

Evaluation of a continuous monitoring and feedback initiative to improve quality of anaesthetic care: a mixed-methods quasi-experimental study

Jonathan Benn, Glenn Arnold, Danielle D'Lima, Igor Wei, Joanna Moore, Floor Aleva, Andrew Smith, Alex Bottle and Stephen Brett



***National Institute for
Health Research***

Evaluation of a continuous monitoring and feedback initiative to improve quality of anaesthetic care: a mixed-methods quasi-experimental study

Jonathan Benn,^{1*} Glenn Arnold,² Danielle D'Lima,¹
Igor Wei,¹ Joanna Moore,¹ Floor Aleva,³
Andrew Smith,⁴ Alex Bottle⁵ and Stephen Brett⁶

¹Imperial Patient Safety Translational Research Centre, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK

²Department of Anaesthesia, Imperial College Healthcare NHS Trust, London, UK

³IQ Scientific Institute for Quality of Healthcare, Radboud University Medical Centre, Nijmegen, the Netherlands

⁴Department of Anaesthesia, Royal Lancaster Infirmary, Lancaster, UK

⁵School of Public Health, Imperial College London, London, UK

⁶Centre for Perioperative Medicine and Critical Care Research, Imperial College Healthcare NHS Trust, London, UK

*Corresponding author

Declared competing interests of authors: none

Published July 2015

DOI: 10.3310/hsdr03320

This report should be referenced as follows:

Benn J, Arnold G, D'Lima D, Wei I, Moore J, Aleva F, *et al.* Evaluation of a continuous monitoring and feedback initiative to improve quality of anaesthetic care: a mixed-methods quasi-experimental study. *Health Serv Deliv Res* 2015;**3**(32).

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hsdr>

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its preceding programmes as project number 11/1015/21. The contractual start date was in March 2012. The final report began editorial review in May 2014 and was accepted for publication in November 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2015. This work was produced by Benn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Health Services and Delivery Research Editor-in-Chief

Professor Ray Fitzpatrick Professor of Public Health and Primary Care, University of Oxford, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor John Norrie Health Services Research Unit, University of Aberdeen, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

Evaluation of a continuous monitoring and feedback initiative to improve quality of anaesthetic care: a mixed-methods quasi-experimental study

Jonathan Benn,^{1*} Glenn Arnold,² Danielle D'Lima,¹ Igor Wei,¹ Joanna Moore,¹ Floor Aleva,³ Andrew Smith,⁴ Alex Bottle⁵ and Stephen Brett⁶

¹Imperial Patient Safety Translational Research Centre, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK

²Department of Anaesthesia, Imperial College Healthcare NHS Trust, London, UK

³IQ Scientific Institute for Quality of Healthcare, Radboud University Medical Centre, Nijmegen, the Netherlands

⁴Department of Anaesthesia, Royal Lancaster Infirmary, Lancaster, UK

⁵School of Public Health, Imperial College London, London, UK

⁶Centre for Perioperative Medicine and Critical Care Research, Imperial College Healthcare NHS Trust, London, UK

*Corresponding author j.benn@imperial.ac.uk

Background: This study evaluated the impact of a continuous quality monitoring and feedback initiative in anaesthesia.

Objectives: To conduct a quasi-experimental evaluation of the feedback initiative and its effect on quality of anaesthetic care and perioperative efficiency. To understand the longitudinal effects of passive and active feedback and investigate the mechanisms and interactions underpinning those effects.

Design: Mixed-methods evaluation with analysis and synthesis of data from longitudinal qualitative interviews, longitudinal evaluative surveys and an interrupted time series study.

Intervention: Continuous measurement of a range of anaesthetic quality indicators was undertaken in a London teaching hospital alongside monthly personal feedback from case summary data to a cohort of anaesthetists, with follow-up roll-out to the whole NHS trust. Basic feedback consisted of the provision of passive monthly personalised feedback reports containing summary case data. In the enhanced phase, data feedback consisted of more sophisticated statistical breakdown of data, comparative and longitudinal views, and was paired with an active programme of dissemination and professional engagement.

Methods: Baseline data collection began in March 2010. Implementation of basic feedback took place in October 2010, followed by implementation of the enhanced feedback protocol in July 2012. Weekly aggregated quality indicator data, coupled with surgical site infection and mortality rates, was modelled using interrupted time series analyses. The study anaesthetist cohort comprised 50,235 cases, performed by 44 anaesthetists over the course of the study, with 22,670 cases performed at the primary site. Anaesthetist responses to the surveys were collected pre and post implementation of feedback at all three sites in parallel with qualitative investigation. Seventy anaesthetists completed the survey at one or more time points and 35 health-care professionals, including 24 anaesthetists, were interviewed across two time points.

Results: Results from the time series analysis of longitudinal variation in perioperative indicators did not support the hypothesis that implementation of basic feedback improved quality of anaesthetic care. The implementation of enhanced feedback was found to have a significant positive impact on two postoperative pain measures, nurse-recorded freedom from nausea, mean patient temperature on arrival in recovery and Quality of Recovery Scale scores. Analysis of survey data demonstrated that anaesthetists value perceived credibility of data and local relevance of quality indicators above other criteria when assessing utility of feedback. A significant improvement in the perceived value of quality indicators, feedback, data use and overall effectiveness was observed between baseline and implementation of feedback at the primary site, a finding replicated at the two secondary sites. Findings from the qualitative research elucidated processes of interaction between context, intervention and user, demonstrating a positive response by clinicians to this type of initiative and willingness to interact with a sustained and comprehensive feedback protocol to understand variations in care.

Conclusions: The results support the potential of quality monitoring and feedback interventions as quality improvement mechanisms and provide insight into the positive response of clinicians to this type of initiative, including documentation of the experiences of anaesthetists that participated as users and codesigners of the feedback. Future work in this area might usefully investigate how this type of intervention may be transferred to other areas of clinical practice and further explore interactions between local context and the successful implementation of quality monitoring and feedback systems.

Funding: The National Institute for Health Research Health Services and Delivery Research programme.

Contents

List of tables	xi
List of figures	xvii
List of boxes	xix
List of abbreviations	xxi
Plain English summary	xxiii
Scientific summary	xxv
Chapter 1 Introduction	1
Chapter 2 Structure of the report	3
Chapter 3 Background research and theory	5
The need for quality monitoring	5
Perioperative quality indicators	5
Data feedback interventions	7
Quality improvement in health care as an approach	11
Chapter 4 Research aims	13
Chapter 5 The intervention: a continuous monitoring and feedback initiative for anaesthetic care	15
Main approach and design rationale	15
<i>Increasing integration of the post-anaesthetic case unit form with the post-anaesthetic case unit nursing workflow and post-anaesthetic case unit patient information/data-logging requirements</i>	16
<i>Increasing availability of comparative databases, contextual information for individual patient records, quality indicator data available and analytic capability</i>	16
<i>Theoretically and evidence-informed practice resulting from the research effort</i>	16
<i>Responses made to feedback from the anaesthetist user group</i>	17
User feature requests: formative input from the evaluation	17
Main iterations to the data feedback intervention	18
<i>September 2010: version 1.0</i>	18
<i>March 2011: version 1.1</i>	18
<i>May 2011: version 1.2</i>	19
<i>October 2011: version 1.3</i>	19
<i>February 2012: version 2.0</i>	19
<i>April 2012: version 2.1</i>	20
<i>June 2012: version 3.0 (final report implementation)</i>	20
<i>July 2013: version 3.1</i>	20
<i>October 2013: version 3.2</i>	20

Chapter 6 Qualitative evaluation	21
Qualitative process evaluation of the initiative	22
<i>Introduction</i>	22
<i>Methods</i>	22
<i>Results</i>	24
<i>Discussion</i>	50
Mechanisms of data feedback effectiveness	52
<i>Introduction to relevant theories</i>	52
<i>Methods</i>	53
<i>Results: empirically informed critical appraisal of theory</i>	53
<i>Discussion</i>	67
Discussion and synthesis of qualitative findings	71
<i>Value of qualitative work for this type of research</i>	71
<i>Unanswered questions and limitations</i>	71
<i>Suggestions for future research</i>	72
<i>Overall conclusions and recommendations</i>	72
Chapter 7 Evaluative survey study	75
Introduction	76
Methods	78
<i>Development of the measurement instrument</i>	78
<i>Data collection</i>	80
<i>Analysis</i>	81
Results 1: psychometric analysis	83
<i>Demographics</i>	83
<i>Internal consistency of scales</i>	83
<i>Scale validation</i>	84
<i>Exploratory factor analysis</i>	85
<i>Discussion</i>	87
Results 2: predictors of perceptions of data feedback usability	88
<i>Descriptives</i>	88
<i>Regression analysis</i>	89
<i>Discussion</i>	91
Results 3: longitudinal evaluation based on perceptions of data feedback	93
<i>Primary site (St Mary's Hospital)</i>	93
<i>Charing Cross Hospital</i>	96
<i>Hammersmith Hospital</i>	99
Conclusions from the survey study	111
Chapter 8 Quasi-experimental evaluation of effect on anaesthetic quality and perioperative outcome indicators	115
Introduction	115
Methods	116
<i>Research design</i>	116
<i>Research aims</i>	117
<i>Intervention timeline</i>	120
<i>Sample and inclusion criteria</i>	123
<i>Research measures and data collection</i>	124
<i>Statistical analysis</i>	128
Results	132
<i>Observed sample</i>	132
<i>Patient temperature on arrival in recovery</i>	138
<i>Patient-reported quality of recovery</i>	144

<i>Postoperative pain</i>	149
<i>Postoperative nausea</i>	156
<i>Surgical site infection rate</i>	162
<i>Thirty-day mortality following surgery</i>	165
Discussion	169
<i>Summary of primary findings against hypotheses</i>	169
<i>Implications of secondary analysis for effect on perioperative quality</i>	172
<i>Data reliability</i>	181
<i>Study limitations</i>	183
Chapter 9 Impact on perioperative productivity	185
Introduction	185
Methods	186
<i>Study design</i>	186
<i>Qualitative data</i>	186
<i>Quantitative data</i>	187
<i>Analysis</i>	187
Results	187
<i>Factors influencing patient transfer efficiency</i>	187
<i>Effectiveness of feedback</i>	190
<i>Evaluation of the initiative</i>	192
Discussion	192
Chapter 10 Synthesis and conclusions	195
Chapter 11 Study recommendations	199
Recommendations for future research	199
Implications for practice	202
Summary of recommendations for development	205
Acknowledgements	207
References	211
Appendix 1 Qualitative interview schedule: first time point	223
Appendix 2 Qualitative interview schedule: second time point	225
Appendix 3 Evaluative survey: feedback on quality of anaesthetic care (St Mary's Hospital)	227
Appendix 4 Raw survey item scores by study epoch and site	231
Appendix 5 Recovery unit data collection form	235
Appendix 6 Extended project summary	239

List of tables

TABLE 1 Summary of main findings from systematic reviews which assess the impact of performance feedback on quality of care delivery or professional practice, and which provide information concerning the characteristics of effective feedback from review of empirical evidence	9
TABLE 2 Primary and secondary aims	13
TABLE 3 User feature requests that were requested during the formative qualitative investigation	17
TABLE 4 Simplified overview of interview topics with example questions asked to interviewees	23
TABLE 5 Overview of quantitative results with frequency breakdown by category	24
TABLE 6 Requested changes evident in time point 1 process evaluation	25
TABLE 7 Relevant theories extracted from the literature to be considered through qualitative analysis	54
TABLE 8 Key mechanisms of data use	57
TABLE 9 Research recommendations linked to key theories	69
TABLE 10 Data collection time points by site and level of feedback	81
TABLE 11 Internal consistency of survey scales	84
TABLE 12 Rotated component matrix	85
TABLE 13 Latent factors attributable to the variance in perceived effectiveness and usefulness of quality indicators and feedback	86
TABLE 14 Rotated component matrix	86
TABLE 15 Latent factors attributable to the variance in perceived positive departmental culture	87
TABLE 16 Respondents' characteristics: descriptive information about the clinicians that were included in the analysis	89
TABLE 17 Descriptive statistics of survey items: percentage responses to categorical variables, with number of responses shown in brackets, split by acute health-care organisation	89
TABLE 18 Descriptive statistics of survey items: mean responses to scale variables, with SD shown in brackets, split by acute health-care organisation	90

TABLE 19 Model summary: overview of statistics illustrating model fit for each of the five study hypotheses	91
TABLE 20 Coefficients: model parameters for final fitted model with all variables entered	92
TABLE 21 Overview of participants at each time point	93
TABLE 22 Items C1–4 (quality indicators)	94
TABLE 23 Items C5–9 (feedback)	94
TABLE 24 Items C10–15 (data use)	94
TABLE 25 Items C1–15 (overall effectiveness)	94
TABLE 26 Items D1–16 (workplace climate)	94
TABLE 27 Overview of participants at each time point	96
TABLE 28 Items C1–4 (quality indicators)	97
TABLE 29 Items C5–9 (feedback)	97
TABLE 30 Items C10–15 (data use)	97
TABLE 31 Items C1–15 (overall effectiveness)	97
TABLE 32 Items D1–16 (workplace climate)	97
TABLE 33 Overview of participants at each time point	99
TABLE 34 Items C1–4 (quality indicators)	100
TABLE 35 Items C5–9 (feedback)	100
TABLE 36 Items C10–15 (data use)	100
TABLE 37 Items C1–15 (overall effectiveness)	100
TABLE 38 Items D1–16 (workplace climate)	100
TABLE 39 Descriptives by item and time point	103
TABLE 40 Descriptives by item and time point	106
TABLE 41 Descriptives by item and time point	108
TABLE 42 Table linking anaesthetic/perioperative practice area with main evaluative hypotheses, causal mechanisms and outcomes	118
TABLE 43 Table defining dependent variables used as research end points within the evaluation	125

TABLE 44 Table of recorded case frequencies at St Mary's Hospital main theatres and for all Imperial sites (including St Mary's Hospital main theatres) by surgical speciality	135
TABLE 45 Table depicting summary gender statistics for cases performed by the pilot anaesthetist group at St Mary's Hospital main theatres and across all Imperial sites, including St Mary's Hospital main theatres	136
TABLE 46 Frequency and valid proportions of ASA score assignments for surgical cases performed both at St Mary's Hospital main theatres and across the whole trust	137
TABLE 47 Longitudinal summary of mean patient temperature on arrival in recovery and proportion of patients with temperature < 36 °C by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	138
TABLE 48 Main evaluative hypotheses	140
TABLE 49 Model parameter estimates for mean patient temperature, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	140
TABLE 50 Model parameter estimates for proportion of patients with temperature < 36 °C, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	141
TABLE 51 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup	141
TABLE 52 Longitudinal summary of mean QoR score and proportion of patients with QoR score above 14 by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	144
TABLE 53 Main evaluative hypotheses	146
TABLE 54 Model parameter estimates for mean QoR, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	146
TABLE 55 Model parameter estimates for proportion of patients with QoR scores above 14, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	147
TABLE 56 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup	147

TABLE 57 Longitudinal summary of proportion of patients with QoR pain score of 2 ('mostly free from severe or constant moderate pain' during the recovery period) and with pain on arrival scores in the range 0–1 ('none' or 'mild') by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	150
TABLE 58 Main evaluative hypotheses	152
TABLE 59 Model parameter estimates for proportion of patients reporting freedom from pain during recovery (QoR Scale), including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	152
TABLE 60 Model parameter estimates for proportion of patients reporting no or mild pain on arrival in recovery, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback version 3 and change in trend following onset of feedback version 3	153
TABLE 61 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup	153
TABLE 62 Longitudinal summary of proportion of patients with QoR PONV score of 2 ('mostly free from nausea, dry retching or vomiting' during the recovery period) and with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period) by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	156
TABLE 63 Main evaluative hypotheses	158
TABLE 64 Model parameter estimates for proportion of patients with QoR PONV score of 2 ('mostly free from nausea, dry retching or vomiting'), including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	159
TABLE 65 Model parameter estimates for proportion of patients with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period), including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback version 3 and change in trend following onset of feedback version 3	159
TABLE 66 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup	160
TABLE 67 Longitudinal summary of proportion of patients with a SSI-specific diagnostic code by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	162
TABLE 68 Main evaluative hypotheses	163

TABLE 69 Model parameter estimates for the proportion of patients with a SSI-specific diagnostic code, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	164
TABLE 70 Longitudinal summary of proportion of patients with 30-day mortality following surgery by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)	165
TABLE 71 Main evaluative hypotheses	167
TABLE 72 Model parameter estimates for the proportion of patients with 30-day mortality following surgery, including beta weight coefficients and associated significance of <i>t</i> -statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3	167
TABLE 73 Summary table of main quasi-experimental findings for primary analytic aims, by anaesthetic/perioperative quality area. The observed effects of implementing basic feedback within a feedback-naive cohort and then escalating the intensity of that feedback (the 'enhanced' feedback condition) are described, having controlled for longitudinal trends in the data	170
TABLE 74 Summary table of findings from ITSA of primary evaluative measures and sensitivity analyses. Significant effects are recorded as either positive improvement or negative detrimental, relative to the effect of the feedback intervention. Temporal effects are additionally described as either stepwise shifts in level between experimental conditions, or change in slope denoting a shift in the rate of change in a variable over time between experimental conditions. For the secondary sensitivity analysis, the findings relative to those of the primary analysis for each quality indicator are described	173
TABLE 75 Summary table of model fit statistics for ITS models fitted under the various primary and secondary analytic aim categories. Within each cell, the proportion of variance in the dependent variable explained by the overall model is indicated (synonymous with the square of the multiple <i>R</i> -statistic, expressed as a percentage), along with the <i>F</i> -statistic from ANOVA (a comparable indicator of the degree to which the model fits the data) and the associated significance of <i>F</i>	179
TABLE 76 Summary of findings from ITSA for primary research hypotheses	196
TABLE 77 Summary of findings from ITSA structured by anaesthetic quality indicator and main evaluative hypothesis	244

List of figures

FIGURE 1 The health-care system as a basic cybernetic feedback loop based on monitoring quality indicators	8
FIGURE 2 Mean scores by survey section and time point: St Mary's Hospital	95
FIGURE 3 Mean scores by survey section and time point: Charing Cross Hospital	98
FIGURE 4 Mean scores by survey section and time point: Hammersmith Hospital	101
FIGURE 5 Intervention timeline variants depicting relationship between the intervention timelines used in the evaluative ITSA and the main observed iterations in the development of the feedback intervention that took place within the project	121
FIGURE 6 Timeline of key data items present in case-level data set (research quality indicators data set merged with trust patient administration system and theatre administration data)	122
FIGURE 7 Case frequency by anaesthetist (pilot group only) and main intervention phase (study epoch)	133
FIGURE 8 Monthly rate of accumulation of perioperative cases within the research data set by contributing Imperial site	134
FIGURE 9 Frequency histograms depicting volume of cases by age for (a) full multisite research data set and (b) St Mary's Hospital main theatres	136
FIGURE 10 Time series for mean patient temperature recorded on arrival in recovery following surgery	139
FIGURE 11 Time series for proportion of patients arriving in recovery following surgery with temperature < 36 °C	139
FIGURE 12 Time series for mean QoR Scale score completed prior to discharge from PACU	145
FIGURE 13 Time series for proportion of patients with QoR Scale scores above 14	145
FIGURE 14 Time series for proportion of patients with QoR pain score of 2 ('mostly free from severe or constant moderate pain') during the recovery period	151
FIGURE 15 Time series for proportion of patients with pain on arrival in recovery score in the range 0–1 ('none' or 'mild')	151
FIGURE 16 Time series for proportion of patients with QoR PONV score of 2 ('mostly free from nausea, dry retching or vomiting' during the recovery period)	157

FIGURE 17 Time series for proportion of patients with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period)	157
FIGURE 18 Time series for the proportion of patients with a SSI-specific diagnostic code	163
FIGURE 19 Time series for the proportion of patients with 30-day mortality following surgery	166
FIGURE 20 Schematic of information workflow developed following appointment of researcher to oversee data collection in PACU and development of feedback system (March 2011)	182
FIGURE 21 Primary site items C1–15 (overall effectiveness)	231
FIGURE 22 Primary site items D1–12 (workplace climate)	232
FIGURE 23 Additional site 1 items C1–15 (overall effectiveness)	232
FIGURE 24 Additional site 1 items D1–12 (workplace climate)	233
FIGURE 25 Additional site 2 items C1–15 (overall effectiveness)	233
FIGURE 26 Additional site 2 items D1–12 (workplace climate)	234

List of boxes

BOX 1 Primary and secondary aims from original protocol	21
BOX 2 Primary and secondary aims from original protocol	75
BOX 3 Examples of the main areas covered by a 16-item scale designed to measure departmental climate for quality improvement. Individual items were rated on an 8-point Likert agreement scale	82
BOX 4 Primary and secondary aims from original protocol	115
BOX 5 Research hypotheses formulated to investigate the role of feedback at different levels of intensity on perioperative process and outcome indicators	117
BOX 6 Research hypothesis specified to investigate how lower-ranked anaesthetists responded to the feedback intervention	120
BOX 7 Primary and secondary aims from original protocol	185

List of abbreviations

AAGBI	Association of Anaesthetists in Great Britain and Ireland	ITSA	interrupted time series analysis
ANOVA	analysis of variance	NICE	National Institute for Health and Care Excellence
ASA	American Society of Anesthesiologists	NIHR	National Institute for Health Research
CLAHRC	Collaboration for Leadership in Applied Health Research and Care	PACU	post-anaesthetic care unit
EDC	electronic discharge checklist	PONV	postoperative nausea and vomiting
GA	general anaesthetic	QoR	quality of recovery
GMC	General Medical Council	SD	standard deviation
IMPAQT	Improving Anaesthetic Quality	SSI	surgical site infection
IT	information technology	WWT	ward wait time
ITS	interrupted time series		

Plain English summary

The experience of pain, nausea, comfort and well-being in the immediate postoperative recovery period is critical to patient satisfaction with surgical care. The quality of the anaesthetic care delivered by the anaesthetist is important in controlling pain and nausea, and may influence longer-term outcomes. Despite this, anaesthetists rarely receive routine feedback on the care that they provide. This is partly due to the challenges of measuring quality of anaesthesia, but additionally it is due to a lack of information systems capable of providing effective feedback. We know from research studies that feedback which is timely, continuous, trusted, non-judgemental and supported by local resources and action-planning processes can improve professional practice and standards of care. In order to learn more about whether or not feedback could be effective for anaesthetists, we implemented and evaluated an initiative over 3 years to provide 44 anaesthetists with basic and subsequently enhanced personalised feedback. We attempted to determine (1) whether or not this initiative improved anaesthetic quality measures over time, as evidenced by statistical trends, and (2) whether or not anaesthetists would engage with the feedback and view it as useful, through conducting surveys and interviews. We found that providing comprehensive personalised feedback to anaesthetists as part of a long-term programme, which they had codesigned, was effective in improving measures of postoperative pain, nausea and quality of recovery from surgery, as well as engaging the local professional group. We concluded that this feedback initiative could be of broader benefit to health-care professionals and patients if implemented elsewhere.

Scientific summary

Background

Effective monitoring and feedback is essential if clinical teams and individuals are to understand variations in care, detect and respond to opportunities to improve standards and evaluate the impact of changes to services.

Anaesthetists as a professional group have a high degree of patient contact in the perioperative pathway yet receive little routine feedback on patient experience or outcomes such as pain and postoperative nausea. Feedback on such outcome measures often occurs irregularly in acute care organisations through clinical audit projects, but these information streams are discontinuous and not geared towards continuous monitoring and improvement. Recent reviews have highlighted that, from the anaesthetist's perspective, current perioperative quality indicators lack sensitivity and specificity. Furthermore, there exists limited evidence concerning the reliability and validity of measures that can be used to monitor patient satisfaction with anaesthetic care. A clear need exists for the development of routine monitoring and feedback of quality of anaesthetic care, from the postoperative period, to support improvement in anaesthetic practice.

Studies show that providing feedback to clinicians can be an effective improvement intervention and results in generally small to moderate positive effects on professional practice. Initiatives that use feedback are more effective than those that do not, and feedback paired with an educational strategy or implementation plan is more effective than using simple passive feedback alone. Research suggests that a large range of characteristics of feedback may influence its effectiveness as a quality improvement mechanism.

Objective

To conduct a comprehensive, mixed-methods, quasi-experimental evaluation of the impact of a departmental continuous quality monitoring and feedback initiative for quality of anaesthetic care, within a London teaching hospital over a 3-year period.

The intervention

The feedback initiative was developed and implemented as part of the Collaboration for Leadership in Applied Health Research and Care North West London portfolio of quality improvement projects and was given the title IMPAQT (Improving Anaesthetic Quality). Based on industrial quality improvement models, the initiative was conceived as a continuous quality monitoring and feedback programme for anaesthetists. It comprised continuous measurement of anaesthetic quality indicators in the post-anaesthetic care unit (PACU) of the primary site coupled with continuous monthly feedback of personal-level case data to 44 consultant anaesthetists. Baseline data collection of anaesthetic quality indicators began in March 2010 at St Mary's Hospital main theatre suite. The intervention model was then implemented in two main phases: (1) implementation of basic, passive monthly feedback using a simple summary statistical report (from October 2010), and (2) implementation of an enhanced feedback protocol (from July 2012 until the end of the project in November 2013).

Basic feedback comprised the provision of monthly personal data summaries in tabled form with limited longitudinal and comparative graphical representations. In the enhanced phase of the programme, a more statistically sophisticated report was developed in response to user requests, including monthly detailed

case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data and institution-wide dissemination. Basic data feedback was enhanced with broader professional engagement activities including regular presentation of statistical results, consultative interviews by the research team, topic-focused engagement and facilitated peer interaction on specific specialty areas (e.g. pain management after gynaecological surgery).

Methods

Qualitative evaluation

A longitudinal, qualitative work stream was used, which ran parallel to the intervention work and took a realist evaluative perspective on the project. The realist position provides a framework for identifying not only what outcomes are produced by an intervention, but how they are produced and how the intervention interacts with varying local conditions to produce outcomes. In total, interviews were conducted with 24 consultant anaesthetists, six surgical nursing leads and five perioperative service leads, in two phases.

