission rate
Safety climate

9. **How should your ward make that information available to you and your family??**

Ward displays
Leaflets
Smartphone / other device
Webpage

10. Preference for information seeking

Information-seeking sub-scale

	Disagree strongly	Disagree slightly	Neutral	Agree slightly	Agree strongly
As you become sicker you should be told more and more about your illness					
You should understand completely what is happening inside your body as a result of your illness					
Even if the news is bad, you should be well informed					
Your doctor should explain the purpose of your laboratory tests					
It is important for you to know all the side effects of your medication					
Information about your illness is as important to you as treatment					
When there is more than one method to treat a problem, you should be told about each one					

11. Have you previously had to complain about care or healthcare staff, nurses or doctors? What made you complain? How? PALS / informally / in writing?

Imperial College London

Imperial College Healthcare NHS Trust

Patient / Carer Information Sheet

Patient and carer information needs on the medical ward: an interview study to explore perceptions of ward performance

Researcher: Dr Samuel Pannick
Supervisors: Professor Nick Sevdalis, Professor Thanos Athanasiou, Dr Susannah Long

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

About the study

Wards are encouraged to display more and more information about their performance, so that staff and patients can see how well they are working. We don't know whether this information is useful for patients and their carers, how they feel about it, or what would make it more useful. This study is part of a PhD (doctorate) on medical ward care.

Why have I been asked to take part?

You have been asked because you are currently an inpatient on this ward, or you are the main carer for a current patient on this ward. Staff on this ward hoped you might feel well enough to help us with this study today.

Do I have to take part?

No. If you do decide to take part you will be given this information sheet to keep and will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This will not affect your care in any way.

What will happen if I take part?

You will be asked some questions about what makes a good ward, the information we display on the ward, and how you might use that information. The discussion will be taped and we might also take some notes. The interview will last approximately 30 minutes. We may also interview your main carer, if you have one, if they are also happy to tell us how they use ward information.

What are the possible benefits of taking part?

We will try to use this study to deliver ward information to patients and carers in ways that they find more helpful.

What are the possible downsides of taking part?

We do not anticipate any major downsides to taking part.

Is taking part anonymous and confidential?
Yes. All information gathered will be treated as confidential. If we quote your opinions at any point it will be done anonymously. No personally identifiable information will be used if we publish this research.

What will happen to the results of the research study?

The findings will be published in a scientific journal. If you would like to be notified about the publication date and journal, please let the interviewer know, and he will be able to provide further information (see below).

Who is organising and funding the research?

The NIHR Imperial Patient Safety Translational Research Centre and the Imperial College Healthcare Charity are organising and funding this research.

Who has reviewed the study?

The study has been reviewed and approved by Westminster Research Ethics Committee.

Who can answer my questions about this study?
You can talk to the researcher about any questions or concerns you have about this study:

Dr Sam Pannick
Email: s.pannick@imperial.ac.uk
Phone: 0207 594 3149

What if something goes wrong?
If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study then you should immediately inform the Investigator (Dr Sam Pannick, as above). The normal National Health Service complaints mechanisms are also available to you. If you are still not satisfied with the response, you may contact the Imperial AHSC Joint Research Compliance Office.

Imperial College London

Imperial College Healthcare NHS Trust

Appendix 3. Consent form, v2.1 (19th January 2016)

CONSENT FORM

"Patient and carer information needs on the medical ward: an interview study to explore perceptions of ward performance"

Participant ID:

Please put your initials in each box

1. I confirm that I have read and understand the information sheet (V2, 07/01/16) for this study and have had the opportunity to ask questions which have been answered fully. ☐

2. I understand that my participation is voluntary and I am free to withdraw at any time, without giving any reason, without my medical or legal rights being affected. ☐

3. I understand that the interview will be recorded, but my views will be kept confidential and only shared anonymously for research purposes. I understand that my personal details will not be shared with anyone outside of the research team. ☐

4. I give permission for the researchers to anonymously reflect my views in scientific publications relevant to this research. ☐

5. I agree to take part in this study. ☐

_____	_____	_____
Your Name	Signature	Date

_____	_____	_____
Researcher	Researcher Signature	Date

Health Research Authority

London - Westminster Research Ethics Committee
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0207 104 8012

26 January 2016

Dr Samuel Pannick
Clinical Research Fellow
Imperial College London
Room 503
Medical School Building
Norfolk Place
W2 1PG

Dear Dr Pannick

Study title:	**Patient and carer information needs on the medical ward: an interview study to explore perceptions of ward performance**
REC reference:	**16/LO/0196**
IRAS project ID:	**198908**

Thank you for your submission of 20 January 2016, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Ms Rachel Katzenellenbogen, nrescommittee.london-westminster@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Poster]	1	07 January 2016
Covering letter on headed paper [Cover letter]	1	04 January 2016
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Verification of insurance]		13 July 2014
Interview schedules or topic guides for participants [Topic guide - appendix 1 of protocol]	3	07 January 2016
Letter from funder [Grant letter]		11 December 2014
Letter from sponsor [Sponsorship confirmation]		08 January 2016
Other [Regarding study registration]	1	20 January 2016
Other [Study protocol and patient information sheet with tracked changes]	3.1	19 January 2016
Participant consent form [Participant consent form]	2.1	19 January 2016
Participant information sheet (PIS) [Participant information sheet]	2	07 January 2016
REC Application Form [REC_Form_11012016]		11 January 2016
Summary CV for Chief Investigator (CI) [Pannick CV]	1	08 January 2016
Summary CV for supervisor (student research) [Athanasiou CV]	1	08 January 2016
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Synopsis]	1	08 January 2016

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

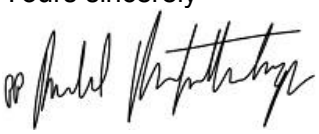
You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

16/LO/0196	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project.

Yours sincerely



Dr Alan Ruben
Chair

Email: nrescommittee.london-westminster@nhs.net

Enclosures: "After ethical review – guidance for researchers"

Copy to: Ms Ruth Nicholson, Imperial College