Survey evaluation

A longitudinal evaluative end-user survey study was undertaken, with three time points corresponding to baseline, basic feedback and enhanced feedback conditions, across three participating sites. The items included scales designed to quantify the effectiveness of current quality indicators, data feedback and the usefulness of the feedback for improvement, along with attitudes to quality improvement within the local working environment. In total, 70 individual anaesthetists completed the survey at one or more time points.

Quasi-experimental evaluation

Evaluation of the impact of the anaesthetics quality monitoring and feedback initiative on anaesthetic quality indicators and perioperative outcomes utilised a single-group longitudinal design, with multiple study epochs. Interrupted time series analysis was used as the primary evaluative model, with interrupts representing multiple intervention time points corresponding to the onset of basic and enhanced feedback protocols. The perioperative indicators modelled included patient temperature on arrival in recovery (two metrics), patient-reported Quality of Recovery (QoR) Scale score (two metrics), postoperative pain (two metrics), postoperative nausea (two metrics), surgical site infection (SSI) rate and 30-day postoperative mortality rate. The study anaesthetist cohort comprised 50,235 cases, performed by 44 anaesthetists over the course of the study, with 22,670 cases performed at the primary hospital site.

Productivity analysis

Interviews were conducted with the perioperative service manager, the lead nurse for the PACU and six surgical nursing leads from the primary site to identify and interpret themes related to productivity. Ward wait time (WWT) was compared pre and post feedback, defined as the interval between the receiving ward being contacted after the patient was deemed ready for discharge from the PACU and the handover of the patient.

Results

Qualitative evaluation

The results provided a rich understanding of the causal mechanisms of effectiveness for monitoring performance and making improvements to practice based on quality indicators, along with a developmental perspective on acceptability and engagement over time. Clinicians clearly agreed with the rationale for the initiative, recognising the existence of a problem and the need for a solution. Clinicians emphasised that the right quality indicators needed to be selected with the right characteristics (i.e. they must be specific, relevant and meaningful) in order to promote the necessary level of trust in the data and

demonstrate fitness for purpose. The interviewees explored the translation of information into action at two levels of the health-care system: the departmental level and the individual clinician level. Crucially, the mechanisms of effective data use were different at each level.

The issue of anonymity was important to end-users and appeared to demonstrate a process of maturity that was longitudinally dependent and tied to end-users' growing confidence in the intent of the feedback system. At the individual level, with a desire to receive normative feedback, people wanted to identify and contact high performers in order to obtain support and ideas for behaviour change. Dealing with case-mix variations (and the intraprofessional issues it gave rise to) was identified as a critical success factor for initiatives of this type. Our analysis suggests that a combination of normative comparison (i.e. genuine peer benchmarking) and individual trends over time may have the greatest effect. The need to transform hard data into usable information and the experience of health-care professionals in doing so can be viewed as a powerful message emerging from this study.

Interviewees clearly identified a role for this initiative in revalidation and participating in quality monitoring and acting on the results is an identified dimension of good medical practice. The connection of the initiative with revalidation and appraisal appeared to significantly increase levels of engagement throughout the evolution of the project.

Survey evaluation

The results suggest that anaesthetists perceive a range of factors as important in determining the usefulness of feedback. Specifically, the local departmental context and its support of quality improvement is an important determinant of how instrumental feedback from monitoring quality indicators is likely to be. Furthermore, feedback that is tailored to be relevant to the personal professional practice of the individual clinician is an important predictor of usefulness. In terms of the feedback content and design characteristics that anaesthetists value most, the perceived credibility of the data and the local relevance of the quality indicators are paramount.

In the longitudinal evaluation, the survey data from the primary site demonstrated a significant improvement in perceptions of quality indicators, feedback, data use and overall effectiveness of quality monitoring between baseline and implementation of basic feedback. For the majority of the survey measures, there was a significant improvement at the secondary study sites between the baseline condition and the implementation of the enhanced feedback protocol.

Interrupted time series analysis

The observed response to the implementation of basic feedback in the quality indicators assessed was, on the whole, limited, and the hypothesised benefits of implementing basic feedback were generally not observed in the data. While the average weekly temperature of patients arriving in recovery increased by 0.082 °C in response to the onset of basic feedback, in contrast, the weekly proportion of patients arriving in recovery with temperature under 36 °C increased between the baseline and basic feedback condition, too, by 3.3%.

The second study hypothesis concerned the effect of implementing enhanced feedback in a group that had been receiving routine basic feedback previously. Escalating the intensity of feedback through implementation of an enhanced feedback protocol had a positive effect across a greater range of measures than implementation of basic feedback. After implementation of enhanced feedback, patients were, on average, warmer on arrival in recovery by 0.064 °C. Both the mean patient-reported QoR Scale score and the proportion of patients reporting high-quality recovery showed small improvements in the rate of change between basic and enhanced feedback conditions (change in trend for mean scale score = 0.009 on a 17-point scale; change in trend for proportion of patients = 0.001%).

The two measures of postoperative pain demonstrated consistent positive responses to the implementation of the enhanced feedback protocol, with significant improvement in both level and rate of change in the proportion of patients reporting freedom from severe pain (change in level = 7.2% of patients; change in trend = 0.004%) and those reporting no or mild pain on arrival in recovery (change in level = 12% of patients; change in trend = 0.003%). The proportion of patients with nurse-reported absence of nausea similarly increased by 5.8% in response to the implementation of enhanced feedback, coupled with an improvement in the rate of change in this measure (change in trend = 0.001%), although no significant effect was detected for patient-reported freedom from postoperative nausea. No significant effect of the implementation of enhanced feedback on SSI rate was detected. Thirty-day postoperative mortality appeared to show a complex response to the implementation of enhanced feedback, with a modest increase in level (0.8%) coupled with a modest improvement in the rate of change over time (change in trend = 0.001%).

The observed positive effects of enhanced feedback on patient temperature data, mean QoR score, and measures of both postoperative pain and postoperative nausea were robust after covariate analysis in which longitudinal variation in disease severity, patient age and gender were controlled. When further statistical models were fitted based on stricter case inclusion criteria, limited to elective general anaesthetic cases, the implementation of enhanced feedback was still associated with significant improvement in level and slope in both proportion of patients with nurse-reported freedom from nausea and proportion of patients with freedom from pain on arrival in recovery. Analysis of the impact of feedback on a subgroup of anaesthetists defined by being ranked in the lower 50th percentile of scores during the baseline period confirmed that the beneficial effects of enhanced feedback were replicated and strongest for this subgroup, particularly in the areas of postoperative pain management, control of postoperative nausea and overall patient-reported QoR.

Productivity analysis

Although most perioperative ward leads described the initiative as useful and recognised the importance of active clinical engagement, significant improvements in WWT were only reflected in three of the eight wards studied. Qualitative analysis of stakeholder interviews suggested that although the stakeholders valued the feedback initiative, further organisational changes would be needed to progress improvement in ward transfer efficiency, including improvements to the local bed allocation and discharge process. Although there was strong consensus of support for the initiative, the interviews highlighted the need for an infrastructure to support change and shared goals targeted through a system-wide approach that included broader understanding of hospital dynamics.

Conclusions and implications for health care

Taken as a whole, the findings from this evaluation provide rich information concerning the effects of a comprehensive, long-term anaesthetic quality monitoring and feedback initiative on multiple dimensions of service performance. Furthermore, they provide insight into the process of development that took place within this initiative, of interactions between context, intervention and user, and document the experiences and perceptions of the anaesthetists that participated as end-users and codesigners of the feedback. Productive future directions for research include (1) investigation of whether or not this model of a continuous feedback intervention will port successfully to other clinical specialties, (2) in-depth cost-benefits analysis of a continuous quality monitoring model compared with conventional discontinuous audit, and (3) how variations in context relating to the maturity of local electronic record systems affect implementation and outcome of similar initiatives.

The research findings give rise to the following specific implications:

1. Quality monitoring and feedback interventions represent a potentially important quality improvement mechanism, especially where investment is made in their long-term development and sustainment.
2. The design of feedback and its perceived intent, fitness for purpose and context of use are all important considerations for success.
3. It is essential to not only involve end-users in the development of the feedback system at conception, but to foster an ongoing sense of ownership of the data and a willingness to interact with them.
4. It is important to pair passive data dissemination with support, active engagement and opportunities for intra- and interprofessional dialogue, concerning how to respond to evidence of variations and opportunities for improvement.
5. Continuous feedback can make the natural variation inherent in human-intensive processes, such as health care, visible to improvement efforts. In so doing, subjective and intangible phenomena, such as patient satisfaction, may be objectified for more constructive conversations, enhanced shared decision-making and better control.
6. In the development of monitoring and feedback systems, appropriate attention must be given to how data are used and converted into information for specific user groups, such as clinicians, rather than simply focusing on what to measure and how reliable those measures are.
7. The success of data feedback interventions should be evaluated using multiple dimensions, including social, organisational and professional outcomes, in addition to clinical end points.
8. While downstream postoperative outcomes may be insensitive to the effects of an anaesthetic quality feedback intervention, process-of-care measures, such as those associated with postoperative pain management, nausea and perioperative normothermia, are more receptive.
9. Within the health informatics field, considerable scope exists beyond this project to further test evolving theory and practice from improvement science and industrial process control related to how data can be used to support continuous improvement in process-based operations.
10. The trend towards a shift away from intermittent, snapshot audits of practice in favour of a continuous monitoring and continuous improvement model within health-care organisations should be the subject of further investigation in terms of its implications for patients and quality of care.

Funding

Funding for this study was provided by the Health Services and Delivery Research programme of the National Institute for Health Research.

Chapter 1 Introduction

In the UK, recent policy reports have set a clear agenda to improve the way in which health-care organisations report on the quality of care delivered to patients, alongside the more conventional financial aspects of service performance. With UK acute care trusts now required to produce quality accounts to report on quality of care delivered to patients, there is a need to support specific clinical service areas in the development of criteria for monitoring quality of care.

The NHS Next Stage Review report *High Quality Care for All*¹ aims to make organisations accountable for quality through a focus on the measurement and reporting of quality indicators representing safety, effectiveness and patient experience.¹ More recently, these principles form a core for planned changes to the health service to link payment of providers to performance on clinical quality outcomes as an incentive for continuous improvement in standards of care.² The Francis Report called for information that is accessible and useable by all, allowing for effective comparison of performance by individuals, services and organisations.³

Though recent policy has focused on organisational-level reporting of quality indicators using quality accounts, other research has suggested that robust measurement systems to evaluate progress in improving the quality and safety of care at the clinical systems level are lacking.^{4,5} The General Medical Council's (GMC) work towards the development of revalidation processes for doctors has seen broad consultation and the establishment of task groups in the national clinical bodies to develop criteria and processes that are acceptable to each clinical profession. Within the anaesthetics specialty, the Royal College of Anaesthetists, the Association of Anaesthetists for Great Britain and Ireland (AAGBI) and the National Institute of Academic Anaesthesia are engaged in work to support the evolving revalidation agenda. This work is being driven by a number of specialist working groups, co-ordinated by the newly formed Health Services Research Centre hosted by the Royal College. A committee has been set up to investigate and develop national frameworks for revalidation of anaesthetists, including specific working groups in the areas of patient and peer feedback (multisource feedback), and quality indicators and outcomes. There is a clear need for health services research to support these activities and the development of the revalidation agenda through understanding how an effective quality indicators framework, data feedback process and the necessary service organisation structures can promote continuous improvement in clinical practice.

Considering the requirements for effective processes to feed back data from anaesthetic quality indicators to clinicians is currently a topical focus. Recent specialty focus on revalidation for anaesthetists has highlighted the requirement for metrics that can provide supporting evidence for individual professional revalidation.^{6,7} This requirement places considerable demands on local quality monitoring systems, which will vary in capability. The perception of the revalidation process and the use that is made of personal professional data within it will be important and it has been suggested that multisource feedback is used constructively to support professional development and is separate from the assessment process used to judge doctors.⁸ The Royal College of Anaesthetists and the AAGBI have published guidance on revalidation and the necessary supporting documentation for evidence of fitness to practice.⁹ Feedback from quality indicators can be used as key support within the multisource feedback required for revalidation. Indeed, participating in quality monitoring and acting on the results is included as one aspect of good medical practice.¹⁰

Chapter 2 Structure of the report

This report begins with a review of the relevant literature and a thorough description of the feedback intervention that was implemented. The remaining sections of the report are then reported by substudy within the mixed-methods design, beginning with qualitative investigation of the initiative, followed by analysis of longitudinal survey data designed to support the evaluation. The report then proceeds with the results from interrupted time series analysis (ITSA) of the impact of the initiative on anaesthetic quality and perioperative outcome indicators.

Each substudy section is designed to be self-contained, including description of the methodological details pertinent to each stream of enquiry, summary of the results or development of qualitative theory and discussion of the findings. The final two sections of the report are designed to draw together the results across the studies that constitute the mixed-methods design and synthesise the key findings and conclusions, prior to a discussion of the important implications for research and practice.

Chapter 3 Background research and theory

The need for quality monitoring

Effective monitoring of the quality of service delivery is central to the capacity of an organisation, unit or individual to maintain and improve standards of care. International studies have found broad variations in care quality and well-publicised examples of the consequences for quality of care and patient safety exist, where unreliable systems and variable professional practices are inadequately monitored.^{3,11} Monitoring is essential if clinical units and individuals are to (1) understand the factors underlying variations in care, (2) detect and respond to opportunities to improve standards and (3) evaluate the impact of changes to services.

Drawing on the rationale and past research outlined above, there is a clear need for work in this area to support the current national service agenda to promote improvement in quality of care and revalidation for clinicians across the professions. There is mounting economic pressure for productivity in the perioperative workflow, which focuses effort on the intraoperative stages of care and theatre utilisation efficiency. Anaesthetists as a professional group have a high degree of patient contact in the perioperative pathway and yet receive little routine feedback on patient experience or outcomes specific to the quality of the anaesthetic care process. There is a need for anaesthetists to receive quantitative feedback from the postoperative stage on the quality of care they deliver to patients and the patient experience. Indeed, owing to tight integration within the surgical process and the development of the modern anaesthetist's role in the perioperative pathway, anaesthesia poses some specific and potentially unique challenges in terms of providing feedback on quality of care to departments and individual clinicians.

Feedback on postoperative nausea or pain control often occurs on a periodic basis through successive audit cycles, but these information streams are discontinuous and may not be geared towards continuous improvement actions. Feedback on the quality of recovery (QoR) experienced by patients on waking from surgery often occurs only on an ad-hoc basis, through personal interactions with patients and recovery room staff or when anaesthetists have the opportunity to personally follow-up patients in between busy theatre list schedules or are requested to attend a patient by recovery room staff. There are few reported examples of systematic, routine monitoring of quality indicators in the recovery room to provide rapid, continuous feedback on the quality of anaesthetic care delivered. Similarly, few evaluative studies of personal professional monitoring programmes for anaesthetists or trainees exist.^{12,13}

Perioperative quality indicators

As a precursor to consideration of how data might best be used to monitor quality of care in anaesthetic services, it is important to consider how quality can be defined and measured. Contemporary practice in development and application of quality indicators to monitor and improve care owes much to the early work of Avedis Donabedian.¹⁴⁻¹⁶ Donabedian made the distinction between structure, process and outcome metrics according to the relationship and proximity of a variable to the desired performance result.¹⁴

In accordance with policy developments, considerable work has been undertaken in specialty areas, including anaesthesia, to identify the potential metrics by which quality of service delivered can be reported and evaluated. Within anaesthetics in particular, there is a lack of reliable, evidence-based quality indicators to quantify quality of care, patient experience and process efficiency, and capable of guiding service development efforts and improvement in professional practice. The close integration of the anaesthetist's function with that of the surgeon and other proceduralists within the perioperative pathway means that identification of routine outcome metrics that are sensitive to variations in quality of

anaesthesia is difficult. A recent review of quality indicators for anaesthesia concluded that conventional perioperative morbidity and mortality data largely lack the sensitivity and specificity necessary for analysis of variation in quality and safety of anaesthesia.¹⁷ In Haller's review,¹⁷ 108 quality indicators in use within the anaesthetics research literature were identified. Around half of the indicators looked specifically at anaesthesia; the other half also measured surgical or postoperative ward care. Only 1% of indicators looked at structure of care; the majority (57%) measured outcome and 42% measured process of care.

Mortality and serious morbidity attributable to anaesthesia has decreased significantly over the last 50 years to the point where mortality is considered a poor quality indicator because it is both rare and frequently related to factors outside the anaesthetists' control. Anaesthesia has been considered at the forefront of improving safety in health care¹⁸ and, currently, with around 2,300,000 general anaesthetics (GAs) performed every year in the UK, the risk of death and major complications after surgery in the general surgical patient population is relatively low: < 1% of all patients undergoing surgery die during the same hospital admission, with perioperative mortality occurring in only 0.2% of healthy elective patients.¹⁹

While variation in perioperative morbidity and mortality are influenced by a range of patient, surgical and anaesthetic factors, variation in the quality of anaesthetic care may be more directly assessed in the immediate postoperative period, in which the patient's experience of recovery is closely linked to the quality of the anaesthetic and the selection of analgesic and antiemetic properties. Although outcomes such as mortality are concrete and therefore easier to measure, outcomes such as patient experience and satisfaction require quantification of subjective perceptions, usually along multiple dimensions. It may be this reason that underlies the observation that there are few validated indicators available that incorporate the patient's perspective on quality of anaesthetic care.¹⁷ Indeed, Haller reported that only 40% of clinical indicators had been validated beyond face validity.¹⁷

In terms of patient satisfaction with anaesthesia, a number of postoperative patient satisfaction questionnaires have been developed and validated.²⁰⁻²⁶ Myles^{26,27} developed a nine-point QoR Scale, measuring patient-reported dimensions of quality, for use as a perioperative outcome measure and for clinical audit. The scale included items derived from a larger questionnaire used to determine patient priorities, and included aspects such as general well-being, support from others, understanding of instructions, respiratory function, bowel function, nausea and pain, among others.²⁸ Psychometric analysis found that the scale showed good internal consistency and validity, with factors such as invasiveness of surgery, duration of hospital stay and gender all being significant predictors of QoR scores.^{27,29} Capable systems for routine monitoring of quality of anaesthetic care as experienced by perioperative patients lags some way behind research-based development of patient satisfaction measures, however, representing a clear gap for operationalisation and translation of research evidence into clinical practice.

A large body of established evidence, linked to national guidelines, supports the impact of various anaesthetist perioperative practices such as patient temperature control, antibiotic administration and management of anaesthesia on care quality and perioperative outcomes [e.g. National Institute for Health and Care Excellence (NICE)³⁰].

Two of the most important dimensions of QoR in the postoperative period are postoperative nausea and vomiting (PONV) and postoperative pain.³¹ PONV is both common and among the most important factors contributing to a prolonged postoperative stay following ambulatory surgery.³² PONV and incisional pain during recovery has a strong negative influence on patient satisfaction, and studies have shown that vomiting and nausea are among the most undesirable complications from the patient's point of view³¹ as well as posing clinical risks. Pain is commonly measured in a post-anaesthetic care unit (PACU) by a variety of means, including visual analogue scales, numerical rating scales, verbal rating scales and behavioural scales.³³ Research suggests that 10.9% of surgical patients experience severe pain.³⁴ A range of risk factors have been identified for PONV, and guidance has been produced for antiemetics administration to patients that are moderate and high risk.³⁵ Aside for the importance patients place on experience of the immediate postoperative period, patient-reported pain and PONV represent important anaesthetic quality indicators

for the attending anaesthetist, as they are interrelated and dependent on the balance between analgesic and antiemetic properties, matched to the patient's characteristics and procedural requirements.

Many quality metrics are routinely measured by staff working in the postoperative environment; temperature on arrival in recovery and time spent in recovery are important aspects of a patient's experience, and are readily quantifiable. Evidence demonstrates that perioperative normothermia results in a reduced incidence of wound infection³⁶ and has been established as a NICE guideline.³⁰ Despite this, a recent UK review of perioperative care for surgical patients found that 34% of hospitals had no policy for the prevention of perioperative hypothermia.²⁰

Research on the development of effective quality indicators for clinical practice suggests that they should be transparent, reliable, evidence-based, measurable and improvable.³⁷ However, it is clear that there are certain challenges in the measurement of quality of care in anaesthesia which must be overcome, and some consensus is emerging as to what may be useful and reliable basic metrics for anaesthetic quality, collected in the recovery period.

Further study to identify sensitive and useful quality indicators for anaesthesia is warranted. From a practical perspective, while variation attributable to the anaesthetic component of care may be determinable in perioperative morbidity and mortality data through well-powered research studies, the requirements of local clinical units differ. Anaesthetic quality indicators must be useful when monitored longitudinally in small numbers of cases, with limited opportunity for case-mix adjustment. Quantification of dimensions associated with the patient's experience of immediate postoperative recovery may provide valid and reliable outcome measures for anaesthesia if evidence and consensus based.

Data feedback interventions

Monitoring involves collecting data on important quality indicators. However, identifying practical metrics that are valid and reliable indicators of the quality of service delivery is only one part of the challenge. Effective quality control requires clinical units to not only develop reliable data collection mechanisms, but to put in place systems and processes for effective feedback and use of the data to support quality improvement. Considerable research and development has been undertaken to define the processes by which valid, reliable and useful clinical quality indicators can be defined, using systematic approaches such as review of scientific research and expert consensus.³⁷⁻³⁹ The question of how the resulting quality indicators can best be used in practice, however, is not currently supported by a coherent body of literature. The conventional clinical audit model often fails to deliver sustainable, effective change^{40,41} and clinicians do not routinely use or engage in quality improvement.⁴²⁻⁴⁴ Crucially, quality indicators generate data representing variation in an underlying parameter of the care process (plus measurement error). How those data are turned into information and useful, actionable intelligence to support quality assurance and improvement is an important issue and is as central to the design of effective quality monitoring systems in health care as defining the right measures in the first place.⁴⁵ Within the patient safety domain, for example, the lessons from incident reporting suggest that simply focusing on data collection and building large databases, without feedback to the front line, is insufficient to sustain engagement and action.⁴⁶

The term 'feedback' is most often used to describe the act of providing knowledge of the results of behaviour or performance to the individual.^{47,48} Within a health-care context, information feedback has been defined as 'any summary of clinical performance of health care over a specified period of time, given in a written, electronic or verbal format'.⁴⁹ A related and useful perspective is that of control systems engineering or cybernetics, in which feedback refers to the process by which data describing the output from a system are returned to an earlier stage of the system to modify their behaviour in order to affect future outputs (*Figure 1*). Crucially, information by itself is not feedback by the engineering definition; rather, feedback must

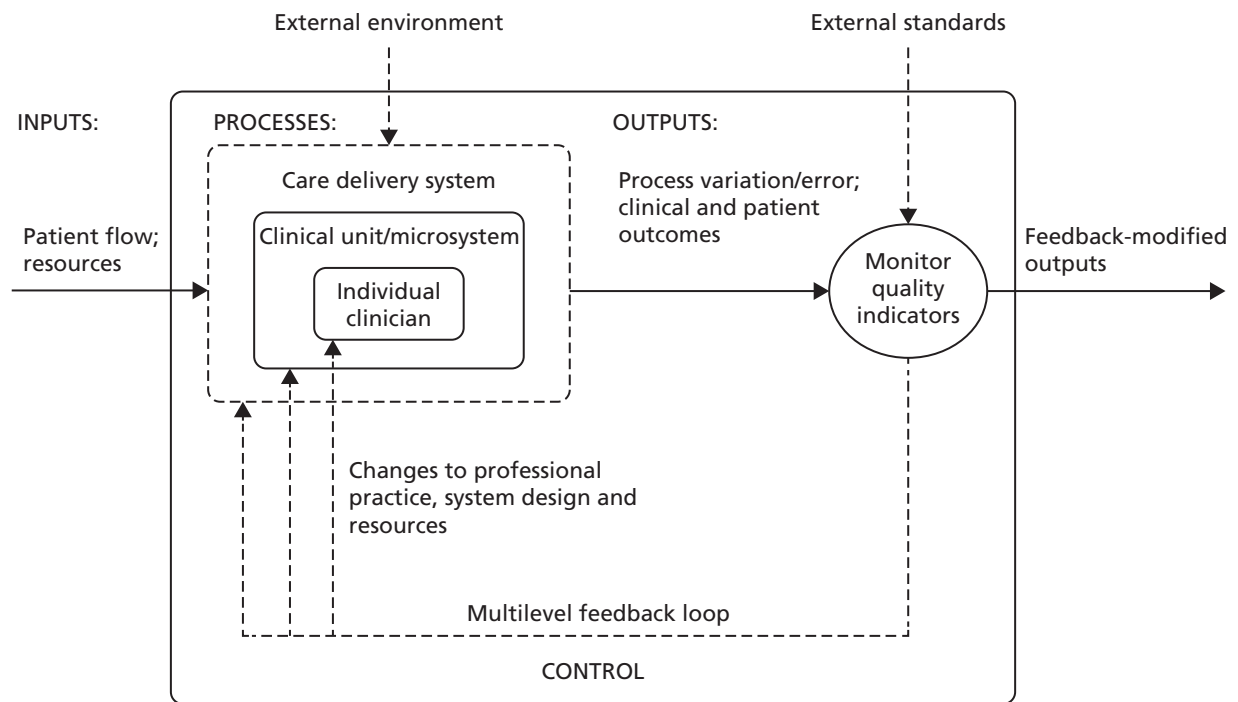


FIGURE 1 The health-care system as a basic cybernetic feedback loop based on monitoring quality indicators. Information concerning outputs from the current system is fed back to an earlier stage to modify processes in order to achieve enhanced future.

incorporate some action or response to close the identified gap.⁵⁰ From an organisational perspective, feedback from operational experience over time is an important mechanism of organisational learning, resulting in incremental and large-scale modification to care systems and processes.⁵¹

In recent studies, many different feedback strategies are applied in clinical settings, but the type of feedback strategies as well as results produced are often heterogeneous.⁵²⁻⁵⁴ The current available evidence is certainly not conclusive, but, taken as a whole, providing feedback results in generally small to moderate positive effects on professional practice⁴⁹ and process-of-care measures may be more sensitive to data feedback initiatives than outcomes.⁵² The literature shows that initiatives that do not use feedback reports are less effective than those that do use feedback reports, regardless of whether or not the feedback is accompanied by an implementation plan.^{53,55} A number of published systematic reviews are relevant to the efficacy of performance feedback to improve professional practice and quality of care. These are summarised in *Table 1*.

The practical value of this literature lies in the descriptions it provides of the range of potential feedback mechanisms that have been developed to improve clinical practice and the conclusions drawn about the value of specific system characteristics in terms of their contribution to effective feedback capable of engendering improvement. Van Der Veer's review identified different barriers and success factors concerning feedback. The most frequently cited barrier was lack of trust in data quality, followed by lack of intensity of feedback and lack of motivation. Success factors included sufficient timeliness (time between data collection and the forthcoming feedback report), dissemination of information, trust in data quality and having a confidential or non-judgemental tone.⁵²

Jamtvedt found that low baseline compliance with recommended practice and higher intensity of audit and feedback were associated with a greater effectiveness of the feedback intervention.⁴⁹ Other important characteristics of feedback seem to be the source and the duration of the feedback.⁵⁶ Physicians are more influenced by an authoritative, credible source that will continue to monitor the physicians' performance over several years.⁵⁶ Mugford concluded that information feedback was most likely to influence clinical

TABLE 1 Summary of main findings from systematic reviews which assess the impact of performance feedback on quality of care delivery or professional practice, and which provide information concerning the characteristics of effective feedback from review of empirical evidence

Review	Year	Effect of feedback	Characteristics of effective feedback
Van der Veer <i>et al.</i> ⁵²	2010	Process-of-care measures are more positively influenced by feedback than outcome of care measures. Adding elements to a feedback strategy positively influences its effectiveness	The most mentioned factors influencing the effectiveness of feedback were (trust in) data quality, motivation of the recipients, intensity of feedback, organisational factors, outcome expectancy of recipients, timeliness, dissemination of information and confidentiality/non-judgemental tone
De Vos <i>et al.</i> ⁵³	2009	Effective strategies to implement quality indicators do exist, but there is considerable variation in the methods used and the level of change achieved	Feedback reports in combination with an educational implementation strategy and/or the development of a quality improvement plan seems to be most effective in improving quality. Analysis of barriers led to the identification of the following barriers in quality improvement: unawareness, lack of credible data, no support management to units and lack of hospital resources
Hysong <i>et al.</i> ⁴⁵	2009	The effect size estimate of 0.40 (96% confidence interval) suggests that audit and feedback has a modest though significant effect on the outcome of interest	Identified characteristics that augment feedback effectiveness are correct solution information and written feedback delivery. Verbal and graphic feedback delivery attenuated feedback effectiveness. There seems to be a trend in that both individual- and group-level feedback may be beneficial but data could not significantly confirm this
Jamtvædt <i>et al.</i> ⁵⁴	2006	Audit and feedback can improve professional practice, but effects are variable. Effects are generally small to moderate	Low baseline compliance with desired practice and higher intensity of feedback were associated with higher adjusted risk ratios across studies, and therefore are more effective in improving quality
Veloski <i>et al.</i> ⁵⁶	2006	Forty-one studies evaluated the independent effect of feedback. Of these, 32 (74%) demonstrated a positive impact on physician performance	Feedback can change physicians' clinical performance when provided systematically over multiple years by an authoritative, credible source
Mugford <i>et al.</i> ⁵⁷	1991	Feedback of information most probably influences clinical practice if it is part of an overall strategy which targets decision-makers who have already agreed to review their practice	Information feedback is likely to have a more direct effect on practice if presented close to the time of decision-making

practice if the information was presented close to the time of decision-making and the clinicians had previously agreed to review their practice.⁵⁷ De Vos found that feedback reports in combination with an educational implementation strategy or quality improvement plan seemed to be most effective. The following common barriers to quality improvement were identified: unawareness, lack of credible data, no support management to units and lack of hospital resources.⁵³

Such findings from systematic reviews of primary empirical sources are echoed by experience in quality improvement projects, which identify effective feedback that supports improvement in care as being sustained or continuous, timely, locally relevant, credible, non-punitive and supportive of remedial action.⁵⁸ Other reviews of the evidence linked to the efficacy of professional behaviour change interventions have found that there are 'no magic bullets', that information dissemination alone was rarely effective and that moderate positive results can be achieved using more complex interventions.⁵⁹ Similarly, synthesis of evidence on professional education and quality assurance found that simple passive intervention approaches were generally ineffective and unlikely to result in behaviour change, whereas multifaceted

interventions involving educational components were found to be more effective.⁶⁰ This finding echoes that of reviews of data feedback interventions, in which multifaceted approaches that involve adding elements to basic data feedback were found to be most effective.⁵²

The distinction between passive data dissemination and more active forms of feedback is of importance to the effectiveness of a data feedback intervention, with several reviews concluding that maximum impact was achieved through embedding data feedback in targeted reporting, quality improvement strategies, educational programmes or similar multicomponent initiatives. Mugford defines passive feedback as the unsolicited provision of information with no stated requirement for action. Active feedback occurs where the interest of clinicians has been stimulated and engaged in aspects of practice, perhaps through the process of agreeing standards, involvement in continuing education or consideration of the implications of the information for improving care.⁵⁷

Integrating the lessons from the diverse literature described previously into coherent guidance for practice in the anaesthetic service area is a somewhat difficult task. Taken at face value, an effective feedback strategy should be timely, intensive, originate from a trustworthy and credible data source, be confidential and non-judgemental, and be supported by the broader organisation, supplied continuously over time and integrated within a broader quality improvement framework. Future research into the application of quality indicators in anaesthesia must take a holistic view of quality monitoring systems that incorporates the whole feedback cycle from data collection, through analysis and dissemination, to actual use of the data to improve practice.

From consideration of the range of characteristics that previous studies have suggested to be features of effective feedback systems, it is apparent that these systems are sociotechnical in nature, governed by a range of human, social, organisational and design factors.⁶¹ The organisational context and culture into which a quality monitoring programme is implemented is, therefore, clearly important to its success, especially in terms of whether or not there is open disclosure and constructive discussion on performance issues within the professional group. How a community of end-users views the system and its output, and how acceptable they consider the system to be, is likely to affect adoption and use of the technology.⁶² Performance measurement systems, in particular, are likely to be sensitive to end-user perceptions of utility, fairness and the opportunities inherent within the system for misuse of the data, from different perspectives. Such considerations require multidisciplinary study that includes clinical, health services, human factors and psychological investigation.

In terms of the effects of performance feedback on individual clinicians and clinical units, research evidence suggests that positive changes in systems and practice can result, especially where feedback from quality indicators is sustained and linked to a quality improvement framework.^{13,49,52,54} Other reviews of the evidence linked to the efficacy of professional behaviour change interventions have found that there are 'no magic bullets',⁵⁹ that dissemination alone is rarely effective and that moderate positive results can be achieved using more complex interventions. Similarly, synthesis of evidence on professional education and quality assurance interventions aimed at changing professional behaviour found that passive intervention approaches were generally ineffective and unlikely to result in behaviour change.⁶⁰ Multifaceted interventions involving educational components were found to be more likely to be effective. There is generally a lack of evidence in this area underpinning the understanding of the causal mechanisms that drive professional behaviour change in health care and a clear need exists for future work in this area to identify the effective modifiers, barriers and facilitators.⁶³

A growing body of research across a range of diverse areas suggests that selection of appropriate quality indicators and providing actionable feedback linked to quality improvement mechanisms can support detection of problem areas and timely action to improve effectiveness, efficiency, safety and the patient experience. Various models for feedback from quality indicators have been proposed and implemented, and yet there is little robust evidence for the efficacy of any one specific model or its fit within the local clinical service context. From both the health services and research perspectives, there is a need to build on

existing work on effective use of data and quality indicators to drive local service improvement and to better understand the requirements for effective quality monitoring and feedback processes at the clinical departmental level to support personal professional development among doctors.

Quality improvement in health care as an approach

In the UK and elsewhere, a number of national and local initiatives have been launched to improve care through the application of industry-derived quality improvement methods, for example Quality Collaboratives, The Safer Patients Initiative, The Productive Series and Collaboration for Leadership in Applied Health Research and Care (CLAHRC). An alternative approach to the question of how clinical units can best use data on quality of care to improve local systems is provided by popular quality improvement programme models which utilise industrial process control principles.⁶⁴⁻⁷³ Drawing on the evolving improvement science discipline,^{74,75} in the UK and elsewhere, a number of national and local initiatives have been launched to improve care through the application of industry-derived quality improvement methods.^{76,77} Such programmes adopt a Continuous Quality Improvement philosophy^{78,79} and often employ statistical process control as a measurement and evaluation model to guide improvement activities.

Process control has received some attention as a means of monitoring variations in clinical practice at the individual, unit and organisational levels within various clinical disciplines, including anaesthesia.⁸⁰⁻⁸⁴ The dominant rationale for this approach is that continuous, longitudinal monitoring of local clinical processes, in near real-time, can detect and correct significant variations in care in a more proactive manner while providing useful information on the reliability of a process over time.⁶⁶ Such information on longitudinal variation is often masked by reporting aggregated data infrequently. Plotting data from process measures longitudinally on run charts or control charts permits the detection of non-routine underlying causes of variation as giving rise to violations in control limits.⁶⁴ Special causes of variation may be identified and addressed through application of quality improvement methods to alter the underlying process in some way, with the dual aim of improving both the reliability and performance of the process.^{73,85}

Although some evidence exists supporting the efficacy of process control as an approach to monitoring and improving care,⁸⁶ reviews of quality improvement models in health care have suggested that the effects of dominant improvement models which utilise the technique are likely to be highly context-specific.⁸⁷⁻⁸⁹ The large volume of published case reports suggest that initiatives using a quality improvement model can be highly effective within a conducive local context⁹⁰ or tailored interventions.⁹¹ The penetration of statistical process control as a means of monitoring and evaluating the impact and sustainability of serial interventions, such as quality improvement initiatives, remains underutilised in health care and there is limited mention of the method in the anaesthetics research literature. Varughese reports one example of a multicomponent quality improvement initiative in paediatric anaesthesia, which utilises run charts and control charts to evaluate the local impact of improvement work.⁸⁰

Perhaps one of the most interesting features of continuous quality improvement and statistical process control as an approach is the ways in which it contrasts with conventional clinical audit as it is often applied in practice. The conventional clinical audit model may be limited in its ability to support or drive quality improvement activities, due to a range of design, organisational and other barriers.^{40,41} Discontinuous or periodic audit provides an aggregated snapshot of practice at a single time point and, therefore, cannot account for prior or baseline trends without multiple follow-up periods. As data collection is not continuous, there may be issues with the reliability of measures and the possibility of a Hawthorne effect within the teams being observed, meaning that the knowledge of being measured actually changes behaviour and poses a threat to external validity. Common experience is that audit can be a data collection exercise, undertaken by junior staff, that fails to prompt corrective action. Where action is taken, there may be a lack of follow-up measurement to evaluate the impact on care and, subsequently, to determine if any gains made have been sustained. Through embedding continuous process monitoring in routine operations following the industrial model, many of these limitations are overcome due to the

creation of a data-rich environment in which to assess current practice over time and evaluate the impact of changes to the care system. In continuous monitoring programmes, the onset and offset of data collection introduce variations due to the Hawthorne effect. Feedback over extended time periods has, additionally, proven to be more effective,⁵⁶ although this must be at the expense of considerable additional investment in data collection and administration.

Chapter 4 Research aims

Table 2 states the primary and secondary research aims that were included in the original research protocol for this project and an explanation of how they link to the existing literature in this area.

TABLE 2 Primary and secondary aims

Research aim	How will it contribute to the existing literature	Key references
To evaluate the impact of a departmental continuous quality monitoring and multilevel feedback initiative on the quality of anaesthetic care and efficiency of perioperative workflow within a London teaching hospital over a 2-year period. Data will be analysed at multiple time points over the course of the project in order to generate both formative and summative information to support development and evaluation of the initiative	<p>Test whether or not the characteristics we are using as part of our initiative are effective</p> <p>Provide research evidence to support others in developing feedback initiatives of this type</p>	Bent <i>et al.</i> , ¹² Bolsin <i>et al.</i> ¹³
To employ a quasi-experimental time-series design to provide robust evidence concerning the impact of a serial data feedback intervention on anaesthetic quality indicators while controlling for baseline variance	Evaluate the effectiveness of this quality improvement methodology for these purposes	Varughese <i>et al.</i> , ⁸⁰ Thor <i>et al.</i> ⁸⁶
To contribute to the evidence base for valid and reliable quality indicators for anaesthetic care, including effective patient experience measures	<p>Identify which quality indicators and patient experience measures are effective/ineffective for these purposes</p> <p>Identifies characteristics of effective quality indicators and patient experience measures as perceived by end-users</p>	Haller <i>et al.</i> , ¹⁷ Myles <i>et al.</i> , ²⁶ Wollersheim <i>et al.</i> , ³⁷ Bradley <i>et al.</i> ⁵⁸
To document the main features of the data feedback intervention as it develops through the CLAHRC programme for replication at other sites, including definition of metrics, data processes, feedback format and action mechanisms	Understand whether or not our intervention is effective and how others can learn from it	Bent <i>et al.</i> , ¹² Bolsin <i>et al.</i> ¹³
To assess the perceived acceptability and utility of this information system for individual end-users, the clinical department and other organisational stakeholders, using a formative, mixed-methods design	<p>Comment on how useful these data will be for the purposes of revalidation, i.e. do clinicians trust in and want to use this type of data as evidence for their fitness to practice? If not, then what data do they want/need?</p> <p>Provide a sociotechnical view of the system</p>	Bradley <i>et al.</i> , ⁵⁸ Eason, ⁶¹ Holden and Karsh ⁶²
To collaborate with relevant specialty groups within anaesthesia and perioperative care, including the Royal College of Anaesthetists, the AAGBI and the National Institute of Academic Anaesthetists, to support the agenda around revalidation and quality indicators within this specialty	Contribute to the development of quality indicators for the purposes of revalidation and the perioperative specialty agenda to develop standards and guidance for a core set of perioperative quality measures	Moonesinghe and Tomlinson, ⁶ Rubin ⁷

Chapter 5 The intervention: a continuous monitoring and feedback initiative for anaesthetic care

The intervention implemented and evaluated in this study (and as part of the CLAHRC North West London quality improvement project on which it is based) was conceived as a continuous quality monitoring and feedback initiative for anaesthetic care. Drawing on an industrial approach to quality measurement and improvement, the initiative took the form of continuous audit of anaesthetic quality indicators in the PACU (recovery room) coupled with continuous feedback of personal-level case data to consultant anaesthetists and perioperative professionals. In the later phases of the programme, basic data feedback was enhanced with broader professional engagement activities and rapid, responsive development of the feedback model in response to user feature requests, in order to increase the capability of the feedback to stimulate improvement in professional practice, learning from case experience and quality of anaesthetic care delivered by the perioperative system.

Baseline data collection of anaesthetic quality indicators began in March 2010 at St Mary's Hospital main theatre suite, with basic feedback (version 1 of the report) introduced 6 months later in October 2010. Enhanced feedback (version 3 of the feedback report) was introduced in July 2012 and ran until the end of the project in November 2013. During the course of the project, many minor iterations were made to the feedback that consultant anaesthetists received, both due to development in the available quality indicator data set and in response to specific requests and feedback from the anaesthetist group. The project developed following the CLAHRC programme quality improvement model, which draws on established improvement science in health care and advocates rapid-cycle iterative development of quality improvement solutions.

Basic feedback consisted of the provision of monthly personal data summaries in tabled form for a limited number of summary quality metrics, compared with group-level averages without adjustment. Limited longitudinal and normative comparisons were included in graphical format. In contrast, the enhanced feedback phase of the intervention employed a design rationale that was driven by more active engagement with users and with a view towards providing specific data and statistical perspectives geared towards supporting personal learning from case experience. Provision of enhanced feedback included, in addition to the basic feedback content, monthly detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data and institution-wide dissemination. During the enhanced feedback phase, engagement with the anaesthetist group was much more active, involving regular presentation of statistical results at meetings, consultative interviews by the research team for formative evaluation of the preferred features of the feedback for end-users and more focused engagement and facilitated peer interaction on specific specialty areas in which potential quality issues were identified (e.g. pain management after gynaecological surgery). During the enhanced feedback phase, the scope of data collection was increased to include multiple sites, which increased the prominence of the quality monitoring and feedback activities within the broader perioperative department across the trust as a whole.

Main approach and design rationale

Development of the feedback reports proceeded iteratively over time in response to several key categories of design rationale.

Increasing integration of the post-anaesthetic case unit form with the post-anaesthetic case unit nursing workflow and post-anaesthetic case unit patient information/data-logging requirements

The PACU nurses were required to maintain both a patient log book and, in later stages of the project, a computer-based theatre administration system database ('TheatreMan'; Trisoft, Peterborough, UK), along with an electronic patient experience tracker (tablet PC-based). These requirements were in addition to the PACU form data collection to support the Improving Anaesthetic Quality (IMPAQT) initiative. Only the last two requirements could be completed in a flexible way at the patient bedside, and it soon became apparent that the PACU nurses were using the PACU form as a flexible record of patient recovery data, which was massed for multiple patients and then periodically entered into the log book and PC-based databases. This was an efficiency measure for staff, which overcame the need to repeatedly move between the patient bedside and the PACU PC workstation in order to enter recovery data. Our project team responded by ensuring that data items and definitions on the PACU form were compatible with the other data entry requirements and to make it clear which metrics were directly comparable (duplicate items) and which were unique to the IMPAQT initiative. The design rationale was that this would both reduce workload and enhance efficiency in terms of PACU workflow, while improving the reliability of our data entry system for the IMPAQT project.

Increasing availability of comparative databases, contextual information for individual patient records, quality indicator data available and analytic capability

Development of the PACU form to include additional/revised data items and enhanced definitions and guidance, usually in response to evidence that there was variation in consistency in the way in which the scales and items were being completed or in response to an identified need for additional metrics or more fine-grained analysis. With the introduction of the electronic theatre administration system (June 2011) and subsequent iterations made to that system by trust information services, the IMPAQT project had access to additional data that were automatically uploaded from the hospital patient administration system as well as enhanced theatre workflow data. Standardised contextual information at case level, such as specialty category and procedure type, allowed finer-grained analysis of caseload and disaggregation of QoR indicators by specialty in later versions of the anaesthetist feedback report. It should additionally be noted that, at first, with the introduction of this database, the additional data entry requirements and recovery metrics that needed to be collected interfered with PACU staff's ability to accurately and reliably generate data against the original IMPAQT project items, and so the PACU form was streamlined to reduce the data collection burden, as described above, in an effort to improve data capture reliability. Subsequently, however, the availability of the TheatreMan database, along with the presence of a National Institute for Health Research (NIHR)-funded researcher dedicated to the project, provided the opportunity to both validate the PACU form-based data set on which the IMPAQT project relied and complete missing data items identified in the IMPAQT database.

Theoretically and evidence-informed practice resulting from the research effort

As part of broader research undertaken at the Imperial centre and strategic collaborations with academic partners, the IMPAQT project team was able to draw on a range of health services research theory and practical experience in order to inform the design of the data feedback programme. Research into effective forms of feedback from quality monitoring identified by a systematic review suggests that effective feedback is timely, continuous, specific to the local context, non-punitive, is accompanied by a sense of local ownership of the data and is issued by trusted/credible sources. Similarly, the longitudinal perspective adopted in the data feedback reports was informed by quality improvement methodology (specifically, theory in the area of process control), which holds that useful data for improvement are generated through sampling small units repeatedly over time rather than large aggregated data collection time points. Evolving theory and research evidence in the science of improvement suggests that data for improvement have a number of distinct characteristics that distinguish them from data for judgement, research or audit (Solberg), and the IMPAQT anaesthetists' reports were compatible with this perspective.

Responses made to feedback from the anaesthetist user group

The IMPAQT project involved close collaboration with the anaesthetist user group, and every opportunity to actively engage with the clinicians who were receiving and using the data was made. This occurred formally, through interviews with individuals and meetings with groups led by the research team, and informally through dialogue between individuals and the lead clinician for the IMPAQT project (GAR), who often received queries and feedback from the recipients of the monthly reports. The requested enhancements to the feedback reports included features such as elaborated details and contextual information for statistically outlying cases, and specific breakdown of case types/specialty to support personal case log books and to support professional appraisals. Analysis of user feature requests was a component of the formative qualitative investigation that was conducted as part of the evaluative research. Illustrative findings from this component of the work are described in the subsequent section.

User feature requests: formative input from the evaluation

In terms of future developments, 75% of respondents expressed an interest in a longer report with a more detailed breakdown that would increase the comprehensiveness of the quality metrics. Sixty-one per cent of respondents felt that in order for the reports to reach their potential they needed to contain a combination of both longitudinal and normative approaches to enable end-users to benchmark their performance both against their own baseline and within a comparable peer group. *Table 3* illustrates the main types of feature requests that were requested during the formative qualitative investigation, which involved consultation with the anaesthetist group and individuals that volunteered to be interviewed by the research team.

TABLE 3 User feature requests that were requested during the formative qualitative investigation

Requested change	Related codes from original analysis
More detail/specificity in the reports	<p>Reports can afford for data to be added as long as they are meaningful</p> <p>Feedback reports are not detailed enough to be used for revalidation and need to consider external factors</p> <p>Reports are too simplistic</p> <p>Information could be added to the report and it would not become too long</p> <p>The introduction of more complex information onto the reports would allow anaesthetists to identify what they need to change to improve</p> <p>Individuals should be able to request further information from the providers of the reports</p> <p>Measures need to be specific enough to make improvements based on them</p> <p>Request for more information on reports</p> <p>Need to be able to funnel down further in the reports to patient-specific information</p> <p>Need for more detail regarding the pain indicator</p>
Trainees to receive separate reports rather than being included under their consultant	<p>Trainee data should not be automatically recorded under the on-call consultant</p> <p>Data being recorded inaccurately under the anaesthetist who is on call rather than the anaesthetist on duty</p> <p>It would be useful to compare trainees with consultants</p>

continued

TABLE 3 User feature requests that were requested during the formative qualitative investigation (*continued*)

Requested change	Related codes from original analysis
Comparison based on case mix	<p>Metrics that take case mix into consideration need to be incorporated into the feedback reports</p> <p>Information on caseload is useful</p> <p>Information about an individual's case mix needs to be more specific</p> <p>Without reference to case mix the feedback reports are not interpretable</p> <p>People are more likely to act on comparisons if case mix is considered</p> <p>Need to understand case mix in order to make valid comparisons</p>
Data presented at an individual level over time to enable the identification of trends	<p>Anaesthetists want to see their own personal feedback over time</p> <p>Individual feedback is more useful than normative feedback</p> <p>It would be useful to be able to instantly see your own feedback over time</p> <p>Need combination of normative feedback and individual feedback over time</p>
Simplification of the key messages	<p>Summary pages are useful</p> <p>Graphs displaying caseload need to be simplified</p> <p>Reports should be as simple and to-the-point as possible</p>
A more visual approach	Preference for graphics over numbers and statistics

Main iterations to the data feedback intervention

The following sections of the report provide a detailed overview of the main developmental sequence of the data feedback component of the intervention.

September 2010: version 1.0

Initial version: the monthly anaesthetist feedback report is based on a rolling 12-month snapshot of the data ('year-to-date'). It includes summary statistics for patient demographics and average/summary scores for pain/PONV/temperature of patients in the last month on the first page. Both basic time series and distribution charts were included based on key indicators: pain, PONV (PACU form data, simple three-point rating scales based on QoR audit tool). Time series depicts monthly score for individual anaesthetist compared with the average for the department. The report also included the ward wait time (WWT) data summaries and breakdowns based on the PACU form data.

March 2011: version 1.1

Moderate iteration (interim/transitional version): this includes rankings of anaesthetists among colleagues in response to feedback on the first version of the report. It also adds a case logbook with breakdown of personal caseload by specialty type to support anaesthetists' requirements to maintain a case log record for appraisal purposes. Quality indicator score distributions for the anaesthetist group were focused on (i.e. a comparative view) in the reports, and the longitudinal (time series) charts depicting personal variation over time were dropped from this version of the report in order to accommodate the changes. These were subsequently reintroduced at the request of the research team with the view that the longitudinal view was important and compatible with the principles of measurement for improvement and process control, drawing on improvement science theory. In this version of the report, the data source switched from the PACU data set (which was maintained by an anaesthetic trainee at the time) to the theatre administration system. Pain and nausea metrics were recorded slightly differently in the theatre administration system, and pain scores in particular were now recorded using an 11-point visual analogue scale (the PACU data source was

subsequently brought into alignment). The PACU data were also removed to allow focus on anaesthetic quality indicators. QoR scores were included.

May 2011: version 1.2

Minor iteration: personally identifiable position within the group distribution. The format was as above with some additional graphical views of personal caseload by specialty and over time. Time series graphs to enable longitudinal comparison with own scores over time were reintroduced. Crucially, in this iteration, identification codes were introduced for each individual anaesthetist so that an individual could identify his- or herself and where he or she fell within the group distribution compared with his or her colleagues (who were not identifiable to the individual). This is the first version of the report that includes full comparative and longitudinal perspectives. This report additionally flagged variation in quality of anaesthesia as an issue in the introductory text, as evidenced by variations in pain scores. Initial TheatreMan data were, however, incomplete and unreliable in places. QoR score data collected in recovery were not included in this version of the report as they did not appear to show any variation across the group (i.e. a potential 'ceiling effect' in the QoR measure).

October 2011: version 1.3

Minor iteration: expansion of the data set to include data collected in day surgery and the Western Eye Hospital, to ensure that anaesthetists who did a significant proportion of their cases in these locations received adequate data coverage. There were additionally some minor changes to the way in which the descriptive summary statistics were presented. These were included within more detailed text descriptions and explanations within each relevant section of the report, rather than presented in a summary table at the start of the report, largely owing to the need to define the origin of each data source in more depth now that multiple database sources were utilised. The log book section similarly expanded to accommodate multiple data sources, and the focus of the graphical data presentations was streamlined to depict the proportion of cases that fell out of acceptable threshold (e.g. proportion of cases arriving in recovery with core temperature < 36 °C), rather than to summarise based on the statistical mean.

February 2012: version 2.0

Major iteration: introduction of detailed breakdown by site of origin and comparison of pain scores by specialty. This version of the report represents a more mature format that consolidated several key features of previous versions into a single format, including the breakdown of personal and departmental summary data by specialty, which allows the user to account for case mix to a degree (i.e. compare themselves with a score based on comparable colleagues' cases only). Key features include the development of a detailed breakdown by site of origin and comparison of pain scores by specialty. Note that this version of the report did not contain any longitudinal data views (time series charts). These were reintroduced partially in April 2012 and more fully later. This version of the report functioned as a pre-release mock-up of the intended final report design, developed with formative input from the evaluation and designated 'version 3'. Version 2 was, therefore, an early release designed to gauge end-user reactions and opinions concerning the new format. The final version 3 report was implemented in June 2012.

The main structural sections and metrics included in this version of the report are as follows:

- cover sheet and message
- 12-month personal caseload by anaesthetist and site
- caseload breakdown by gender, age and American Society of Anesthesiologists (ASA) score (ASA's classification of fitness of cases before surgery)
- percentage of cases by specialty
- distribution of anaesthetists' patient temperature data (proportion of cases over 36 °C)
- distribution of anaesthetists' patient nausea data (proportion of cases without PONV)
- distribution of anaesthetists' patient pain data (proportion of cases with visual analogue pain scale scores below 4)
- proportion of anaesthetist's patients in pain by specialty, compared with departmental proportion.

April 2012: version 2.1

Minor iteration: as above but with addition of two longitudinal time series charts based on pain data.

June 2012: version 3.0 (final report implementation)

Major iteration: previous version consolidated and addition of a number of enhanced functions and features requested by anaesthetist group in response to exposure to various version 3 candidates since February 2012.

Version 3 of the anaesthetist feedback report represented a large development effort with enhanced input from the research team and from formative data from the evaluation project. Version 3 represented the consolidation of key features from previous iterations along with intensive development of the underlying database structure, Excel macro for report generation (Microsoft Excel, Microsoft Corporation, Redmond, WA, USA) and addition of numerous enhanced data views and features derived from qualitative analysis of the requests and requirements of end-users interviewed as part of the evaluation project in early 2012. Crucially, version 3 of the report included full cross-sectional (peer comparison) statistics for all quality indicator measures, full longitudinal time series graphs tracking trends over time, flagged cases and contextual details to allow identification of specific case instances that violated set thresholds for quality indicators, and comprehensive specialty and site comparisons, to allow adjustment for case mix and selection of a comparable peer subgroup.

Version 3 report structure:

- 12-month caseload breakdown by anaesthetist and site (Surgical Innovation Centre, Western Eye Hospital, day surgery unit, Queen Elizabeth The Queen Mother Hospital).
- Annual logbook of cases broken down by surgical specialty as recorded in Theatre administration software.
- Patient ASA score, age and gender breakdown.
- Proportion of all patients with temperature over 36 °C in last 12 months (cross-sectional comparison with peer group and longitudinal data over time by month). Breakdown by site over last quarter. Flagged cases and case details of patients that fell below threshold for temperature.
- Proportion of all patients free from PONV in last 12 months (cross-sectional comparison with peer group and longitudinal data over time by month). Breakdown by site over last quarter. Flagged cases and case details of patients that fell below threshold for PONV.
- Proportion of all patients with pain below score 4 in last 12 months (cross-sectional comparison with peer group and longitudinal data over time by month). Breakdown by site over last quarter. Comparison of pain scores with peer group by specialty (i.e. case mix adjusted). Flagged cases and case details of patients that fell below threshold for pain.
- QoR Scale score distribution by anaesthetist.

July 2013: version 3.1

Minor iteration to version 3 format and site categories to reflect inclusion of data from other non-St Mary's Imperial sites.

October 2013: version 3.2

Minor iteration to the way data were recorded in TheatreMan in response to requests to distinguish between cases in which a consultant was present and cases in which they were 'supervising consultant'. No major changes to report format.

Chapter 6 Qualitative evaluation

This chapter reports the rationale, methodology and findings from the qualitative work stream of the evaluation. The chapter is divided into three key sections. Section 1 contains the qualitative process evaluation of the feedback initiative. The second section is a critical appraisal of the applicability of a number of pre-selected theories to the mechanisms of data feedback effectiveness. Finally, in section 3 we present core conclusions and recommendations from the work stream as a whole. *Box 1* contains all key research aims from the original project protocol, with those of greatest relevance to the qualitative component presented in bold.

The qualitative work stream of the project addresses these aims by providing a comprehensive requirements analysis for the sociotechnical design of the system based on multiple end-user and broader stakeholder perspectives. These outcomes provided formative input to the development of the feedback reports throughout the life of the project. The interviews assessed the perceived acceptability and utility of the information system for individual end-users, the clinical department and other organisational stakeholders. They also explored perceptions of the impact of the programme on local culture and attitudes towards quality of care and quality improvement including any associated organisational change issues. This component of the evaluation enabled us to qualitatively explore the mechanisms by which individuals and groups use the data to change practice by capturing specific narratives and use case scenarios. Alongside this it also supported the identification of key barriers and enablers to the successful development, implementation and utilisation of this type of quality monitoring and feedback system within a specific service context.

BOX 1 Primary and secondary aims from original protocol

Research aims

To evaluate the impact of a departmental continuous quality monitoring and multilevel feedback initiative on the quality of anaesthetic care and efficiency of perioperative workflow within a London teaching hospital over a 2-year period. Data will be analysed at multiple time points over the course of the project in order to generate both formative and summative information to support development and evaluation of the initiative.

To employ a quasi-experimental time series design to provide robust evidence concerning the impact of a serial data feedback intervention on anaesthetic quality indicators while controlling for baseline variance.

To contribute to the evidence base for valid and reliable quality indicators for anaesthetic care including effective patient experience measures.

To document the main features of the data feedback intervention as it develops through the CLAHRC programme for replication at other sites, including definition of metrics, data processes, feedback format and action mechanisms.

To assess the perceived acceptability and utility of this information system for individual end-users, the clinical department and other organisational stakeholders, using a formative, mixed-methods design.

Qualitative process evaluation of the initiative

Introduction

In this section we report the qualitative process evaluation of the quality monitoring and feedback initiative. The purpose of the evaluation was to assess whether or not clinicians would engage with an initiative of this nature, whether or not they would find the quality indicator data provided useful and whether or not they would use the feedback to change their professional practice. Given the broader specialty focus on clinician revalidation and establishing quality indicators for anaesthesia, understanding the causal mechanisms by which an intervention of this type might result in changes to practice and outcomes is an important aim and precursor to further evaluative work. From a practical point of view, it can additionally provide important information concerning the role that the local context for implementation plays, which might influence the degree to which the intervention will generalise to other contexts. If the professional culture within a clinical department is reflective, with open and constructive sharing of past performance data in order to learn and strive for excellence, then the effects of any feedback initiative are likely to be enhanced compared with a unit which is characterised by closed or punitive dialogue on variations in care. Such an approach has been termed 'process evaluation' in the improvement science literature and advocates investigation of social, cultural and organisational factors in addition to effective intervention design.⁹²

Methods

Design

The design was a longitudinal, qualitative work stream which ran parallel to the intervention work and took a realist evaluative perspective on the project. The realist position provides a framework for identifying not only what outcomes are produced by an intervention, but how they are produced and how the intervention interacts with varying local conditions to produce the outcomes. Complex, multifaceted interventions are subject to phased implementation and intensive iteration. We therefore expected effects to be serial and cumulative over time. We also expected levels of engagement and impact of the initiative to vary as a result of ongoing interaction with contextual and organisational preconditions.

Participants

Participants for the interviews were perioperative service leads (including the lead nurse for the PACU), consultant anaesthetists and surgical nursing leads from the primary site at which the initiative took place. Potential respondents were approached to give feedback to assist in developing the programme and as users of the information it provided. These perioperative service stakeholders represent an array of actors in the processes of the previously described feedback initiative and enabled us to explore the different mechanisms of change that were under way. Forty-four consultant anaesthetists were contacted to participate along with specific perioperative service lead, surgical ward lead and recovery nursing roles that were selected based on professional position and level of expertise.

Data collection

Interviews were conducted by a research team including one senior social science researcher, two research associates in quality and safety, and two clinicians in training undertaking a research placement. Interviews determined the perceived value of specific quality indicators in anaesthesia and impact of feedback design. Furthermore, factors optimising engagement with the initiative were investigated as well as the mechanisms by which data were used to create behaviour change. Multiple interview schedules were used by the research team (dependent on the professional role of the participant, the time point at which they were being interviewed and their previous participation in the project). *Table 4* provides a simplified overview of the topic areas with example questions covered. The research team received strong clinical input into the design of the interviews. The initial interview schedule was piloted with a senior consultant anaesthetist using a cognitive walkthrough technique in which the interviewers' questions were first answered, and then discussed in depth in terms of wording, relevance and duplication. *Appendix 1* reproduces the final interview schedule for consultant anaesthetists at time point 1 and *Appendix 2*

TABLE 4 Simplified overview of interview topics with example questions asked to interviewees

Topic	Example questions
General views on feedback	In your view, what are the most important aspects of quality of care relevant to anaesthetics practice? Do you think anaesthetists/PACU staff/ward staff generally get adequate feedback on these aspects of quality of care?
Evaluation of the current initiative	What are your general thoughts about this initiative and the feedback reports that you receive? What was your initial reaction to seeing your data? How do you use the information contained within the reports?
Departmental perspective	What is the potential value of this initiative to the department? How do you think the department itself should use the data?
Project stakeholder questions	What are the implications of this initiative for the anaesthetics specialty? Can you see a role for initiatives of this type in revalidation?
Future development	Are there any measures, features or functionality that you would like to see included in future versions of the reports? What further support could be provided for anaesthetists/PACU/wards to use these data to improve care?
Broader context	Do you see any barriers to engagement with and utilisation of this initiative? Is there anything about the organisation or context in which you work that might make a system like this one more or less successful? Do you think there is an atmosphere of transparency here among the clinical group/nursing directorate concerning quality?
Longitudinal component	Have your views about the feedback reports changed or developed in any way over time? Have your perceptions about anonymity changed over time?

reproduces the final interview schedule for consultant anaesthetists at time point 2. The research team engaged in ongoing reflexivity throughout the data collection process. This involved individual researchers continuously reflecting on and discussing their own personal influence on the interviews that they were conducting and how this impacted on their understanding and interpretation of the data. The fact that the interviewers were already engaged in the operational process of delivering feedback to end-users raised potential issues of subjectivity and bias. This was counteracted by ensuring that multiple researchers engaged in the interview process and that any arising issues associated with the action research style approach to the project were discussed and reviewed at regular steering group meetings.

Analysis

Analysis was led by a research associate in quality and safety and took an inductive approach, informed by principles of grounded theory and social constructionism.^{93,94} This enabled themes to emerge naturally from the data without the influence of prior theory. Interview recordings were transcribed, read and reread until the data were familiar. The transcripts were then open-coded into units of meaning using NVivo software (version 10, QSR International, Warrington, UK). Units of meaning were later coded and grouped into broader themes and subthemes. The emerging qualitative template was reviewed and discussed with strong clinical input from three consultant anaesthetists and a junior doctor, as well as ongoing academic input from the senior social sciences researcher. Results were presented to the project steering group on two occasions to gain further senior academic perspectives on the work, and findings with high relevance to the development of the feedback reports were regularly sent to operational leads. A number of

iterations were developed until no new categories of meaning were derived. Content analysis was also employed where appropriate to quantify responses.⁹⁵

Results

Interviews lasted between 30 and 60 minutes each. In total we analysed approximately 22 hours of interviews with 24 consultant anaesthetists, six surgical nursing leads and five perioperative service leads.

Content analysis

The results of the content analysis are displayed in *Table 5*. Results are shown in figures as well as percentages to accommodate the fact that not all participants responded to all questions. Ninety-five per cent of respondents felt that they did not generally get systematic and objective feedback on the quality of care that they provided. However, 57% of respondents described specific changes that they had made to their practice as a direct result of the current initiative, and 90% of respondents agreed that the reports had strong potential for use as part of the upcoming revalidation process.

In terms of future developments, 75% of respondents expressed an interest in a longer report with a more detailed breakdown that would increase the comprehensiveness of the quality metrics. Sixty-one per cent of respondents felt that in order for the reports to reach their potential they needed to contain a combination of both longitudinal and normative approaches to enable end-users to benchmark their performance both against their own baseline and within a comparable peer group.

Formative evaluation (feature requests)

Throughout the interviews a number of requests were made for iterations to the feedback reports. These were fed back to operational leads and where possible were integrated into the development of the project. *Table 6* lists the requested changes with links to their original qualitative codes and example quotations from the time point 1 analysis.

Main analysis against research aims

The results of the process evaluation are presented below with reference to the key qualitative research aims and questions that we set out to address in the original protocol. These themes emerged inductively from the data and provide a realist perspective on experiences of receiving feedback with reference to the influence of the local context. The analysis presented below represents three levels of thematic analysis, with 58 low-level codes identified, which were subsequently structured into 26 mid-level categories and finally subsumed within five high-level themes for reporting purposes. An overview of the coding and

TABLE 5 Overview of quantitative results with frequency breakdown by category

Question	Response
Do you think anaesthetists generally get adequate feedback on these aspects of quality of care?	Yes: 1 (5%)
	No: 18 (95%)
Are there any examples of changes you have made to your practice?	Yes: 8 (57%)
	No: 6 (43%)
Would you rather see your data compared with others', your data displayed over time, or both?	Normative: 1 (8%)
	Individual: 4 (31%)
	Both: 8 (61%)
Would you be interested in a longer report with a more detailed breakdown/analysis of your data?	Yes: 9 (75%)
	No: 3 (25%)
Can you see a role for initiatives of this type in revalidation?	Yes: 9 (90%)
	No: 1 (10%)

TABLE 6 Requested changes evident in time point 1 process evaluation

Requested change	Related codes from original analysis	Example quotations
More detail/specificity in the reports	Reports can afford for data to be added as long as they are meaningful	<i>I think the, as a number it's quite a good reflection of patient experience but it's not necessarily specific enough for me to be able to address anything in my practice to do something about it. Like, it's not that it's a bad thing that I try and improve all the different components of the Quality of Recovery scores</i> <i>Anaesthetist 10, time point 1</i>
	Feedback reports are not detailed enough to be used for revalidation and need to consider external factors	
	Reports are too simplistic	
	Information could be added to the report and it would not become too long	<i>I think it's far too generalistic and too simplistic to be useful to me</i> <i>Anaesthetist 3, time point 1</i>
	The introduction of more complex information onto the reports would allow anaesthetists to identify what they need to change to improve	<i>What would be a value as a first step is to take those factors that we absolutely know to be related to PON-V and try and equate for them. So if the expected incidences of nausea and vomiting in my patient population from the available literature is 25%, and mine is 20%, I'm good. If it's 15% and I'm 20% I'm bad. So the absolute number is a good indicator but with more information goes more complexity. So the simple stuff would be look at a few things known to cause nausea and vomiting. The more complex is how does my personal technique affect that</i> <i>Perioperative service lead 3, time point 1</i>
	Individuals should be able to request further information from the providers of the reports	
	Measures need to be specific enough to make improvements based on them	
	Request for more information on reports	<i>I think anything which provides a talking point is useful. So anything which gives you another variable to look at would be useful</i> <i>Anaesthetist 5, time point 1</i>
	Need to be able to funnel down further in the reports to patient-specific information	
	Need for more detail regarding the pain indicator	<i>Where you've got pain scores individual. So like I had one case or something that was really really sore and I didn't remember that. So maybe if they had been red-flagged I could have got the notes and had a look back and thought that was the one who screamed at everything, or something</i> <i>Perioperative service lead 1, time point 1</i>
		<i>When you say more of my own data, you mean more detailed, more of a breakdown of it? Yes I am always keen to see anything like that, yes. I am always a bit of a numbers man like that, I am always keen to see it</i> <i>Anaesthetist 1, time point 1</i>
		<i>Yeah, I think . . . I'm not sure how you'd do that but if you just gave the hospital number or something we could kind of look them up and go, 'Oh yeah it was that person', or whatnot but . . . Because I was surprised because generally the problem cases you remember</i> <i>Perioperative service lead 1, time point 1</i>

continued

TABLE 6 Requested changes evident in time point 1 process evaluation (*continued*)

Requested change	Related codes from original analysis	Example quotations
		<p><i>I think I've, I think I've probably reached a plateau with this feedback now and I almost need more from it to be able to change my practice in the future. So from what I've got here I've done what I can to change it, and I seem to be reaching a stable plateau. So really I'd be more interested in, of these pain patients, which patients it is that I really need to look at now. Are they stray patients or is there a trend in the patients that are painful? And is it all of the men that I anaesthetise that are painful or is there a particular procedure or something?</i> <i>Anaesthetist 10, time point 1</i></p>
		<p><i>I'd like to see procedure-related pain scores on here so I could identify of the 40 patients last month that had an average pain score or whatever, which ones were the ones that were sore. Because if it's just one particular type of operation that I do that's sore then I can't get that out of this information as it currently is</i> <i>Anaesthetist 10, time point 1</i></p>
		<p><i>You want more information about why they're in pain and was it dealt with and did it go away, that sort of thing</i> <i>Perioperative service lead 1, time point 1</i></p>
		<p><i>It is not reflective of my practice because half the time I may not even know the patient is being anaesthetised, if you see what I mean. If I am at home or on call and they are anaesthetising it, they are doing their standard anaesthetic as a trainee, and I might not even know the trainee! So it is probably not strictly fair</i> <i>Anaesthetist 6, time point 1</i></p>
		<p><i>Yes, when I look through the list and I look at what cases I have done, I see I have done quite a few, for example gynaecology cases, and other cases. What I think is happening is they are probably getting done on nights I am on call, I think, I am not sure, but I guess that is what is happening. So they are put under me, although I didn't actually anaesthetise them. That is the only thing I can think of. I don't know!</i> <i>Anaesthetist 6, time point 1</i></p> <p><i>I'd quite like to see it to include the registrars as well, so not just the consultants. Which would be a bit more variable, obviously, but I think it would be quite interesting for people</i> <i>Perioperative service lead 2, time point 1</i></p> <p><i>No, I think it could be useful. I mean, if it's possible to do, it could be useful and interesting. I think people might argue that trainees do a slightly different case mix to consultants and different kinds of cases, but it might be useful. I think anything which provides a talking point is useful. So anything which gives you another variable to look at would be useful</i> <i>Anaesthetist 5, time point 1</i></p>

TABLE 6 Requested changes evident in time point 1 process evaluation (continued)

Requested change	Related codes from original analysis	Example quotations
Comparison based on case mix	Metrics that take case mix into consideration need to be incorporated into the feedback reports	<i>You could. There is all sorts of things, I mean there is the EURO-QUAL score for example, those sort of things. EuroSCORE which I think gives a reflection of the complexity of the case and maybe something along those lines, you could build into this system which would be interesting</i> <i>Anaesthetist 6, time point 1</i>
	Information on caseload is useful	
	Information about an individual's case mix needs to be more specific	
	Without reference to case mix the feedback reports are not interpretable	<i>I think if you did bring in some measure of the difficulty of the sort of cases, I think that would be interesting. I don't think ASA is good enough, though</i> <i>Anaesthetist 6, time point 1</i>
	People are more likely to act on comparisons if case mix is considered	
	Need to understand case mix in order to make valid comparisons	<i>Yeah . . . these ones might just be paediatrics or something, because paediatrics always seem to wake up screaming regardless of what you do with them. I don't know, so it might be useful to kind of know what these people do, but then you've kind of got to be careful to maintain confidentiality I suppose, otherwise it just gets . . .</i> <i>Perioperative service lead 1, time point 1</i>
		<i>But you don't know what these ones lower are do you, major abdominal surgery or something like that, they have a bit more pain</i> <i>Anaesthetist 1, time point 1</i>
	<i>I'd like to see procedure-related pain scores on here so I could identify of the 40 patients last month that had an average pain score or whatever, which ones were the ones that were sore. Because if it's just one particular type of operation that I do that's sore then I can't get that out of this information as it currently is</i> <i>Anaesthetist 10, time point 1</i>	
	<i>Yeah, and maybe look at their case-mix as well. For instance, if someone was doing an awful lot of trauma cases, then it's quite difficult to get them warm if you've got a surgeon, or if you've got them bleeding out in front of you. So that could skew the data, but you can only interpret that intelligently if you have the case-mix in front of you</i> <i>Perioperative service lead 2, time point 1</i>	
	<i>I think you need to – well, you probably need to tease this out a bit more. So, I mean, obviously patients who are short cases aren't necessarily going to get cold. So if you've come out as tops for your patient staying nice and warm but you only do day-care short cases, that doesn't necessarily mean that if you did a different type of case you would be as good. I don't really feel that I can make a tremendous amount of conclusions about whether my practice is good or bad</i> <i>Anaesthetist 3, time point 1</i>	

continued

TABLE 6 Requested changes evident in time point 1 process evaluation (*continued*)

Requested change	Related codes from original analysis	Example quotations
		<p><i>... if I had five aneurisms and I compared all my aneurisms to, you know, what I've done in the past all the other aneurisms, and that would be actually an interesting, or my tonsils compared to everyone else's tonsils. I think that would be much more interesting than a generic comparison against everyone and everything and every patient</i></p> <p style="text-align: right;"><i>Anaesthetist 4, time point 1</i></p>
		<p><i>I think this kind of data is globally OK, it's of interest, but it doesn't give me as much either, incentive really because it's all very interesting, OK, this month I'm eight, next month I'm two, next month I'm 10, you know, maybe it's got to do with actually I was away on holiday for most of the time. And I only anaesthetised two people compared to 40 people like I normally do. So, do you know what I mean? Just it's, it's the power of the information is just how it varies, it's the same if you compare, if to get all my aneurisms and everyone else's aneurisms and I found that I was actually the, you know, all my patients, that I'm the worst. Then I would actually then, that would incentivise me more to go and see how everyone else does it, how to improve it, etc., etc.</i></p> <p style="text-align: right;"><i>Anaesthetist 4, time point 1</i></p>
		<p><i>It might do. In fact, for instance, if I do children's ENT [ear, nose, throat] procedures, or children's urology procedures, then it might be useful to compare it to other people who also do children's urology procedures and see how we compared, so we've got very like with like. That might make it easier, but I accept that it's quite a difficult one to do</i></p> <p style="text-align: right;"><i>Anaesthetist 5, time point 1</i></p>
		<p><i>I think it should be adjusted according to complexity of patients because I think it is a bit unfair otherwise. Because otherwise people who do day cases all the time are clearly going to be amongst the best anaesthetists in the department, do you know what I mean? And I think that is an issue. So I think maybe you need to look at that</i></p> <p style="text-align: right;"><i>Anaesthetist 6, time point 1</i></p>
		<p><i>Well, I'd love to compare myself to somebody doing a very similar list, and if I'm worse than they are then obviously I'd be very upset about it [laughs]. But I would probably do something about it</i></p> <p style="text-align: right;"><i>Perioperative service lead 2, time point 1</i></p>

TABLE 6 Requested changes evident in time point 1 process evaluation (continued)

Requested change	Related codes from original analysis	Example quotations
Data presented at an individual level over time to enable the identification of trends	Anaesthetists want to see their own personal feedback over time	<i>So you're, I'm surprised, you can't compare, so, and I think the only thing would be an issue with the cases that you're given month by month but you can't compare, as you say, your own data</i> <i>Anaesthetist 4, time point 1</i>
	Individual feedback is more useful than normative feedback	
	It would be useful to be able to instantly see your own feedback over time	<i>So if I'm number 10 I can't see over March, June, July, August, September how it, kind of, changes over the period of time</i> <i>Anaesthetist 4, time point 1</i>
	Need combination of normative feedback and individual feedback over time	<i>I would like to compare my practice and then, if I've made a change, how it impacts on my work</i> <i>Anaesthetist 3, time point 1</i>
		<i>Not recently, no, no. And actually, that's where it might be quite good to have the previous reports' scores in there somehow. Yeah, that would be good. OK</i> <i>Perioperative service lead 2, time point 1</i>
		<i>I suppose what we lose is the opportunity with this to reflect over how have we done over the last few months. How has each month compared on month on month? As a cohort, has there been any change? I'm not sure if this helps to identify that. It probably does, but I don't think we present that. That would be useful to know, whether the whole department scored differently last month in one thing. That should be quite useful</i> <i>Anaesthetist 5, time point 1</i>
		<i>I think that it would be helpful to possibly have a thing each month, like showing you how you've done over the months. Without referring back to the previous month, you don't quite know that you've changed from the month before</i> <i>Anaesthetist 5, time point 1</i>
	<i>I would be happy that changing something will just get the trend going better. But of course if it goes 100% that would be very good but I think the importance is to have monthly reports, related to the fact that you can see whether you are getting better with what you changed or not. If you still need to change something more or is it enough?</i> <i>Anaesthetist 2, time point 1</i>	

continued

TABLE 6 Requested changes evident in time point 1 process evaluation (*continued*)

Requested change	Related codes from original analysis	Example quotations
Simplification of the key messages	Summary pages are useful	<p><i>No. I think that this feedback needs to be in the most concise easy-to-read version so someone can open it as an e-mail, flick through it and identify if there are any areas that need attention and if so, which areas and what they can do about it. And I think anything complicated would just stop people from opening it</i></p> <p style="text-align: right;"><i>Anaesthetist 10, time point 1</i></p> <p><i>I think this one's too wordy, I think, I don't think we've got this log book stuff right. It's now going to come out as a column chart, I don't know that we need to be looking at the number of procedures in a month and the previous month, I think we can just look at annual data and people know that they've got relatively, that they've got a relatively consistent workload month in, month out. And I think I might try and get rid of some those columns so that what is there becomes easier to interpret</i></p> <p style="text-align: right;"><i>Anaesthetist 10, time point 1</i></p>
	<p>Graphs displaying caseload need to be simplified</p> <p>Reports should be as simple and to the point as possible</p>	
A more visual approach	Preference for graphics over numbers and statistics	<p><i>It depends what we are looking at, I suppose. It depends, charts for some and numbers for others. It is difficult. Charts are always nice to . . . they are straightforward to look at, aren't they, I suppose. Charts I would say then, yes</i></p> <p style="text-align: right;"><i>Anaesthetist 1, time point 1</i></p> <p><i>First of all I like the fact that it is a graphic, it is not a number. I found this much more effective</i></p> <p style="text-align: right;"><i>Anaesthetist 2, time point 1</i></p>

thematic structure is provided below, before commentary and example quotations for each of the five high-level thematic areas.

- Value of feedback for clinicians.
 - Need for this feedback initiative – why is it important?
 - This project represents the first step towards effective feedback on anaesthetic care.
 - Anaesthetists want to deliver high-quality care for their patients.
 - Importance of measuring patient experience and satisfaction.
 - It is important that anaesthetists receive feedback on the quality of care that they are providing to their patients.
 - Levels/existence of feedback before initiative begun.
 - It is not standard practice to quantify how patients recover after anaesthetic.
 - Anaesthetists at this trust generally did not receive feedback.
 - Conceptualisation of ‘the good anaesthetist’.
 - Vision of anaesthetic practice – anaesthesia viewed as a form of art.
 - Using feedback was associated with professionalism.
 - Role of efficiency versus quality.
- Selection of quality indicators and reporting format.
 - Conceptualisation of quality of care.
 - Suggested additional/alternative metrics.
 - People have different views on what quality of care in anaesthetics is.
 - Quality of care covers a broad range of factors and some of them are very difficult to capture/measure.
 - Specificity of feedback.
 - Request for more information on reports.
 - Meaningfulness of data.
 - There is a need for a greater focus on outcome measures.
 - Ability to control outcomes of quality indicators.
 - Importance of nausea as a quality indicator for feedback.
 - Importance of pain as a quality indicator for feedback.
 - Trust in the metrics.
 - Nausea measure needs increased accuracy.
 - Data are not always linked to the correct anaesthetist – data quality issues.
 - The way that we measure pain is subjective.

- Format of reports/data.
 - Comparisons would be more useful if case mix was considered.
 - Need combination of normative feedback and individual feedback over time.
 - It would be useful to be able to instantly see your own feedback over time.
 - Need for anonymity.
 - In favour of normative feedback.
- Presentation of reports/data.
 - Preference for graphics over numbers and statistics.
- Application of feedback to departmental quality improvement and professional behaviour change.
 - Role of the department – department-level involvement in the initiative.
 - There is a practical function to feedback for service managers.
 - Contrast between quality improvement and performance management
 - Data must be identifiable at some level if they reflect potential patient safety issues.
 - Severe outliers need to be dealt with via governance procedures.
 - Reports should not be viewed punitively.
 - Case mix needs to be incorporated in order to use feedback reports for any type of performance management.
 - Ease of translation from data to improvement/role of feedback for quality improvement.
 - Variation on judgement as to when an improvement is necessary.
 - Examples of feedback in action – practical examples of improvement linked to feedback initiative.
 - Conceptualisation of the improvement process.
 - Conceptualisation of own performance.
 - Feedback reports provide reassurance to anaesthetists.
 - People generally think that they are performing better than they actually are.
 - Feedback reports quantify/objectify an anaesthetist's conceptual understanding of their own performance.
 - Affective reaction to receiving feedback – how do people feel about the feedback that they are receiving?
 - Having feedback reports increases the motivation of anaesthetists to improve quality of care.
 - People are comfortable with the collection of performance data.
 - Anaesthetists care about their feedback reports and want to do well on them.
 - Cognitive reaction to feedback – how do people think about the feedback that they are receiving?
 - Feedback reports promote thoughts about practice and potential improvement.

- Need for additional support/involvement.
 - Anaesthetists need further support translating feedback into improvements.
- Practical application of feedback reports.
 - Feedback reports useful for revalidation/appraisal.
- The context for feedback initiatives.
 - External influences on QoR.
 - Other members of the team influence QoR.
 - Influence of specific operation on QoR.
 - Effect of patient factors on QoR.
 - Demands of feedback initiative on time and resources.
 - Maintaining the feedback increases workload.
 - Effect of the individual on the perception of feedback. Personal characteristics of the recipient.
 - Feedback reports serve different purposes for different people.
 - Some people will not even look at the feedback data that they are being provided with.
 - Effect of levels of transparency within the organisation.
 - There is currently a high level of transparency in relation to this project in the organisation.
 - Methodological issues surrounding feedback initiative.
 - Issues surrounding the effect that sample size (i.e. the number of cases that each anaesthetist does) has on feedback reports.
- Cumulative and serial effects of the intervention
 - Emotional impact of feedback decreases.
 - Clinicians develop more positive attitudes towards receiving feedback.
 - Initial scepticism fades.
 - Increase of data to work with.
 - Benefit finding.
 - Iterations made to feedback reports.
 - Ongoing data quality issues.
 - QoR score data.
 - Need for future developments.
 - Need to ensure that data are collected at the right time.
 - Requests for 6-monthly/yearly summary data.
 - Requests for improved benchmarking.
 - Need for greater organisational transparency to surround the initiative.

The value of feedback for clinicians

This section has been aggregated from three mid-level categories and nine low-level codes. It comments on the overall perceived utility of having this type of information system and its acceptability to end-users in its initial format. Categories centred on the availability of feedback before the initiative was introduced, and the interaction between receiving feedback and conceptualisations of professional identity, provide further explanation of the extent to which the initiative was of value.

Interviewees were asked about their general perceptions of the value of providing routine feedback on quality of care to health-care professionals. They unanimously stated their support and agreement for the motivating principles underlying the initiative; that there was a need to monitor and provide intelligence on current quality of care to the responsible health-care professional in a timely and useful way:

I think it is very important because you really don't know, you walk away, you don't know whether the patient is vomiting after half an hour, and is back in theatre, nobody really tells you so I discover sometimes after that my patient actually was sick because I don't see him being sick once I wake him up and I walk away.

Anaesthetist 2, time point 1

We've got access to data now; we know how long it takes for every single patient to be collected from recovery and I can communicate to staff and investigate any issues surrounding, you know, any delays.

Surgical nursing lead 2, time point 1

The main reason that the clinicians found the initiative of value was because they saw it as facilitating improvement:

No, it's brilliant; and I think feedback is very important for us to improve and look back on our practice and to change things that aren't working properly.

Perioperative service lead 2, time point 1

I have no qualms with it being used because if we haven't got the information and the evidence then how can you improve? So no, it needs to be done and I hope it carries on . . .

Surgical nursing lead 5, time point 1

It was also associated with the concept of being a 'good' clinician. The effective use of the feedback to change behaviour and make improvements was associated with professionalism in the sense that it represented acting on the needs of your patients in a systematic and rigorous way:

It is professionalism taking into account that if you don't treat pain properly, you probably need to do something more.

Anaesthetist 2, time point 1

The whole point that we're here is to improve things and to make the patient flow, patient pathway, patient experience much better.

Surgical nursing lead 1, time point 1

The significance of the initiative was often linked to the fact that levels/existence of formal feedback before the project began were extremely low. In fact, 95% (18 out of 19) of respondents felt that they did not generally get systematic and objective feedback on the quality of care that they provided. Instead, interviewees

reported having to rely on anecdotal feedback from patients and informal discussion with colleagues. This did not provide them with the opportunity to modify their behaviour based on accurate and reliable information:

There's been no history of individualised feedback, so having data that relates to my own practice is phenomenally useful.

Anaesthetist 10, time point 1

Well yeah, they give me the actual times which I wouldn't be 100% aware of if it wasn't for the reports.

Surgical nursing lead 6, time point 1

In that sense the project was framed as a first step towards effective feedback on perioperative care and therefore represented a change in itself:

My intense support for this project is the fact [that] this is a start. This is showing it can be done and we can build on this and create much greater things for the future.

Perioperative service lead 3, time point 1

Selection of quality indicators and reporting format

This section has been aggregated from six mid-level categories and 17 low-level codes. It comments on end-user preferences for the selection of metrics and the presentation of feedback reports. Qualitative categories emphasise the importance of specificity and meaningfulness when providing feedback that accurately represents clinicians' conceptualisations of quality of care in anaesthesia.

Clinicians discussed characteristics of monitoring and feedback that increased its usability and effectiveness. It was evaluated that quality indicators and the data that they provide should be meaningful and trustworthy in order to increase engagement. It was emphasised that there is individual variation and ambiguity in definitions of quality in the area of anaesthesia:

Quality of care with anaesthetics depends on who you talk to . . . So its quality very much depends on the, you know, beauty is in the eye of the beholder, and that's very true of quality.

Anaesthetist 10, time point 1

The view was put forward by a number of interviewees that 'quality of care' covered a broad range of areas and some of these were extremely difficult to capture and measure effectively. Clinicians suggested a number of factors that were not being measured as part of the initiative but that they perceived to be relevant and comprehensive:

I mean these are the things that are easy to overlook because people think it is not important, it is not going to cause a major problem to me but it is just something patients remember afterwards, when they go home and they say 'oh I had a splitting headache'. It is like constipation, that is the thing that bothers them for the following week. We get them out of theatre and out of the recovery and then we think that they are OK but then they have headaches, sore throats or constipation, these little things that patients remember, so anything that we can do to improve that makes a big difference I think.

Anaesthetist 1, time point 1

Maybe I'd like to see how many patients were called for and weren't ready. That would be quite valuable from my point of view. I don't know whether you can do that or not.

Surgical nursing lead 3, time point 1

Interviewees suggested that future feedback should be developed to report on non-technical as well as technical skills and highlighted a need for measures of the quality of preoperative as well as postoperative care to provide a more holistic representation of the care received:

I think one thing they don't perhaps assess is more . . . they do assess a clinical thing and if the patient having difficulty with pain, and if they have a temperature, but not very much about your communication skills, or how you communicate in the team so I think that's something that might also be useful how you lead your team in theatre, how the people perceive you as the team leader or the team member, because that could also be quite useful.

Anaesthetist 7, time point 1

The third thing is that I think some anaesthetists are actually very slow [laughing] in actually giving an anaesthetic and they fuff around and they can't decide which type of area to use and although it doesn't directly affect the patient, it affects the working environment. Yeah? I don't know if that enters into sort of quality of care, I think it does because it is not a calm environment for the patient when they come in, because they feel the tension. Because the nurses don't know what the anaesthetist is doing, the anaesthetist is taking ages with the anaesthetic.

Anaesthetist 9, time point 1

Interviewees highlighted the importance of being able to control the outcomes of the quality indicators that are being monitored in order to maintain motivation to engage with the project. They wanted to be able to identify differences in the data when they modified their behaviour, and emphasised the significant role of perceived improvability in encouraging changes:

Whatever you do sometimes they are still sick, although I think we can make a difference to it.

Anaesthetist 1, time point 1

The current indicators of PONV and pain were perceived to be both meaningful and important quality indicators because of the insight that they provided to anaesthetic care:

So from that point of view, yeah, I think actually the one that nags me the most is this one, because I don't think anyone should wake up sore.

Perioperative service lead 1, time point 1

Yes. I've found post-op nausea and vomiting as a very clear outcome and it's got very clear, it's got a very clear treatment to control perioperatively. So it's very easy to know what to address to improve it.

Anaesthetist 10, time point 1

However, concerns were raised around the measurement of both of these metrics. Issues of subjectivity in interpretation were raised alongside a need to consider the psychological component of pain and nausea perception. Therefore, the reliability of these measures as a basis for improvement actions was questioned:

And it's about what people expect. And so if they expect it to have no pain whatsoever and they had a bit of pain, now that's a catastrophe. But if they expected it to be hugely painful then that's a different number. If they've got, for example, an abscess with a lot of tension in it, they've got pain before you started, even now they've got a skin incision over that, they've taken away that information, their pain is down. It's very difficult to measure, and huge, so it's about how people perceive it themselves and how they behave.

Anaesthetist 3, time point 1

Data quality issues were also raised as something that reduces the level of trust that clinicians have in the quality monitoring system:

It's been very useful, although as I say I do sometimes dispute whether it's accurate by what time we leave the ward and what time, hence we did our own survey, and showed quite big discrepancies in that.

Surgical nursing lead 5, time point 1

In particular, interviewees reported cases of data not being linked to the right anaesthetist when trainees perform a case under the supervision of a consultant:

If I am at home or on call and they are anaesthetising it, they are doing their standard anaesthetic as a trainee, and I might not even know the trainee! So it is probably not strictly fair.

Anaesthetist 6, time point 1

In terms of the level of data that are fed back to individuals, clinicians emphasised the need for effective specificity and detail. Seventy-five per cent of respondents expressed an interest in a longer report with a more detailed breakdown that would increase the comprehensiveness of the metrics. Requests were made for more detailed information on feedback reports to enable clinicians to funnel down further to case-/patient-specific information and as an aid to recall and learning. This was viewed to be particularly important for the pain and WWT indicators:

Because if you do 99 things well and 1 thing bad, you kind of can't remember the bad thing and you think, 'Oh, maybe it didn't happen', whereas if you had the information on that and you went, 'OK, so that day I didn't do that'. That, you'd learn from it.

Perioperative service lead 1, time point 1

It mainly just gives the timings, doesn't it? I don't think it gives any reasons. Like for example when a recovery nurse has to bring the patient back up to the ward, it doesn't really say what the reason for that was. Maybe some more information or a comments section might be good, like maybe we were short-staffed that day.

Surgical nursing lead 4, time point 1

Further evaluation was directed at the presentation and format in which the feedback reports were delivered to individuals. Mixed views were expressed around the value of normative (peer comparison) feedback versus individual trends over time to support improvement:

I think, ultimately, though a timeline reflecting practice over time and how that's changed or doesn't change, might be more meaningful than a month-on-month review.

Anaesthetist 5, time point 1

But if you can see a trend then you can kind of predict, well, Thursdays are not a very good day, we need extra staff to really give that extra push to get patients up and to do recovery. You can factor that into the establishment.

Surgical nursing lead 1, time point 1

I have never ever seen myself graded against others in the department before so actually that was quite good, and I actually . . . it is nice to see where you are in the department as a whole in terms of those, I mean because those measures are fairly . . . you know, they are there so it is fairly iron tight.

Anaesthetist 6, time point 1

However, 61% of respondents felt that in order for the reports to reach their potential they needed to contain a combination of both of these approaches to enable end-users to benchmark their performance both against their own baseline and within a comparable peer group:

Yes, I think, for me to improve my practice I would need to first have my comparable data over a month or over a year. And also how does my data compare to other anaesthetists that do exactly the same thing? And I think then you'd get a more accurate idea of how you can improve or whether you should improve or whether you need to improve.

Anaesthetist 4, time point 1

Interviewees felt that normative comparisons would become much more useful if case mix was better accounted for. Comparing one's own performance with that of others who do not have a similar case mix was viewed as disengaging and demotivating as it did not provide meaningful information:

The difference is comparing yourself . . . you need to compare like to like. It's pointless comparing my practice with a colleague who does nothing like me, who does different kinds of cases, different kinds of pathologies. So that's a difficult one; you need to compare like with like.

Anaesthetist 5, time point 1

I think it's very difficult to compare to other wards because everywhere is different and has its own set of problems. So I look at my ward, and I don't really care how I'm performing against other areas because everyone's different.

Surgical nursing lead 6, time point 1

The importance of anonymity was also emphasised as a factor that naturally increases people's engagement with the project. Interviewees felt that the removal of anonymity could potentially put the success of the initiative at risk by increasing end-user resistance:

Only consultant resistance and, I suppose, almost embarrassment at having your own figures published. And I think that's where having it anonymised works quite well.

Perioperative service lead 2, time point 1

Clinicians expressed a preference for graphics and figures over numbers and statistics. These were viewed to be more effective at successfully transmitting useful information to the recipient:

First of all I like the fact that it is a graphic, it is not a number. I found this much more effective.

Anaesthetist 1, time point 1

Especially, you know, graphics, they are very, you know, for someone to see it, it's very easy to spot the difference and, you know, what's going on.

Surgical nursing lead 2, time point 1

Application of feedback to departmental quality improvement and professional behaviour change

This section has been aggregated from eight mid-level categories and 17 low-level codes. It comments on the mechanisms through which groups and individuals have interacted with and used the data that the feedback reports provide them with. It captures specific narratives and uses case scenarios while evaluating the impact of the programme on staff capability to use data from quality indicators effectively. Categories are based around attaining the balance between quality improvement and performance management as well as affective and cognitive reactions to individualised feedback.

The feedback initiative was perceived to be useful for quality improvement at the departmental or clinical unit level, particularly in terms of providing evidence for reporting changes in overall performance over time:

That would be useful to know, whether the whole department scored differently last month in one thing. That should be quite useful.

Anaesthetist 5, time point 1

It's good to see what they are and where the Trust is and where you need to improve, and against the national. And it does make competition dare I say it within the NHS but it's not always for the better but it just gives you, like I say, a quantitative to where you are and where you need to be.

Surgical nursing lead 5, time point 1

Having the data to be able to evidence claims about the state of care being delivered locally was perceived as useful for a clinical unit or department embedded within a broader health-care system:

Well, first of all, I can ask for a big bonus because, let's be honest, if I can show that my team have decreased nausea and vomiting, pain, increased temperature over time, that's a result. That's the continuous quality improvement thing. If I can show that during the period of this process we managed to improve in these outcomes, I can then ascribe the improvements in outcomes to the process. May not necessarily be true but I could, and that's fantastic. That would be lovely.

Perioperative service lead 3, time point 1

Yeah. They will send me a message saying, 'Well done [surgical ward],' and the staff, I don't receive messages back but that's just good for them to see what we're doing and where we're at.

Surgical nursing lead 4, time point 1

This also provided a firm basis for connecting with senior levels of the organisation and requesting the necessary support for further improvement:

I would get the heads of nursing for each of the CPGs [Clinical Programme Groups]. And then sit down with them, say to them that you would like to sit down with their lead nurses and just go through some of the data. And say, you know, we send this out to you every month, do you read it, do you take it on board, how do you think we can start making these changes, because the impact on recovery is sometimes very significant.

Perioperative service lead 4, time point 1

There was much discussion around the need for an effective balance between quality improvement and performance management. Some interviewees felt that reports should not be viewed punitively or associated with performance management if they were going to be successful in engaging people in reflection and improvement:

People have tried to do quality improvement processes by being more confrontational and ended up with absolutely nothing out of it, so I think things are improving.

Anaesthetist 10, time point 1

However, there was also a strong view expressed by the department leads that when it came to patient safety there was a responsibility to act on data that indicated low-quality care or that provided evidence of poor compliance with best practice guidelines:

But I think what we ought to do is sit down, CPG [Clinical Programme Group] by CPG, with the heads of nursing and some of the senior nurses and go through the data, the ones in particular who are not doing so well. So you leave [high performing surgical wards] out of it and you pick the rest of them and you sit down and say, OK folks, how can we come to an agreement that we make this better, how can we make pick up from recovery to be seen as as important as some of the other things that you deem are more important. And I think that's the way forward, to be honest with you. Because you send that information out, and I do know that people read it because a couple of people have rung me up and said, oh, I haven't received it, and so I do know that some of them read it, but I don't think it's at the forefront of their brain. And it's a symptom of other problems, which they're probably trying very hard to solve.

Perioperative service lead 4, time point 1

It was thought to be important that anonymity could be bypassed if there was a risk of unsafe care being delivered:

I think you have to have a crackable code if somebody can make the case that patient safety may be at risk if it's uncrackable.

Perioperative service lead 3, time point 1

Clinicians expressed a number of different reactions to receiving individual-level feedback on the quality of care that they provide. Interviewees reported that peer group comparisons had a motivating effect when it came to changing personal practice, and made individuals aware of what was possible in terms of high performance:

And if we see we are down here in the lower ranks of quality in terms of nausea, vomiting and pain relief, that's a tremendous incentive to move ourselves up to there. And if everybody is so motivated to move ourselves up, then the median is going to get pushed up and up and up and up.

Perioperative service lead 3, time point 1

I'm quite competitive so I wanted to make it better. And then it was the question, why can't we do it as others, you know?

Surgical nursing lead 2, time point 1

There was a strong sense that people had a genuine desire to perform well on the feedback reports. This was linked to the fact that negative feedback could result in a feeling of alarm and disappointment for the recipient:

I think if I ended up in the bottom or sort of below at least that line, I'd be like, 'Oh my gosh, what am I doing wrong?' ... Well no one likes to be criticised do they.

Perioperative service lead 1, time point 1

I think for me I was really shocked to see the time the patient stayed in recovery. I think that was a big eye-opener, definitely.

Surgical nursing lead 1, time point 1

The reports were perceived to be effective at automatically promoting thoughts about practice and potential improvement. Clinicians felt that the data encouraged them to pause and consider how they may need to alter their practice for the benefit of the patient:

And you look at it initially and you think, 'No, that can't be right. How can I be down here? Down at the bottom'. And the mature response is, 'Well, actually, perhaps I am. Let's go and really have a look at those patients and let's see if I can improve'.

Perioperative service lead 3, time point 1

They've made me think. I wouldn't say they've made me change my practice because I think I've always been aware probably not of the amount of minutes the patients have to wait, but there's nothing much I can do to make things any smoother than . . . I think we've tried quite hard to get our patients back.

Surgical nursing lead 3, time point 1

Interviewees discussed the impact of the feedback reports on their conceptualisation of their own performance. Receiving the reports gave them the opportunity to quantify and objectify what was previously an abstract representation of their practice:

Well yes, it tells me . . . it puts a percentage on it which I didn't know before, I had a rough idea, I always thought as I think probably most anaesthetists do they think they do quite well and you know I . . . yes I didn't know the exact figures so it did tell me that, yes. It gave me some more exact . . . it put some more exact science behind it and some figures which, you know, I quite liked actually.

Anaesthetist 1, time point 1

Obviously it's useful. Having data is always helpful when trying to improve practice because then you've got a starting point and then if you make improvements you've got figures that show you improvement, which is always the best way to measure anything.

Surgical nursing lead 6, time point 1

Individuals had strong beliefs around what the improvement process consisted of and how people should be acting on the feedback reports that they receive to monitor variation, prioritise action and improve care:

I think they should look at their numbers and check they are achieving good standards of care really and do something about it if they weren't.

Anaesthetist 1, time point 1

. . . suppose if there was an area that was always scoring low then they'd need to look into why that was happening and see if they could address the reasons.

Surgical nursing lead 4, time point 1

If I was a ward manager I would look at the report and see that I'm doing very badly and I'd look into what is it that is causing me to reflect so badly compared to other wards? And if I'm honest with myself I would eliminate the reasons one by one to improve.

Perioperative service lead 5, time point 1

There was variation on judgement as to when an improvement was actually needed based on the data in the reports. This generally came down to the fact that scores across the board were very high and therefore it fell down to personal preference whether or not improvements were seen as necessary:

But unless I had a patient who was extremely cold or extremely . . . then I'm not too worried. If they're just a little bit cold, well that's not a major concern of mine.

Anaesthetist 5, time point 1

No, I don't really make any changes because they all get Bair Huggers [TM], they all get fluid warmers, they all get all the post-op nausea and vomiting bits. The pain thing, I'm usually on top of, so don't, that's not usually a problem.

Anaesthetist 4, time point 1

Ultimately, many interviewees took the opportunity to discuss specific improvements that they had made to their own or observed in others' practice based on the data that had been fed back to them. Fifty-seven per cent of respondents described specific changes that they had made as an effect of the initiative:

I thought: 'My goodness, I do quite a lot of patients'; 'my goodness, oh, some of them are in more pain than I thought they would be in'. And I did some things to change it; so I changed my own practice a little bit, particularly on the gynaecology patients . . . I do an abortion list on a Thursday, and we were using a diclofenac suppository which doesn't really start working in recovery – it's working about half an hour later; whereas I changed it to an intravenous preparation of ketorolac, which is working in recovery and works quite nicely.

Perioperative service lead 2, time point 1

But what this project has done is it has highlighted to the ward managers, the ward staff, that when recovery calls they feel they have to come because they are now aware of this project and the monitoring. And some of the wards who used to be the culprits for coming down very late sometimes come down very early because they know that if they don't they are being monitored. So it helps us in the theatre department to just turn around our patients as quick as we can, although I should say that most probably because of the bigger unit, delay in collection would not impact our theatre utilising like it used to do in the small unit, but I think it has improved over time.

Perioperative service lead 5, time point 1

It's had an impact there and drastically reduced it by them having to be more organised. Because it used to be a pattern of trying to send patients down here who were actually going home later that day just to make a bed up there, so that's going to stop. So that option won't be there. But it's good to see that HDU [high-dependency unit] has vastly improved, so that has which is great, because it's good for the trust, for the targets, so yeah.

Surgical nursing lead 5, time point 1

It was suggested that individual users of the feedback reports may benefit from further support in the translation of their feedback report into effective information that can be applied to make improvements. This could potentially be provided through increased interaction with colleagues and wider dissemination and discussion of outcomes of the project, which may help individuals to prioritise action, set measurable objectives and monitor progress:

But if there is a problem like that, and you can't see how you can improve it, then you have got to work out what the barrier is and I suppose you might then need to talk to a colleague about that, because if you are having pain problems and you are doing everything you could do, it could be your epidural technique, it could be something.

Anaesthetist 1, time point 1

Not really. But we could have such meetings. It would be nice to be able to meet with managers on [other surgical wards] and just share this data with them, and just hear their side of the story, you know.

Perioperative service lead 5, time point 1

Ninety per cent of respondents agreed that the reports had strong potential for use as part of the upcoming revalidation process. Clinicians identified the benefit of automatically receiving data that demonstrated their caseload and performance:

Yes I think it will definitely be, well like I say you take these numbers to my appraisal and then the next stage is going to be revalidation, and I think that will, absolutely, it is going to . . . you can show how many cases you have done, your case mix and your results to a certain extent so yes I think it will be very useful.

Anaesthetist 1, time point 1

It was also recognised that the reports have a role to play in appraisals as a demonstration of high and consistent performance:

And also, you know, for appraisal and revalidation this is just golden because you can say, 'I'm above average . . . I'm above average . . . I'm above average . . . I'm good'.

Perioperative service lead 3, time point 1

The context for feedback initiatives

This section has been aggregated from five mid-level categories and eight low-level codes. It comments on the key barriers and enablers to the successful development, implementation and utilisation of this type of quality monitoring and feedback system within a specific service context. Categories are combined to represent the interaction between external influences on a patient's QoR, the availability of time and resources to support an initiative of this type and the personal characteristics of the clinicians receiving feedback.

In addition to procedural variations and patient-specific factors, interviewees reported a number of factors that were external to the feedback initiative but which impacted on its success by having a contextual influence. These interactions included the influence of other members of the perioperative team:

I think, actually, our capacity to influence overall patient outcome is immense but because we are part of a very large team it's very difficult to single out what difference that individual anaesthetist makes.

Perioperative service lead 3, time point 1

I haven't got any control in relation to medical staff, obviously. If they're not coming and they're not discharging patients on time, you know, obviously this is the most difficult part of this.

Surgical nursing lead 2, time point 1

The availability of time and resources was also thought to impact on the sustainability of the initiative in the future and the ability to make improvements based on it:

Well, just work. It's quite a lot of work for the recovery nurse, it's a lot of work for whoever analyses it all, so, you know, it's not something I'd be happy to do to sit down and trawl through all those bits of paper, it's an awful lot of work.

Perioperative service lead 1, time point 1

Obviously time is always a barrier, but there's certainly no barriers from a point of view that I don't think they're important, it's just time, and having to share my attention to a number of different areas that all want things improved, and I can't just focus on one thing in trying to improve it, I need to be focusing a little bit on every area to improve so that everybody sees improvement across the board. This is but one aspect of my day.

Surgical nursing lead 6, time point 1

However, transparency within the organisation in relation to the project was reported to be high, which allowed for open discussion and a constructive response to existing variation:

I don't think we're particularly adversarial here, and I think we generally, kind of, discuss things and we're quite open with each other about our data and about how we do things.

Perioperative service lead 2, time point 1

Characteristics of the recipient of the feedback report were thought to influence the way in which it is used. The reports served different purposes for different people, and different personalities were prone to engage in different mechanisms of change:

Any feedback mechanism requires that an individual opens the envelope and has a look at the information and processes it. So there's some people who just won't be interested in it and there's not much we can do about that.

Anaesthetist 10, time point 1

Cumulative and serial effects of the intervention

This section has been aggregated from four mid-level categories and nine low-level codes. It comments on how perceptions of the initiative and levels of engagement with data feedback processes have changed over time as a result of both general habituation to the project and ongoing iterations that were made to the reports. Categories are combined to understand how a decrease in emotional impact could be accompanied by an increase in positive attitudes towards the feedback initiative. This section also provides us with evaluative perspectives on the project from anaesthetists that were not interviewed at time point 1 and, therefore, were not included as part of the above process evaluation. This has resulted in the development of more evolved categories around data quality and requirements for the future development of a sustainable initiative.

Longitudinal data demonstrated that the emotional impact of receiving feedback had decreased over time:

Initially very depressing [interviewer laughs] and then actually it's lost value, I'd say as time goes by, it's lost its impact. Yeah, I'm pretty stationary.

Anaesthetist 20, time point 2

Well, I think we did when they first started [laughs] doing them, because we were going, oh, flipping heck, duh, duh, duh, duh, duh. But, you know, we've got used to it.

Anaesthetist 11, time point 2

However, it was interesting to note that clinicians now felt more positive about receiving the feedback (despite it having less of a core impact):

Yeah, I think it's . . . I think people's position has changed, and I think, surprisingly, very rapidly.

Anaesthetist 5, time point 2

But I think everyone has come around to the fact that it's a good idea and it's definitely the future and the way forward.

Anaesthetist 19, time point 2

A number of reasons for this were evident in the longitudinal data set. First, initial scepticism had faded and clinicians were less likely to be wary of the initiative and its underlying purposes. Individuals naturally became more trusting over time and this led to them feeling more positive about their involvement with the project:

Yes I think they have. I think the initial scepticism, when you receive your first report, it's usually, 'Oh well it's just one of these . . . it's a fad, it's going to – it's not going to last and there'll be another initiative soon' so you sort of dismiss it, which is a – I suppose not an unreasonable approach, because we've seen it all.

Anaesthetist 24, time point 2

I was sort of interested, but a bit sceptical about the quality, to be perfectly honest . . . No, I think once I got used to the data, you know, very positive I think.

Anaesthetist 17, time point 2

Second, individuals had more data to work with as the project went on. This gave them greater scope for extracting meaningful information (i.e. identifying trends and patterns) and making evidenced changes to their professional practice:

Well, initially I only had one month so it was just a point in space. So there was no trending as such . . . So anything like this . . . trends are often more useful than a point in time, and you provide that. So right at the beginning I didn't get that obviously, but it's improved over time.

Anaesthetist 13, time point 2

But, you know, having seen it and seen, you know, over the months that month on month you can, you know, you build a picture as to what's going on. No, I've been impressed, actually.

Anaesthetist 17, time point 2

In fact, at time point 2 there were multiple examples of changes that had been made to practice based on feedback. A selection of use cases are presented below:

I've been quite involved with trying to sort it out. The monthly and 3-monthly data shows that [name of hospital site] main theatres and [name of surgical ward] all the patients were . . . I think something like 50% were hypothermic. And we compared to the other sites and that didn't really fit with what we thought was happening. I guess, quite often, we measure the temperature in theatre, and that's normal, and then it's recorded as 35.5 or something by recovery staff. So we suspected it was a problem with the tympanic temperature probes, but we had to prove it, so we did a little audit measuring patient temperatures in recovery, with three different types of thermometer, and it turns out it's basically a measurement error and the tympanic thermometers were under-reading. But one very useful outcome of this data was that's what prompted the whole thing, because someone was saying that one of my patients is hypothermic – because I'm sure they're not. So it led to a quality improvement drive, if you like, with us, so we've got a new type of thermometer in. But there is a problem still, because Clinical Engineering are making a bit of a pain of themselves and saying that we can't just use these thermometers we've bought, they have to go through various processes with the Trust. So it's a bit frustrating.

Anaesthetist 19, time point 2

Yes, yeah. I mean – so this fits well with the way we design the anaesthetic technique for this service which was step-up, step-down and we do something for 2 months and review the results, whether it's recovery ward and then post take ward rounds. So we've always done that in the bariatric unit, so this fits exactly right with it. We just change one thing and do it for 2 months and review the results. It's not so easy to do, but it is helpful. It would be nice to know the peer groupage, because then we could as a group do things. But yes, but that's how we design the technique, so in principle yes, we have changed a few things.

Anaesthetist 20, time point 2

Yes, I think some of the shorter cases, where I wouldn't previously think it was necessary to actively warm the patients, and I think I realised that even the shorter patients could . . . quicker than I anticipated, I think. So yeah, more warming, along with the other anaesthetists who do gynae, probably increase the amount of . . . increase the pain medication used.

Anaesthetist 17, time point 2

Yeah, I have. I am using – I was already doing it anyway, starting to, and it has made me think I should do it more. I'm being a lot more heavy-handed with opioids, particularly towards the end of the case. I'm thumbing in a lot more.

Anaesthetist 15, time point 2

I have yeah. I now give more opiates – well, not just opiates but more pain relief inter-op.

Anaesthetist 13, time point 2

From the feedback reports, I saw that my bariatric patients were in a bit more pain than anyone else so it just made me think about giving more analgesia than I'd already been giving them and some of my orthopaedic patients.

Anaesthetist 16, time point 2

Yes. I've been thinking about months where my nausea and vomiting have dipped below, and it's triggered me into just thinking about it a bit more carefully. So I possibly do treat nausea and vomiting with more agents or different agents compared to before this data started. So, yeah, it has an influence on that.

Anaesthetist 19, time point 2

A further longitudinal effect was that individuals identified multiple ways in which the data could benefit them as an individual, and this opened their eyes to potential opportunities and incentives for interacting with their data. In particular, clinicians identified ways in which their performance data could be used to support them through the upcoming processes of appraisal and revalidation:

Yeah. It is the hardest thing . . . I'm not sure what's happening in the rest of the country in terms of how people are engaging, but I think the driver is going to be the appraisal process and the revalidation. And I think people might then find it a very useful thing. When they use this at their appraisals, if they found it came in use . . . That it's providing the opportunity to . . . We're meant to be encouraging reflective practice, all that kind of stuff, and it does do that.

Anaesthetist 5, time point 2

Very positive, yeah. Yeah. You know, I think once it started and we realised what we were getting, I think it's really sort of feedback and it's certainly very, sort of, for our sort of, you know, appraisal and sort of revalidation, that sort of thing.

Anaesthetist 17, time point 2

This was thought to be something that would continue and become even more prominent in the future:

But, unfortunately, until people can use it for their own advancement, you will still get some who say, 'Well, what does it mean?' But I think that's going to change very rapidly with the whole revalidation thing. I think people will suddenly be very, very grateful that they have something.

Anaesthetist 5, time point 2

The two iterations to the initiative that were perceived by end-users as having had the most significant effect on usability were the introduction of case-specific information on reports to highlight outliers and the introduction of within-specialty group work.

Case-specific information was introduced following requests from end-users for greater detail on reports during the time point 1 interviews. Reports were developed to include patient information associated with any outliers to support clinicians in recalling case details and identifying why outcomes may have varied. Interviewees found this extremely useful because it enabled them to focus in on the most important areas for follow up and ensure that they had sufficient information to respond appropriately:

I think they've become more useful and maybe more meaningful in terms of, I think, the way things have changed. For example, you'll be told, for instance, three of your patients arrived who were cold, and this is the kind of cases they are, so, actually, I think it's become more personalised, and, actually, I quite like that, it's made it more relevant to me.

Anaesthetist 5, time point 2

I do read it carefully, page to page, and the thing I zoom in on are the particular patients that – you know the ones where they say like, 'You've had four patients with nausea and vomiting' I'll look specifically at those ones and then if there's something strange about one of them I don't particular remember, then I'll look back on my phone and see . . .

Perioperative service lead 1, time point 2

Within-specialty group work was introduced as a response to concerns around inappropriate case-mix comparisons. The group work was a pilot based on a team of anaesthetists specialising in gynaecological surgery. It involved presenting comparisons within the specialty peer group and allowing the highest performer to act as a mentor for the rest of the group. Interviewees extracted much greater meaning from this type of comparison, as they perceived it as having greater validity:

Have they changed? No, the only thing I have seen [project lead] do is the . . . I think he looked specifically at the gynae data compared to everybody else, and that was beautiful because, you know, I could then see what everyone else was doing, but as a big block, 'This is your results from recovery,' didn't help at all because my patients don't go to recovery here, they go to recovery in day care.

Anaesthetist 9, time point 2

You can get that simply from the data because now that [project lead] has started giving us your score for X measurement within the speciality groups then you are comparing not necessarily like with like but something much closer to it. So that I think is very helpful. That, to me, is the most useful part of it, and it also enables me to say that, OK, you are doing fine at the gynae but the general surgery wasn't so great.

Anaesthetist 15, time point 2

Revealing the identity of the highest performer also enabled them to request support and identify potential changes to be made to their own practice through improvement conversations:

Yeah. Well, for instance, it was pointed out there were two particular people who performed extremely well, and their practice was then reflected on, and a group of people were shown . . . Said could you see what they're doing and see if you could change your practice. So that was almost instigated to someone actually changing or being able to use . . . And I think, in a way, one might call it spoon-feeding, but it is quite useful to have that, in a way. Someone needs to . . . I think, for me, what is more meaningful is a personal touch behind it. Somebody is actually saying this is what you need to know, this is where you didn't perform as well. So it does introduce that element of, well, your patients are not doing as well as her patients, instead of just relying on you to go, yourself, hang on a second . . . But that came down to identifying a clear trend, which I don't think there are that many clear trends – but that was a particular one, which seemed to work.

Anaesthetist 5, time point 2

At the time of the second round of data collection, anaesthetists were receiving regular feedback based on the QoR score. There were mixed views about the inclusion of this metric in the feedback reports. Generally, participants felt that it was important to include measures of patient experience but they were unsure about the current metric's level of meaning and accuracy:

Well, what I like is actually the patient satisfaction because that's all I care about, at the end of the day, is whether or not the patients were happy.

Anaesthetist 21, time point 2

I suspect it's skewed by the fact that there is at least one component that's just always good for everybody so, again, it causes a clustering effect, and I suspect that it might be better if you took at least nausea and vomiting out of that.

Anaesthetist 15, time point 2

Well, I think some of them could be quite subjective. I can't remember the questions off the top of my head but ones like, 'first time to go to the toilet', some people give more fluids in the theatre, some people give them less, so, you know, that's not necessarily the right or wrong thing to do in theatre but it would affect how quickly they went to the toilet afterwards.

Anaesthetist 16, time point 2

I think it's very much influenced with who is asking questions, how is questions asked . . . It depends a lot on patient psychology and time in which the question is asked.

Anaesthetist 23, time point 2

At time point 2, anaesthetists continued to report concerns regarding the quality of the data that they were receiving via the feedback reports. This was mainly associated with issues of contamination in terms of having all of your own cases included and none of anybody else's:

I think so. It's hard to be sure, the ones that are missing, because they're not itemised, but I'm assuming most of them are on there. I am aware that sometimes things aren't accurate on TheatreMan, because occasionally I see cases which I know I haven't done or shouldn't even be supervising, so might be entered incorrectly. I'm aware sometimes theatre staff put my name in incorrectly, so I try and double-check it.

Anaesthetist 19, time point 2

Well, basically, if you're the consultant on call and it gets labelled and it comes under your name even though you've not even been told of the case that happened, and then you find when this comes out that you don't know anything about the case anyway.

Anaesthetist 21, time point 2

Where the nursing staff have clearly put in the wrong name. And I think that is an issue, but, you know, bar actually going up to the computer every time and saying, by the way, I am this person, and which is actually quite difficult to do, you know. If you've got a busy list then it's very difficult to do that. So no, because I've put down to-do lists that I know I would never do in a million years [laughs].

Anaesthetist 11, time point 2

Well, I mean, I just remember from one of the last ones I had three patients in severe pain. There were four patients in severe pain and three of the patients in severe pain weren't mine. And I knew they weren't mine because one was an open myomectomy and I've never done an open myomectomy at this hospital. And two were two 15-year-old appendixes which had had my name on because I was the consultant on-call, but I didn't do them.

Anaesthetist 12, time point 2

Anaesthetists emphasised the importance of collecting data at the right time during the perioperative pathway. Some concerns were voiced about whether or not recovery was the right time and place to be collecting data from patients:

Ah, they're still waking up and they're, you know, they're . . . you know, if we were to . . . You know, if you fine motor somebody, fine motor test you now, give you an anaesthetic, fine motor test you tomorrow, your performance tomorrow would be rubbish. You know, it would be the equivalent of being over the limit for driving, for instance. So, you know, they're not, you know, they're sort of coming round, they're confused, you know, particularly as they're older. It's not ideal, but when else can you capture them, other than, you know, trawling around the wards and that's just so labour intensive.

Anaesthetist 17, time point 2

What I would also like which you're not – this is – it would be a completely separate data set is either patient kind of postal survey afterwards. The trouble is, they don't really remember us. They go, 'It was the tall blonde one' or, 'It was the short brown girl' or – do you know what I mean? And they don't really remember us. But it would be nice to have – I've looked at this on the Trust GMC survey it's got suggested things for you know when we go and see the patient and it says, 'Was the person polite and presentable? Were they not smelling of booze and curry? Did they explain the procedure to you? Did you feel comfortable? You were able to ask questions that you came up with the treatment plan which you were happy with? Would you be happy to see this person again? Did you feel confident in them?' All of these quite fluffy, soft things, but soft, fluffy things are normally the most important things, especially in doctor–patient relationships. That would be great. That would be really good, but of course that is a completely different data set and capture to what you're doing here. But that – and then reported back to you as an individual. That would be fantastic.

Anaesthetist 22, time point 2

There was a strong sense that feedback needed to go further and develop more so that it could be used by end-users in greater detail. There was certainly a feeling that the feedback in its current form had reached its potential, and anaesthetists requested further iterations in order for them to be able to continue improving their professional practice. Many interviewees were keen to be measuring quality indicators preoperatively as well as postoperatively:

You know, we spend a very, you know, a surgeon sees a patient often, more than once in clinic beforehand and builds up a relationship there, we get a few minutes before surgery. And it's a pretty major thing that they're having done, that we're being involved in, so to build up sort of a rapport and trust in that short period of time, you would hope that you might get some sort of feedback from that, not just whether they wake up in pain or they're feeling a bit sick afterwards.

Anaesthetist 17, time point 2

Maybe it would be better to ask questions in a way that you know what the patient thought about the quality of preoperative visit and postoperative care, so you have both, for anaesthetist it would be more important to know what they think about preoperative visit.

Anaesthetist 23, time point 2

In terms of presentation of data it was also suggested that anaesthetists would benefit from receiving 6-monthly or yearly summary reports to maximise the usability of the data:

I do think a slightly longer-term summary of data might be useful – that's why I said yearly, but maybe even 6-monthly or quarterly – where are the hotspots of poor recovery or poor analgesic management or poor nausea and vomiting management and poor temperature. I think that would be useful.

Anaesthetist 19, time point 2

Ideas for improved methods of benchmarking were also put forward as a way to reduce the influence of contextual factors on the ability to monitor personal performance. These included a focus on ranking over time to provide end-users with a different type of historic benchmark and the use of vertical comparison to compare anaesthetists working in the same theatre on the same procedures:

The overall one is interesting more because, I think . . . Well, actually, it's funny, because I've never really thought about it before, but if you look at it in terms of where you are . . . So if your caseload is stable and you're always doing the same sort of stuff and you don't do much changes, you might expect that you'd always be in position 18 out of 40 – but if you're in position 18 out of 40 for month after month after month and suddenly you slip to position 36 out of 40, something has changed with your practice. So that might be an interesting thing to look at to see how people have changed over time – if anybody has changed their ratings.

Anaesthetist 5, time point 2

Yeah, I think so, 'cause then you say that, you know, maybe there's three of us who regularly do tonsillectomies and, you know, we're all about the same. Or, you know, if one stood out as being significantly better, you know, I would be, you know, I'd go and talk to them and say what do you do that makes you look so good. Or, on the other hand, I would quite happily if somebody said actually how do you do your tonsils because you seem to be, you know, and just sort of sharing the . . . and then that improves all of us then, and that's, you know, that's what we should be doing.

Anaesthetist 17, time point 2

In terms of local barriers to effective use of feedback, anaesthetists at this time point did not generally perceive the organisation to be a transparent learning environment to work in. Issues with senior management were expressed and these were thought to impact on the organisational culture surrounding the feedback initiative:

I don't know how it's in other hospitals, but I don't think the culture of transparency in learning is a thing in this trust.

Anaesthetist 24, time point 2

Yeah. But that does happen – you try and make a change, you get two or three interested people that drive it through, and then you hit brick walls and red tape. So there's quite a lot of that in this organisation generally. So the worker bees, which are the consultants and nurses and juniors, do meet these obstacles sometimes. So I'd say we might not be supported so much in that – because you would hope somebody enthusiastic would say, 'That's great. You've got off your backsides to find a problem. You've done something about it. We'll do everything we can to rectify it for you' – but it doesn't feel like that.

Anaesthetist 19, time point 2

Discussion

This work stream explored clinician perspectives and experiences of the complex quality monitoring and feedback initiative in perioperative services. The results provide a rich understanding of the causal mechanisms of effectiveness for monitoring performance and making improvements to practice based on quality indicators. They also comment on how such processes develop over time as part of an iterative quality improvement initiative.

Adoption of technology has been linked to how acceptable participants perceive a system to be.⁶² Our results demonstrate that this initiative was clearly wanted by the clinicians and was viewed as a definite first step in the right direction towards making lasting improvements to patient care based on systematic performance data. In this sense the end-users recognised the existence of a problem and the need for a solution (which in itself can be a challenge for new interventions).⁹⁶ This is supported by the fact that very few evaluative studies of personal professional monitoring programmes for perioperative practitioners

currently exist in health-care organisations.^{12,13} Owing to modern productive pressures for longer theatre lists and higher patient throughput, there is considerable pressure on anaesthetists to focus their attention solely on the intraoperative process and not follow up on patients in the later stages of recovery. Recovery-based indicator feedback, therefore, provides a means of learning from anaesthetic outcomes in the immediate postoperative period that was previously delivered through irregular patient contact. In this sense, the whole initiative provides clinicians with ammunition to change the nature of the discourse with their senior managers. By measuring something associated with quality, it automatically makes it possible to talk about it in the same terms as, for instance, efficiency. Therefore, the whole exercise is a useful organisational corrective to put quality back 'on the agenda' from a clinical point of view and increase the visibility and credibility of anaesthesia within the hospital.

Translating data into information

It was emphasised that the right quality indicators need to be selected with the right characteristics (i.e. they must be specific, relevant and meaningful) in order to promote the necessary level of trust in the data and convince clinicians that the initiative is appropriate for their use. The importance of these characteristics is supported by the literature on characteristics of effective feedback.^{52,97,98} In particular, the relevance of using pain and nausea and vomiting as indicators of quality of care was emphasised in this qualitative evaluation. These indicators have been shown to be two of the most important dimensions of quality of recovery and have been empirically linked to prolonged postoperative stay after ambulatory surgery⁹⁹ and to overall patient satisfaction.¹⁰⁰

The soft nature of quality outcomes for the anaesthetic process makes it difficult to capture effectively and measure quality of care. No matter how well measures are formulated, some phenomena due to its experiential and subjective nature will mean that the numbers are still open to interpretation (limiting ability to guide action). Presenting the data back to clinicians in the right format to support them in transforming them into useful information that has meaning and gives out an actionable message is, therefore, equally important. Our analysis suggests that a combination of normative comparison (i.e. genuine peer benchmarking) and individual trends over time may have the greatest effect. The need to transform hard data into usable information and the experience of health-care professionals in doing so can be viewed as a powerful message emerging from this study.

Translating information into action

Even once the data have successfully been interpreted to result in meaningful segments of information, there is still a need to transform that information into practical action that can potentially make a difference to patient care. This represents the move from the conceptual to the practical stage of effective data use. The importance of this transition has been explored in the literature.⁴⁵ Lessons from incident reporting demonstrate the significance of dissemination and action plans on the frontline of health care in order to make data collection worthwhile.^{101–103} Principles established in the fields of engineering and cybernetics are that information by itself is not feedback unless it involves some action or response to close the identified gap.¹⁰⁴ The interviewees explored the translation of information into action at two levels of the health-care system, namely the departmental level and the individual-clinician level; crucially, the mechanisms of effective data use were different at each level.

The area of performance measurement systems is particularly vulnerable to inevitable issues of sensitivity around utility, fairness and unintended consequences.¹⁰⁵ Feedback initiatives should be monitored closely to ensure that they are not misused for 'gaming' purposes and do not result in a narrow focus on the metric rather than the quality of care that it represents.¹⁰⁶ Feedback that is confidential and presented with a non-judgemental tone has been found to be most effective.⁵² There was a strong understanding that, ultimately, patient safety had to be prioritised over the protection of individual employees. However, in order for this to be a fair process any feedback initiative must incorporate a thorough acknowledgement of case-mix variations. Careful selection of the right indicators can reduce the need for case-mix adjustment. Arguably, all patients should arrive in recovery warm and have well-managed pain and nausea postoperatively. If there are differences in risk associated with these measures, then anaesthetists should

be compensating for that through their practice. Realistically, though, pain/nausea is more difficult to control for in certain procedures. This goes some way in explaining why within-specialty group work was so welcomed by end-users at the latter stages of the project. This enabled fair comparison alongside an opportunity to learn from peer experts in an open and constructive way.

At the individual level, conceptualisations of anonymity and its importance in supporting the success of the initiative were contradictory in nature. In terms of implementing a local quality monitoring initiative, there seemed to be a desire for anonymous data in order to get people on board and decrease resistance. However, as a project is initiated and individuals identify its benefits, views on anonymity tend to mature. At the individual level, with a desire to receive normative feedback, people wanted to identify and contact high performers in order to obtain support and ideas for behaviour change. In this sense anonymity could actually be viewed as a barrier to the effective translation of information into improvements. Learning theory states that feedback should also be task rather than individual focused.⁴⁸ By encouraging clinicians to focus on effectively improving quality of care rather than punitively assessing each other's performance, we can ensure that this type of initiative reaches its potential.

There is a need for more support and guidelines in identifying exactly when information needs to be acted on. This currently was assigned to personal and individual decision-making rather than systematic analysis of the data and the meaning that it provides. This was particularly relevant in the light of consistently high scores across the majority of clinicians involved in the initiative. Research has shown that low baseline compliance with desired practice increases the effectiveness of feedback.¹⁰⁷ Thought should be given to how the research team can adapt the mechanisms of our initiative to support clinicians in their experience of the revalidation process and enable them to construct a balance between meeting baseline standards and continuing to reflect on performance and reach one's professional potential. Interviewees clearly identified a role for this initiative in revalidation, and participating in quality monitoring and acting on the results is an identified dimension of good medical practice.¹⁰ In fact, the conceptual connection of the initiative with revalidation and appraisal appeared to significantly increase levels of engagement throughout the evolution of the project.

Receptive contexts for change

There is a clear need to be aware of contextual influences when exploring topics of a sociotechnical nature.⁶¹ The term 'contextual influences' refers to factors that are not associated with or controlled by the initiative itself, but do have an impact on the way in which it works. Studies have demonstrated that organisational context influences perceptions of usability associated with performance data.^{97,108} There is also a further need to ensure protected time and resources and increase transparency across the organisation if initiatives such as this are going to flourish and be ultimately sustainable. In the design of feedback interventions there should be active consideration for how the impact of context will be captured and understood in relation to the outcomes.¹⁰⁸

Mechanisms of data feedback effectiveness

Introduction to relevant theories

Interpretation of findings from the initial inductive analysis suggested that pre-existing theories in the areas of social sciences and health services research may be of relevance to better understanding the fundamental mechanisms of data feedback effectiveness through the eyes of end-users. In identifying which of these theories are most applicable in explaining such processes we may be able to improve the design of future initiatives and maximise the desired outcomes (i.e. greater clinician engagement with performance feedback leading to improvements in patient care). It will also help us to better comprehend the relationship between these mechanisms and the local context in which they are taking place, which is essential for a realist evaluation of this type of complex quality improvement initiative.

The primary goal of data feedback is to encourage recipients to make appropriate changes to their behaviour based on review and interpretation of current performance levels. Theories of behaviour change, such as the theory of planned behaviour,¹⁰⁹ comment on the psychological constructs that predict whether or not we are likely to change our behaviour. One would, therefore, expect there to be a relationship between those established psychological constructs and the way in which individuals engage with and respond to their feedback reports. By identifying, understanding and targeting the appropriate psychological constructs it may be possible to increase the likelihood of behaviour change in future initiatives and therefore have a positive impact on patient care. The initial emotional response to receiving feedback on performance is also likely to have an impact on how people will go on to use it. Psychological theories such as cognitive dissonance theory¹¹⁰ and instrumental learning theory (i.e. operant conditioning)¹¹¹ provide explanations for how clinicians may feel when they receive performance feedback and how these feelings may impact on both thoughts and behaviour.

The introduction of regular feedback reports at the individual level can be regarded as a new technology that people need to adapt and respond to, and in this sense it would be useful to understand and be aware of common patterns in how this takes place. The diffusion of innovation theory^{112,113} and the technology acceptance model¹¹⁴⁻¹¹⁶ may go some way in explaining how and why responses to feedback change over time and the impact that this has on the mechanisms through which they work. It is important to view the initiative from the perspective of groups as well as from that of individuals. Understanding the way in which people are likely to respond to feedback as a collective will also support the design and acceptance of future initiatives. For example, the organisational context will influence the way in which change is received and responded to. The way in which individuals interact with and support each other throughout the learning process is also likely to have an influence on overall outcomes.

The purpose of this additional component of the analysis was, therefore, to identify to what extent these theories were evidenced by our existing qualitative data and thus may be of greatest relevance to understanding the mechanisms of data feedback and developing future initiatives of this kind.

Methods

Following the initial inductive analysis (reported in section 1) the data from all consultant anaesthetist interviews at both time points (including perioperative service leads who were practising as consultant anaesthetists simultaneously) were explored deductively to inform interpretation of meaning in relation to a set of preidentified relevant theories from the literature review. *Table 7* lists the theories that were considered in the analysis and the processes through which they had the potential to explain the mechanisms of data feedback effectiveness. The deductive analysis involved a research associate in quality and safety coding the raw data against key constructs of the individual theories using NVivo and exploring the existing themes (from the prior inductive analysis) to look for relationships with the relevant literature. In this sense the analysis represented a balance between an inductive and a deductive approach. Initial outcomes were reviewed and discussed with clinical input from a consultant anaesthetist and a consultant intensivist, as well as ongoing academic input from the senior social sciences researcher.

Results: empirically informed critical appraisal of theory

Table 8 displays the key mechanisms of data use that emerged from our analysis as being of greatest relevance to the experiences of our anaesthetists alongside the relevant codes to which the data were categorised.

The data from both time points were explored to identify examples of the high-level constructs. These were then explored more inductively to extract greater meaning and insight in terms of the core mechanisms that were at play. Researcher commentary and example quotations are presented against each of the mechanisms below.

TABLE 7 Relevant theories extracted from the literature to be considered through qualitative analysis

Theory	Key references	Research questions
Theory of planned behaviour	Ajzen ¹⁰⁹	<p>What do people perceive as the consequences of using feedback? Are those consequences positive or negative?</p> <p>What are the potential negative or positive consequences of receiving and using data based on your performance?</p> <p>What are the subjective norms? What do other people think about them using feedback to make improvements? What does the department/organisation think? What do other theatre department professionals think?</p> <p>Do anaesthetists feel that they have the resources to use feedback for improvement? What are the barriers and facilitators to achieving this?</p>
Normalisation process theory	May and Finch (2009), ¹¹⁷ May <i>et al.</i> (2007), ¹¹⁸ May <i>et al.</i> (2009), ¹¹⁹ Murray <i>et al.</i> (2010) ¹²⁰	<p>Does the feedback initiative and the way in which it has been implemented make sense to anaesthetists?</p> <p>How have they have been able to cognitively participate in the implementation of the initiative at this organisation?</p> <p>Do they feel that the outcomes of the implementation process have been monitored and reviewed?</p> <p>Any comments on the fact that we are completing these interviews/survey, etc.?</p>
Adoption of innovation model and diffusion of innovation theory	Greenhalgh <i>et al.</i> (2004), ¹¹² Rogers (2004) ¹¹³	<p>What comments have been made on the communication channels used to diffuse and discuss the initiative?</p> <p>Comments on the social system of the anaesthetics department</p> <p>Have any opinion leaders been identified? How has this impacted on diffusion? Any comments on the fact that the project is being led by their peer?</p> <p>Any variation in adoption? Characteristics of individuals and how this links with how much they are using the feedback reports for improvement</p> <p>Any comments on changes over time – how have people’s opinions on the initiative changed over time?</p>
Technology acceptance model (TAM)	Chuttur (2009), ¹¹⁵ Davis (1989), ¹¹⁴ Holden and Karsh (2010) ¹¹⁶	<p>Do people think that using the feedback initiative will enhance their job performance? Is the feedback perceived as useful? In what way is it useful to them?</p> <p>How easy do people think the feedback initiative is to use? Will it take a lot of effort? If so, what is it about it that will require effort?</p>

TABLE 7 Relevant theories extracted from the literature to be considered through qualitative analysis (*continued*)

Theory	Key references	Research questions
Organisational change theory and resistance to change	Gollop <i>et al.</i> (2004), ¹²¹ Gustafson <i>et al.</i> (2003), ¹²² Iles and Sutherland (2001), ¹²³ Pettigrew <i>et al.</i> (2001), ¹²⁴ Scott <i>et al.</i> (2003) ¹²⁵	Do they feel that the feedback initiative addresses customer needs? Do they feel that the advantages of the initiative have been identified and promoted? Do they perceive the design of the feedback to be flexible? Do they perceive involvement, support and goal setting from their leaders? Do they identify evidence of effectiveness during implementation? Were they told how the feedback initiative would work for them in an effective way? Any mention of the implementation being led by key people? Did they feel that goals and priorities of the initiative were clear? Did they perceive there to be a supportive organisational culture surrounding the initiative? Were there strengths and opportunities that were associated with the feedback? Were there weaknesses and threats that were associated with the initiative? Did they identify any driving or restraining forces? Any demonstration of commitment to the project and compliance with its goals? Any comments on ownership?
Sociotechnical systems theory (as applied to information systems development and implementation, including unintended consequences)	Harrison <i>et al.</i> (2007), ¹²⁶ Eason (2007) ⁶¹	Any commentary on interactions between social and technical factors Impact of feedback initiative on local environment Any examples of renegotiation or reinterpretation of the feedback initiative over time Comments on adaptability of the feedback initiative Comments around the meaningfulness of using the feedback initiative – what does it mean for them to be using it?
Theories of professional behaviour change	Ashford <i>et al.</i> (1999), ¹²⁷ Grimshaw <i>et al.</i> (2002), ⁶³ Grimshaw <i>et al.</i> (2001), ⁶⁰ Oxman <i>et al.</i> (1995) ⁵⁹	Any examples of education/training that surrounded the feedback initiative Identification of local opinion leaders that are associated with the project Awareness of cost and resources associated with the project Reference to the literature/research that supports the mechanisms of the initiative Identification of barriers or facilitators to professional behaviour change based on the feedback initiative

continued

TABLE 7 Relevant theories extracted from the literature to be considered through qualitative analysis (*continued*)

Theory	Key references	Research questions
Self-affirmation theory	Steele (1988) ¹²⁸	<p>Any examples of how feedback self-affirms an anaesthetist and emphasises their integrity/worth as an individual</p> <p>Any examples of the feedback being perceived as threatening – defensiveness and message rejection</p> <p>Links between integrity and message acceptance</p>
Cognitive dissonance theory	Festinger (1957) ¹¹⁰	<p>Any examples of how the feedback has conflicted with prior values, ideas or beliefs (i.e. the existence of dissonance)</p> <p>Has this resulted in them altering their cognitions in order to reduce such dissonance?</p> <p>Link this to potential rejection of the feedback reports</p>
Social constructivism/social cognitive theory	Vygotsky (1962), ¹²⁹ Bandura (1986) ¹³⁰	<p>Any examples of social interaction around feedback to establish peer norms</p> <p>Any examples of learning and making changes to practice based on guidance from and modelling of expert peers</p>
Goal setting theory/control theory	Locke (1968) ¹³¹	Any examples of individuals using a goal to guide and assess behaviour
Instrumental learning/operant conditioning/behaviorism	Skinner (1948) ¹¹¹	Any examples of links between feeling rewarded or punished and whether or not behaviour was affected

TABLE 8 Key mechanisms of data use

Mechanism of data use	Example high-level codes	Example low-level codes
Feedback as a component of planned behaviour	Perceived consequences of the initiative	Data could be misinterpreted and used punitively for performance management
	Subjective norms	Changes made based on the data could have negative effects on the patient
Feedback as a socially situated cognition	Perceived behavioural control and associated behaviours	Provided with ongoing access to systematically collected data
		Reassures you of your professional skills
	Examples of peer experts	Demonstrations of self-efficacy
		Resources as a barrier to use
		Level of authority as a barrier to use
		Workload as a barrier to use
		Project leader becomes peer expert
		Highest performer becomes peer expert
		Examples of scaffolding
		Learning from peers
Feedback as a threat to internal consistency	Response to dissonance	Requesting support from peers
		Sense of openness for learning across peer group
	Examples of dissonance	Feedback provides a reason for discussion
		Learning through discussion
		Need to explore alternative communication channels
		Need for a formal process for learning through interaction
		Expectation of high performance
		Positive views of performance not normally challenged
		Having negative feedback contradicts identity as a 'good' doctor
		Feedback has an emotional effect
Emotional response to negative feedback		
Feedback as reinforcement for learning	Feedback reports viewed as a reward	Rejection of a threatening message
		Link between a threatening message and behaviour change
		Sense of professionalism attached to active reflection on performance
Feedback as reinforcement for learning	Feedback reports viewed as a reward	Reward comes through reassurance
		Reward comes through data for revalidation/ CV purposes
		Reward for the department to use as evidence for performance

continued

TABLE 8 Key mechanisms of data use (*continued*)

Mechanism of data use	Example high-level codes	Example low-level codes
Feedback as a motivational tool	Need for a goal to measure performance against Feedback has led to new goals being identified and set	Process of making changes and testing them
Feedback as an innovation	Need for gradual implementation	Stepwise process Gain buy-in Non-threatening

CV, curriculum vitae.

Feedback as a component of planned behaviour

The way in which anaesthetists perceived the consequences of using their feedback reports influenced the way in which they engaged with the initiative as a whole. Examples of both positive and negative perceived consequences and their influence on future behaviour were evident in the data. Interviewees who considered themselves to be highly engaged with the initiative and more likely to change their behaviour based on it tended to focus their attention on positive perceived consequences such as building on their professional ability and improving the experiences of their patients:

Yes I do, I think there are a few things, I think people will look at this and they will look at their own figures and if they . . . it might suddenly bring home to them ‘oh a lot of my patients are feeling a bit nauseous afterwards, what am I doing?’ It will make them think about it. And I think the other thing is, with these figures you can see your comparisons to all your colleagues and I think that definitely will make people think, if they see themselves down a little bit they are going to . . . it will change their practice, I am sure it will because it is a naturally competitive instinct, possibly. And I think it just stops people being complacent about it, because I think that is the biggest thing, especially after you have been in practice for several years, I think people do get complacent and they will just deliver a patient. I am just talking as an anaesthetist, they will deliver a patient off to recovery and I don’t know but I suspect a few people never see them again, they don’t go and see them and don’t try and find out how it is. So this helps a lot I think. And I think it would change people’s practice.

Anaesthetist 1, time point 1

Pain wise, I think it’s really useful because I know that’s really important, we need to get that right and it’s good to know that certain cases you’re doing well and the ones where you aren’t doing well, that it flags those up so you know that the next time you do that what you can try and improve as well.

Anaesthetist 16, time point 2

On the other hand, interviewees who were concerned about potential negative consequences tended to disengage from the initiative. Perceived negative consequences included making changes based on inaccurate data, data being used punitively for performance management and staff becoming preoccupied with the data themselves rather than providing all-round good patient care:

Well, if the data turns out to be either inaccurate or just a statistical elaboration or to be due to a phenomenon that is outside the control of the anaesthetist, a case mix being an example, subsequent location of care might be another one if you look at longer-term stuff, then, yes, of course, you might do a change with unintended adverse consequences.

Anaesthetist 15, time point 2

And even, even stuff like, you know, my concerns about whether the post-op nausea and vomiting numbers are accurate is an important component of that, you know, for me to use feedback I have to think that the feedback is valid otherwise there's no point to me using it.

Anaesthetist 10, time point 1

All data is dangerous isn't it. Data that's not risk-adjusted is dangerous to doctors in this day and age. It would easily be used as a tool to criticise and to remove people from positions and to threaten them. And we see that all the time with the WHO [World Health Organization] Checklist and this sort of data and feedback and it's very easy to do by trusts and misaligned management. I have data concerns, basically.

Anaesthetist 20, time point 2

[S]o long as it doesn't get used in a punitive way which I think would be a concern, not least because I don't think any of us trust the GMC at all not to behave in that way. I certainly don't. I keep being told I'm paranoid but I would be cautious about it simply because I don't trust them.

Anaesthetist 15, time point 2

I worry sometimes that people are interested in data collection rather than patient care. So I think it's still very distracting in recovery because it's not automated. So I feel that care levels do change. People are much more worried about my nausea score than actually dealing with a problem.

Anaesthetist 20, time point 2

That would perhaps concern me that people would say, 'Mustn't let any of my patients have any pain, therefore I'll give them all 10 mg of i.v. [intravenous] morphine just after coming off the table'. And we go into recovery and we might be recording a respiratory break but we're not recording tidal volume, we're not recording end tidal CO₂. You might actually end up having patients stay longer in recovery. I'm not really sure that's terribly useful.

Anaesthetist 3, time point 1

There was variation in the data in terms of the degree to which end-users felt in control of using their feedback to make changes to practice. Some clinicians demonstrated high self-efficacy (i.e. the belief that they have the necessary skills to achieve a goal) when it came to engaging with the project and improving personal performance:

No, I think it's pretty, you know, it's pretty personal data, and I think it's . . . you know, I can either just read it and forget, or I can read it and act on it. I don't think there's any barrier to me acting on it.

Anaesthetist 17, time point 2

Yes, I don't think anything restricts me to look at and analyse it for myself and then use it in my practice.

Anaesthetist 16, time point 2

Not that I'm aware of, no. I mean all the factors involved I can change independently of what other people want me to do. I'm not forced to do anything in particular, if I want to add in something I can.

Anaesthetist 13, time point 2

I know that I'm able to immediately affect the outcome of these measures, so I can do things to make these measures different.

Anaesthetist 10, time point 1

On the other hand, some participants also identified clear barriers to being able to engage with and use the data in an optimal way. Barriers included restrictions on available resources, the influence of other people on their outcomes and limits on the number of changes that can feasibly be made to anaesthetic practice:

Well . . . I don't have any resources at all. I don't have a budget, I'm not in control of my staffing levels, I don't have a drug – or, no I just shout at people and ask them questions and see if they'll do stuff for me! I don't have any resources.

Anaesthetist 22, time point 2

Well there is no decent data set. If there was you'd be using it. It's soft data though; it's not a hard outcome. For me, measuring a temperature scale is a soft outcome, because it's so independent of me. So although we're calling it a 'quality marker', it's independent of me. If the surgery takes longer or the nurse breaks the heating device for the intrathoracic gas, what can I do? I cannot put any more heat in than I try to do, so actually – so my personal score goes down because of the failure of a service, or the cheapness of a service. So if one month, Procurement decides not to buy something, then I'm suffering. So it's a soft outcome – it's not totally under my control. If it was all under my control I'd take it much more seriously. But it isn't. But we're anaesthetists; very little is under our personal control, because as a team, we're part of a very large team, so unlike a surgeon firing a staple, it's soft.

Anaesthetist 20, time point 2

But again some of that – not to put it onto someone else's fault but a lot of that is dependent on the surgeon. If it's a junior surgeon doing the operation on a hernia he can take twice as long, and they're pulling on stuff, and they don't infiltrate the local anaesthetic properly; so there are lots of factors that might contribute to it.

Anaesthetist 14, time point 2

And I think, probably, the most of us get frustrated with the external factors particularly, you know, like you're sending for a patient and so you don't necessarily get to start on time and then your list might overrun and all this kind of stuff.

Anaesthetist 4, time point 1

I don't think it's actually got, you might be the most knowledgeable, experienced anaesthetist but actually the hospital doesn't have the drug or the Bair Huggers are broken, your, you've got inexperienced recovery staff, you know, all these other factors may not necessarily have anything to do with your qualifications or how up to date you are. It's actually got to do with the supporting external factors.

Anaesthetist 4, time point 1

But the best that I can do is put the hugger on the patient and make sure that he's covered and I can't improve much on that even if I want to.

Anaesthetist 23, time point 2

But, you know, when you have a cold theatre to start with and you have done absolutely everything you can to – you know, you give them warm fluids, you put a Bair Hugger on, you can't do any more, and yet their temperature when they hit recovery is still below and you're skewed because of that. And that's frustrating more than anything.

Anaesthetist 11, time point 2

I do remember 1–2 months I dropped down a bit so I thought was there anything I was missing out or something I was doing but I couldn't think of anything because I was doing all the stuff and giving the antiemetics and so it was probably just one of those things.

Anaesthetist 1, time point 1

No, I don't really make any changes because they all get Bair Huggers, they all get foot warmers, they all get all the post-op nausea and vomiting bits.

Anaesthetist 4, time point 1

Feedback as a socially situated cognition

Interviewees expressed clear examples of peers becoming experts throughout the life of the project. The most prominent example of this was linked to the fact that the clinical lead for the project was a consultant anaesthetist who also received feedback on his own performance and, therefore, was undergoing the same process simultaneously. Clinicians reported engaging with this peer expert to request guidance and support in extracting the most meaning from their personal data:

Just informally, you can go and chat to him, he, kind of, understands what we do, he's getting reports himself as well, I presume, and it's just nice to have that person there who I know who's approachable. It's not usually a bit query or anything, it's like, '[clinical lead], I just had this thing that I need to do', or he's like, 'No, it doesn't really mean anything' or, 'Yes'.

Anaesthetist 16, time point 2

So, no, it's been very good discussing it with [clinical lead] and just discussing what he's doing.

Anaesthetist 9, time point 2

Actually, no, but it is one of the things I want to do, not so much – there's a couple of things I do want to catch up with [clinical lead] and ask him about so, no, I haven't, but it is something I intend to, I just haven't seen him for ages.

Anaesthetist 15, time point 2

In this sense the clinical lead for the project took on the role of a local opinion leader and provided assurance to other clinicians that involvement in the initiative was worthwhile and credible:

[O]bviously if you have an anaesthetist who does it you would assume that he or she would understand directly how to handle the data and apply the appropriate amount of sensibility about it, let's put it that way.

Anaesthetist 18, time point 2

Second, through the within-specialty group work that was introduced in the latter stages of the project, opportunities were created to work with an expert peer to improve your performance in a particular specialty. In many cases clinicians reported processes of modelling and scaffolding being used to result in improvements in performance. Expert peers provided them with the appropriate levels of support to promote personal development without taking authority over decision-making and action:

It was really interesting because the person who came out top – it didn't matter who came out top, but, sort of, you could then see their anaesthetic and change your anaesthetic if it was different. Sometimes it just wasn't different and you wondered why, you know, and that was useful, I thought.

Anaesthetist 12, time point 2

Like, for example, when [clinical lead] emailed about the gynae practice, and things have changed according to what he said which is great, so that's made a change that we had an issue with our patients, as a group rather than individuals, and he was able to point that out and things were changed and he was able to say who was doing really well so that people could, maybe, talk to them and say, 'look what you're doing for your patients but I'm not doing'.

Anaesthetist 16, time point 2

But, as we've seen recently in the gynae pain project, it only helps if you unblind and say 'This person does something really, really well. What is it that you do so well?'

Anaesthetist 18, time point 2

Well, I think so, because if you see someone who's right tip-top and they don't have any patients who get sick or anything afterwards and they're doing exactly the same sort of anaesthetics as you, then it would be worth talking to them. What do they do differently to what you do?

Anaesthetist 21, time point 2

Because then you can go to speak to them and go, 'Well what are you doing that I'm getting wrong?' [Laughs] Or people can come to me and obviously ask the same! [Laughs]

Anaesthetist 22, time point 2

Even when expert peers were not formally identified as part of the project, clinicians took it upon themselves to request support from each other when reviewing their feedback and striving to change future professional practice. Communities of practice were evidenced and knowledge sharing was viewed positively:

Well I am not so close to everybody, because it is a big unit, 30 people, but I think the majority of people I work with on a daily basis are brilliant and I think that I can talk with them in the full transparency and I always ask for help and even if I have terrible doubts I will not be ashamed to ask what the hell I do wrong! Just check on me please!

Anaesthetist 2, time point 1

[B]ut if there is a problem like that, and you can't see how you can improve it, then you have got to work out what the barrier is and I suppose you might then need to talk to a colleague about that, because if you are having pain problems and you are doing everything you could do, it could be your epidural technique, it could be something.

Anaesthetist 1, time point 1

I'd probably drop in on them and see what they're doing and see how my practice differs from theirs, and see what I can learn from it.

Perioperative service lead 2, time point 1

Interviewees reported positive experiences with peer-based learning through social interaction, and requests were made for more formal processes to be put in place to encourage and support this in the long term:

And I would like to see at maybe every audit meeting, or every two audit meetings an update on where things are at and how the general things are. And maybe then give people the opportunity to say, this is what I've done. So that would be useful.

Anaesthetist 5, time point 1

I think it's useful getting the e-mails, I think it might be . . . I think it's useful . . . It would be useful to have verbal feedback at the audit meeting every couple of months, maybe, or . . . And it might be useful also to publicise people who've done something as a result of it. To say well, actually, as a result of this, so-and-so has done this to their practice. So having anecdotes sent to people.

Anaesthetist 5, time point 1

But I think the way it's happening is, effectively, eventually more and more people talk about it. So it's a matter of keeping it visible. I wasn't at the presentation [academic lead] did last . . . I said last week or a couple of weeks ago, I was away. But, actually, that kind of thing more often, I think, might be useful.

Anaesthetist 5, time point 1

I suppose, basically, the department meetings. We have an audit meeting once a month where these, sort of, things are discussed, meant to be – this is a fantastic audit, and we can talk about it and how it affects what we do.

Perioperative service lead 3, time point 1

Feedback as a threat to internal consistency

Clinicians reported instances of cognitive dissonance (i.e. when two or more conflicting cognitions occur simultaneously) when reviewing their performance feedback. This was often associated with discrepancies between what the data were telling them and their prior beliefs about themselves as a health-care professional:

We all go through a period of anger, disbelief, this can't be right . . . What's the lovely line. . .? Only 5% of men think that their driving is below average. And yet it has to be 50%. We all think we're great. Of course we do, it's part of being a self-confident doctor. And you look at it initially and you think, 'No, that can't be right. How can I be down here? Down at the bottom'.

Perioperative service lead 3, time point 1

I suppose I was surprised how bad I was, more than anything else.

Anaesthetist 6, time point 1

I thought: 'My goodness, I do quite a lot of patients'; 'my goodness, oh, some of them are in more pain than I thought they would be in'.

Perioperative service lead 2, time point 1

I always look, I look the results, and I'm always disappointed with the results because I think I work very hard and I'm not happy.

Anaesthetist 23, time point 2

The uncomfortable feeling of dissonance was resolved by clinicians in one of two ways. Some chose to reject the data as inaccurate/irrelevant, while others accepted them and attempted to make changes to their practice to ensure that future data improved:

I mean I was appalled to find that I am not the best in terms of postoperative analgesia, and I am not the best in postoperative temperature which actually I was thinking about and I thought well that probably reflects more the type of surgery I do, so if you do cardiac, major complex surgery, and major vascular surgery, they are the sort of patients that are more likely to have more complex pain problems or complex temperature problems, and so on. So I guess I need to be satisfied with my lot!

Anaesthetist 6, time point 1

But like all things, doctors are prone to self-denial, so if one month I have a terrible score, I shall just blame one of the trainees and clear my mind [interviewer laughs]. That's not my problem!

Anaesthetist 20, time point 2

Yes, I started off quite nauseas. I'm old, I use quite a lot of nitrus oxide, I notice that I was down below half way in my nausea and vomiting, I stopped using it and I got above half way. I did not believe before that that nitrus used by an intelligent mature and experienced man would influence PON-V, I was wrong, and that showed it to me. It also – what is interesting, I was convinced as a regional anaesthetist that I was very close to God in terms of analgesia and my patients had no pain. What I was doing was only concentrating on those patients whom I put a block in, and only going back into recovery to see those patients for the wonderful pleasure of getting the accolade from the patients saying, 'I've no pain.' I wasn't going to see the other operations where I couldn't do a block. What that said to me is, actually, some of my patients are in quite a lot of pain. So I went back to see them and they were, and since then I've, basically, given more morphine, quite simply, and I think it has had an influence. It's still not perfect but it's pushed that in that direction.

Perioperative service lead 3, time point 1

I have always thought I was a much better anaesthetist but I suppose one's self-worth is always higher than the actual, so that has been quite interesting actually. It has made me look harder at what I do, I suppose, so it is good from that point of view.

Anaesthetist 6, time point 1

Feedback as reinforcement for learning

There were many examples in the data of feedback providing a direct reward to the person using it. Rewards emerged in a number of different formats and at a number of different levels. On a practical level, receiving regular data on your performance provided support and resource for curriculum vitae building and revalidation preparation:

Well basically when I do . . . it is interesting because you can actually position yourself so when you are doing your CV [curriculum vitae] or when you are doing an application, you can actually say 'well actually I am in the top 10 of the anaesthetic department for quality in terms of these things'. So you can show that data. So actually it is probably quite useful to have as well, in other ways. You probably wouldn't say so if you were in the last 10%! So it has its benefits, I suppose.

Anaesthetist 6, time point 1

Has to be. Revalidation is are you competent? Are you good enough? How excellent are you? You can talk about – we can't talk about – we can't yet talk about big outcomes but this is really, really important stuff. We can say, 'Well, actually, this is my nausea, vomiting, pain, fitness discharge'. This is tangible data that we can demonstrate our excellence or our adequacy or whatever you wish. I think it will be fantastic.

Perioperative service lead 3, time point 1

In some cases, receiving feedback was viewed as a reward in the sense that it provided emotional reassurance to the recipient that they were performing at the appropriate level:

Yeah, certainly in terms of your own personal performance and you can say: 'I'm not rubbish because these are my figures'.

Perioperative service lead 2, time point 1

At the departmental level rewards were demonstrated in the sense that feedback could be used as evidence of collective high performance:

I think the department can use it to show how good it is, and I think that if we can demonstrate quality anaesthesia, I think that is ideal, and if we can show for example the trainees are giving quality anaesthesia, that reflects our teaching as well. So I do think it is good all round.

Anaesthetist 6, time point 1

Well, first of all, I can ask for a big bonus because, let's be honest, if I can show that my team have decreased nausea and vomiting, pain, increased temperature over time, that's a result. That's the continuous quality improvement thing. If I can show that during the period of this process we managed to improve in these outcomes, I can then ascribe the improvements in outcomes to the process. May not necessarily be true but I could, and that's fantastic. That would be lovely.

Perioperative service lead 3, time point 1

In terms of ongoing interaction with and use of feedback reports to make improvements, individuals also experienced a feeling of reward when changes that they made to their behaviour were evidenced in their next feedback report as an improvement:

Because you do need to get the buy-in. Getting the buy-in is hard if people don't see a tangible change. There is very little in there at the moment which is presented as tangible change in terms of a reward to the individual practitioner. And this is something which . . . At the moment, the tangible reward is, well, this will reflect better on me if I do better – 'Oh, isn't this terrible? My patients are not performing very well – that's quite humiliating.' But in terms of reward, the reward would have to be, oh, actually, last month, I was number 18 – actually, now, I'm number 2, and, actually, I'm always . . . It's almost emphasising that element, I think, might be strong and a strong thing to do.

Anaesthetist 5, time point 2

Feedback as a motivational tool

The feedback reports have provided anaesthetists with the opportunity to set goals for the purpose of individual-level improvement based on identification of where they are most needed:

Because it gives me a benchmark, that I could do better, I could do worse. There is something that I need to learn, there is something that I do right, so it gives me an idea.

Anaesthetist 2, time point 1

So these very basic data have called things to my attention. I'm now more obsessive about temperature control because the most objective is temperature, I know I can push that up, and so I now have hot air blowers on the patients in the anaesthetic room if I'm going to be in there for a while rather than leave them cooling off for 15 minutes, because you never catch that 15 minutes up. So, yeah, it's had an impact.

Perioperative service lead 3, time point 1

The importance of being able to test progress against goals was highlighted as being important. Interviewees described the process of making changes and testing them against the feedback reports that they received:

Oh no I said it would be very good that if changing once in a day I got 100% right but I don't expect it to be the case! I would be happy that changing something will just get the trend going better. But of course if it goes 100% that would be very good but I think the importance is to have monthly reports, related to the fact that you can see whether you are getting better with what you changed or not. If you still need to change something more or is it enough?

Anaesthetist 2, time point 1

No. But I think we should be doing it. I think we should be doing it and I think it would be really interesting for an individual anaesthetist if they could identify a particular area that they wanted to work on. Say, they, a particular type of operation they had a problem with pain that we could produce control charts for that particular procedure over time for that individual to see if the changes they were making were having any impact.

Anaesthetist 10, time point 1

I think it is useful for your annual review . . . when you do your appraisal then it's useful for you to be . . . to identify the area where you are maybe not as good as you thought, or . . . and to see which kind of improvements you achieve by changing your practice.

Anaesthetist 7, time point 1

And the other great advantage, the interest to me, not as chief of service but as an individual, is this gives you a fantastically powerful tool in which to say, let us change what we do and we can evaluate the effect of what we have done by looking at this data.

Perioperative service lead 3, time point 1

However, there seemed to be a disconnection between making a change to practice and reviewing future performance against the goal. Interviewees reported instances of making changes to their behaviour but being unable to identify improvements against their goal in their feedback reports:

That's an important piece of information because if everyone's saying that then we've got a problem. Because people are trying to change something and we're measuring it and they're not impacting the outcome. Which either means that what we're measuring is not a, you know, it may mean that what we're measuring isn't an actual accurate reflection of what they're trying to change and some sort of disconnect going on.

Anaesthetist 10, time point

I thought: 'My goodness, I do quite a lot of patients'; 'my goodness, oh, some of them are in more pain than I thought they would be in'. And I did some things to change it; so I changed my own practice a little bit, particularly on the gynaecology patients. Has there been any improvement? Probably not, according to the data.

Perioperative service lead 2, time point 1

Feedback as an innovation

Interviewees emphasised the importance of introducing initiatives, such as data feedback, gradually in order to increase buy-in at the individual level:

I think it's improving and I think this is helping to improve it. I think most departments are going from a level of having no feedback to having some feedback and there's the initial introduction of this feedback was done in a very stepwise, gradual, non-threatening way because we knew that that would cause problems otherwise. And I think that was very successful and I think as a consequence that people have now embraced this information a lot more. People have tried to do quality improvement processes by being more confrontational and ended up with absolutely nothing out of it, so I think things are improving.

Anaesthetist 10, time point 1

You have to convince everyone that this data is solid and it's validated and it's – yeah, you have to basically get everyone on board. Change doesn't happen overnight and it's important to get everyone – it's how to manage change rather than implement change.

Anaesthetist 24, time point 2

Because you do need to get the buy-in. Getting the buy-in is hard if people don't see a tangible change.

Anaesthetist 5, time point 2

Discussion

Summary of main findings

The deductive analysis allowed us to identify the presence of a number of relevant theories in our qualitative data set. It is clear that such pre-existing theories, from the fields of social sciences and health services research, are highly applicable in accounting for the mechanisms of effect for feedback. This is an important finding as it allows us to better contribute to the design of future feedback initiatives with a more informed understanding of the underlying mechanisms of effectiveness. The mechanisms that were found to be of greatest relevance to the experiences of our participants were feedback as a component of planned behaviour, feedback as a socially situated cognition, feedback as a threat to internal consistency, feedback as reinforcement for learning, feedback as a motivational tool and feedback as an innovation.

Relationship to the literature

The need for feedback to be viewed as an innovation that should be implemented gradually over time complements our earlier longitudinal findings. The increase of engagement and impact over time fits with expectations of a complex quality improvement initiative with phased implementation. When designing feedback initiatives researchers should encourage individuals to focus on the positive consequences of using feedback to change behaviour and reduce perceptions of any potential negative consequences. Prior research has demonstrated that the outcome expectancy of recipients influences the effectiveness of performance feedback.⁵² In a sense, the feedback reports encouraged clinicians to consider the effect of their behaviour on consequences to their personal data as well as consequences to the clinical outcomes of the patients that they treat.

The results of the analysis suggest that we should ensure that anaesthetists believe that relevant others across the department/organisation (i.e. their peers) want them to use the feedback and increase perceptions of the feedback reports as easy to use and act on (i.e. targeting self-efficacy). This can be achieved by reducing barriers and increasing facilitators. Lack of hospital resources, untrustworthy data and high baseline compliance have each been highlighted as barriers to effective feedback in the literature.^{98,107}

Individuals should be supported through experiences of dissonance when reviewing feedback to ensure that they resolve inconsistencies through behaviour change rather than message rejection (i.e. behaviour change should become an attractive and obtainable option). This may also interact with ideas around attribution of responsibility/blame, which inevitably vary from anaesthetist to anaesthetist. Receiving and acting on data about one's own individual performance is clearly linked to the concepts of professionalism and excellence. Anaesthetists associate their involvement in the initiative with their professional identity and the need to strive for excellence. The quantification of their performance enabled them to objectively assess their conceptual ideas of excellence, and this sometimes contradicted their professional identity and led to behaviour change. The importance of the search for excellence in anaesthesia has been highlighted in the literature.^{99,132} This work has emphasised a need for perioperative practitioners to be supported by their educators and organisations in achieving a higher level of performance rather than baseline competence. The concept of excellence has also been associated with the ability to seek challenges and learn from them in an ongoing cycle of development.¹³²

The emergence of feedback as a motivational tool comments on the interactions between what is 'actual', what is 'ideal' and what is 'possible'. Feedback can be viewed as a 'prompt' for action when there is a discrepancy between these concepts. There is a clear limit to how much improvement can be made at an individual level (i.e. there are only a set number of actions that can be taken before resource and clinical barriers come into play). However, interviewees generally felt that the individual anaesthetist has a degree of power to use the data in the way that they want to for their own professional development. This finding can

be linked to the emerging awareness of the need for active rather than passive feedback and the importance of goal setting.^{106,133} Passive feedback has been defined as the unsolicited provision of information with no stated requirement for action. Active feedback, on the other hand, occurs where the interest of the clinicians has been stimulated and engaged in aspects of practice, through the process of agreeing standards, involvement in continuing education, or consideration of the implications of the information for improving care.¹³⁴ Facilitative rather than directive feedback has been shown to enhance the effect of feedback for high-achieving groups that are undertaking complex tasks.⁴⁸ The effectiveness of feedback has been linked to the motivation of recipients and the presence of plans and strategies for improvement.^{52,98}

Social interaction around the feedback reports and their use at both the individual and departmental levels should be encouraged. Working as an anaesthetist can be viewed as a relatively solitary professional role. The presence of the feedback reports encouraged camaraderie and a sense of learning community across the department. This was demonstrated through interviewees' support of the identification of expert peers to interact with and learn from through mutual problem solving. It is further evidenced through the need to request formal processes for discussion and interaction rather than viewing this as something that would occur naturally. A lack of support management to clinical units has been evidenced in the literature as a barrier to effective feedback.⁹⁸ The identification and promotion of potential peer experts to support and guide others through processes of modelling, scaffolding and cognitive apprenticeship should be considered in the design of future interventions. The clearest example of the presence of a peer expert in this initiative was the clinical lead for the project who was also a consultant anaesthetist receiving feedback reports. The literature has emphasised the importance of appropriate leadership and, in particular, the need for peer-led feedback rather than feedback provided by an external group.⁹⁶ The source of feedback and the credibility that is attributed to it has been associated with its success.¹⁰⁴ In fact, feedback has been shown to change physicians' clinical performance when provided systematically over multiple years by an authoritative, credible source.¹³⁵

When feedback was viewed as a reward, levels of engagement increased. This could be linked to the fact that anaesthetists do not generally receive regular praise. They are not necessarily recognised and rewarded by their patients, as they are considered to be working 'behind the scenes' of clinical care. The initiative may have gone some way in reducing the anonymity of the anaesthetist. The reports shine light on the work that anaesthetists do which simultaneously increases visibility and provides a form of praise and positive reinforcement.

Recommendations for the design of future initiatives

It is vital that feedback initiatives are grounded in theory as far as possible. Identifying the mechanisms through which end-users engage with feedback goes some way in suggesting which theories should be prioritised in which circumstances. It seems that our initiative may have benefited from more formally arranged social interaction linked to reviewing and acting on feedback reports. Such forums would provide an opportunity for goal-setting and peer-based learning and guidance. *Table 9* presents a number of recommendations for theoretically underpinned feedback initiatives based on the mechanisms that we perceive to be of greatest relevance.

TABLE 9 Research recommendations linked to key theories

Mechanism	Related theories	Key references	Recommendations
Feedback as a component of planned behaviour	Theory of planned behaviour Theories of professional behaviour change	Ajzen (1991), ¹⁰⁹ Ashford <i>et al.</i> (1999), ¹²⁷ Grimshaw <i>et al.</i> (2002), ⁶³ Grimshaw <i>et al.</i> (2001), ⁶⁰ Oxman <i>et al.</i> (1995) ⁵⁹	In order for feedback to be effective we need to: <ul style="list-style-type: none"> ensure that anaesthetists believe that the consequences of using feedback will be positive for them as an individual ensure that anaesthetists believe that relevant others (i.e. their peers) want them to use the feedback ensure that anaesthetists perceive the feedback reports as easy to use and easy to act on. This can be achieved by reducing barriers and increasing facilitators increase education and knowledge surrounding feedback initiative. Use of educational outreach and reminders identify local opinion leaders increase awareness of costs/resources needed to achieve the desired clinical change use relevant theoretical and empirical literature to support change take an active rather than a passive approach increase facilitators to change and reduce barriers to change
Feedback as a socially situated cognition	Social constructivism Social cognitive theory	Vygotsky (1962), ¹²⁹ Bandura (1986) ¹³⁰	In order for feedback to be effective we need to: <ul style="list-style-type: none"> encourage social interaction around the feedback reports and their use at the individual and departmental levels identify and promote potential peer experts (i.e. models) to support and guide others through processes of scaffolding and cognitive apprenticeship
Feedback as a threat to internal consistency	Cognitive dissonance theory Self-affirmation theory	Festinger (1957), ¹¹⁰ Steele (1988) ¹²⁸	In order for feedback to be effective we need to: <ul style="list-style-type: none"> reduce the unpleasant experience of dissonance encourage anaesthetists to reduce dissonance by changing performance/behaviour and not by rejecting the information (feedback reports) as inadequate or incorrect use mechanisms that encourage anaesthetists to self-affirm while receiving feedback that could be construed as threatening

continued

TABLE 9 Research recommendations linked to key theories (*continued*)

Mechanism	Related theories	Key references	Recommendations
Feedback as reinforcement for learning	Instrumental learning Operant conditioning Behaviorism	Skinner (1948) ¹¹¹	In order for feedback to be effective we need to: <ul style="list-style-type: none"> ensure that individuals are rewarded rather than punished for engaging with their reports
Feedback as a motivational tool	Goal-setting theory Control theory	Locke (1968) ¹³¹	In order for feedback to be effective we need to: <ul style="list-style-type: none"> ensure that it creates measurable goals
Feedback as an innovation	Normalisation process theory Adoption of innovation model and diffusion of innovation theory Technology acceptance model (TAM) Organisational change theory and resistance to change	May and Finch (2009), ¹¹⁷ May <i>et al.</i> (2007), ¹¹⁸ Murray <i>et al.</i> (2010), ¹²⁰ Greenhalgh <i>et al.</i> (2004), ¹¹² Rogers (2004), ¹¹³ Chuttur (2009), ¹¹⁵ Davis <i>et al.</i> (1989), ¹¹⁴ Holden and Karsh (2010), ¹¹⁶ Gollop <i>et al.</i> (2004), ¹²¹ Gustafson <i>et al.</i> (2003), ¹²² Iles and Sutherland (2001), ¹²³ Pettigrew <i>et al.</i> (2001), ¹²⁴ Scott <i>et al.</i> (2003) ¹²⁵	In order for feedback to be effective we need to: <ul style="list-style-type: none"> ensure that the implementation process is coherent and makes sense to anaesthetists ensure that anaesthetists are able to cognitively participate in the implementation process ensure that the implementation process is collectively enacted ensure that the outcomes of the implementation process are reflexively monitored and assessed (i.e. has it worked? Is it useful?) identify and use the most effective communication channels to pass messages about the feedback initiative through increase the rate of adoption of the feedback initiative throughout the social system (department of anaesthetists) encourage highly respected anaesthetists (i.e. opinion leaders) to adopt the feedback initiative. This will create an instinctive desire for the innovation inject the innovation into a group of anaesthetists who would readily use the feedback initiative and provide positive reactions and benefits for early adopters of the innovation increase relative advantage, compatibility, simplicity, trialability and observability increase perceived usefulness increase perceived ease of use ensure that the initiative is based on a thorough exploration of the problem and customer (anaesthetist) need identify the advantages to users of the feedback initiative increase flexibility of design increase leader goals, involvement and support provide evidence of effectiveness

TABLE 9 Research recommendations linked to key theories (continued)

Mechanism	Related theories	Key references	Recommendations
			<ul style="list-style-type: none"> • identify key people to lead the change • be clear about goals and priorities • create and encourage a supportive organisational culture • identify and increase strengths and opportunities associated with the feedback initiative • reduce any weaknesses or threats that are associated with the feedback initiative • increase driving forces and reduce restraining forces • increase commitment, enrolment and compliance • increase ownership

Discussion and synthesis of qualitative findings

Value of qualitative work for this type of research

Qualitative work lends itself extremely well to the evaluation of complex quality improvement initiatives. It allows researchers to gain a detailed insight into the underlying processes and mechanisms that are at work from the perspective of end-users. It permits the exploration of how the characteristics of the individual, the local context and the intervention itself have interacted with one another and the effect that this has had on outcomes. These multifaceted relationships are far more difficult to capture through quantitative methodology. By conducting qualitative interviews you have the opportunity to investigate the reasoning and justification behind people's views and decisions. It was particularly important for us to understand how the feedback intervention functions as a technology of change, as this provides essential information concerning theoretical and practical explanations that can contribute to future developments of this and similar initiatives in the field of improvement science. In doing this, the evaluation is able to account for social, cultural and organisational elements as well as those associated with the design of the initiative itself.

Unanswered questions and limitations

The sample studied was a relatively small opportunity sample; however, it did include the perspective of three professional groups (consultant anaesthetists, perioperative service leads and surgical nursing leads) within the perioperative team and, therefore, effectively represented the end-user group within the primary site of our initiative. We therefore were able to confidently provide insight into and generate hypotheses about the phenomenon under study. Follow-up interviews were not conducted with surgical nursing leads. The reason for this was that the original interviews with these participants were undertaken over an extended time period. Only two consultant anaesthetists and one perioperative service lead were interviewed at both time points, which could have restricted our ability to capture longitudinal views. However, many of the people interviewed at time point 2 had been present at the hospital throughout the entire initiative (even if they had not completed an interview at time point 1). Therefore, they can still be considered as representing a longitudinal perspective on the effectiveness of the initiative.

We chose not to apply our deductive theoretical analysis to the surgical nursing lead data from time point 1. The reason for this was that the consultant anaesthetists had more individual-level control over using their feedback to change outcomes and this is what we wanted to explore through the theories. Data delivered to the surgical nursing leads represented a systems view and their use was confounded with a number of important barriers to patient transfer efficiency (reported in *Chapter 9*).

Responses to interview questions may have potentially been influenced by demand characteristics and social desirability issues, such as the sensitivity of the topic. Clinicians may have felt obliged to overplay the extent to which they were engaging with and making changes based on their feedback reports (especially as the programme was clinically led by a peer). It may have been the case that those who did not engage well with the initiative declined the opportunity to interview and therefore biased our results.

The research team were heavily involved in the operational development of the initiative and, as a result, issues of subjectivity and bias in data collection were a possibility. However, the importance of reflexivity in qualitative data collection and analysis was emphasised throughout the process and discussed at regular steering group meetings to raise awareness and combat the issue.

Suggestions for future research

In terms of the design of future initiatives it would be worthwhile to incorporate specific characteristics based on existing theory and then evaluate their success in meeting the aims of the programme. It would also be of interest to conduct a qualitative study that tracked a group of clinicians through the entire life cycle of a performance feedback initiative. This would ideally involve qualitative data collection before the introduction of the initiative to provide baseline information for the purpose of comparison. Owing to the pace of evolution of complex quality improvement initiatives it could be argued that interviewing a small number of participants at regular time points throughout an intervention would be more fruitful than focusing on larger samples.

Overall conclusions and recommendations

Based on a review of the relevant literature in quality improvement, it was hypothesised that the mechanisms of effect of this intervention would operate at two levels: (1) stimulation of local culture change, promoting more open and constructive use of feedback on quality of care delivered at unit level, and (2) stimulation of personal professional behaviour change through normative comparisons with colleagues and enhanced insight into variations and outcomes associated with personal practice. A realist qualitative exploration of the experiences and perceptions of end-users provided the ideal approach for investigating these hypotheses.

Issues associated with local culture and transparency were evident at both time points, which suggests that the intervention was not always existing in an entirely open and constructive environment. There was certainly an ongoing need to avoid associations with performance management if the intervention was to be trusted and engaged with by clinician end-users. However, it was noticeable that anaesthetists became more open with their data over time and were less demanding of complete anonymity. This was demonstrated through their involvement and positive appraisal of specialty-level group work and interaction with peer experts. In fact, there was a high demand for social interaction around the feedback in general and the need for this to be formalised was regularly expressed. This was linked to the fact that the professional nature of being an anaesthetist is generally solitary and, therefore, social interaction and disclosure does not automatically take place as a matter of course. It was considered to be hugely important to have a peer leading the project, which may go some way in explaining why the culture surrounding feedback became more open over time. Clinicians found it easier to place their trust in a fellow anaesthetist who was creating a collective culture for improvement than in somebody who was external to the department.

Interviewees put forward a multitude of use cases that demonstrate utilisation of the performance data to make improvements and test the effect of changes to behaviour on outcomes. This was particularly evident at time point 2, which suggests that the initiative had a cumulative effect. This is to be expected from an iterative quality improvement intervention with phased implementation. As time passes, end-users are more able to identify the benefits in having and using the feedback data, and this leads to an increase in engagement. It was evident in our analysis that benefits go wider than just having the opportunity to improve practice. In particular, levels of engagement were linked to recognising how the data could be

used to satisfy the requirements of revalidation and, therefore, save time and effort on the part of the anaesthetists.

Individuals have certainly engaged with the feedback throughout the project lifespan and clearly value its presence in their department. However, they also recognise a number of flaws and feel that they need more personalised feedback if they are able to effectively improve their professional performance in the long term. The need for personal feedback is a key finding of the project. End-users need to be confident that the data that they are using are clinically meaningful or else they are far more likely to engage in processes of message rejection.

It is clear that feedback is a multifaceted interaction between format, focus and recipient, and this should be taken into consideration in the development of interventions alongside recognition of complexity and context.⁴⁸ Clinicians identify numerous benefits of this type of initiative and identify particular factors which support the effective transition from data, to information, to action and improvement. These systems provide end-users with the evidence that they require to demonstrate their professional ability and identify any areas in which they would benefit from further support and training.

Health-care professionals face barriers at the organisational, departmental and individual level when they engage in behaviour change, and these must be carefully addressed by any feedback-based performance improvement initiative.⁶³ Any initiative will be more likely to achieve its desired results if it is developed and implemented alongside the goals and environment of the profession that it is trying to improve.¹⁰⁰ The findings from this realist qualitative evaluation of user perceptions and experiences are valuable for designing similar initiatives in the future that will effectively support clinicians in developing their practice. Qualitative views highlight key areas of challenge for existing and future systems, which must be overcome if such systems are to realise their full potential.

Chapter 7 Evaluative survey study

This chapter reports the rationale, methodology and findings from the survey work stream of the evaluation. The chapter begins with an integrated introduction and methodology section and is then divided into three key results/discussion sections reporting on different aspects of the survey data. *Results 1: psychometric analysis* reports the development and psychometric validation of a survey instrument to explore perceptions of data feedback. *Results 2: predictors of perceptions of data feedback usability* reports a multiple regression analysis to identify the significant predictors of perceptions of data feedback usability. Finally, *Results 3: longitudinal evaluation based on perceptions of data feedback* comments on how perceptions of data feedback change over time as a result of the initiative. **Box 2** contains all key research aims from the original project protocol with those of greatest relevance to the survey component presented in bold.

The survey work stream of the project addresses these aims by developing and validating a research tool to evaluate the effects of the initiative. The data collected using the tool have been used to serve a number of purposes which all contribute to understanding the characteristics of effective feedback.

BOX 2 Primary and secondary aims from original protocol

Research aims

To evaluate the impact of a departmental continuous quality monitoring and multilevel feedback initiative on the quality of anaesthetic care and efficiency of perioperative workflow within a London teaching hospital over a 2-year period. Data will be analysed at multiple time points over the course of the project in order to generate both formative and summative information to support development and evaluation of the initiative.

To employ a quasi-experimental time series design to provide robust evidence concerning the impact of a serial data feedback intervention on anaesthetic quality indicators while controlling for baseline variance.

To contribute to the evidence base for valid and reliable quality indicators for anaesthetic care including effective patient experience measures.

To document the main features of the data feedback intervention as it develops through the CLAHRC programme for replication at other sites, including definition of metrics, data processes, feedback format and action mechanisms.

To assess the perceived acceptability and utility of this information system for individual end-users, the clinical department and other organisational stakeholders, using a formative, mixed-methods design.

Introduction

Health systems internationally are attempting to understand how best to use the monitoring and regulation of care quality in order to maintain and drive up standards. Health-care professionals require appropriate support to interpret their performance and systematically improve over time. In the UK, clinician revalidation has been introduced as a mechanism to uphold and improve practice through continuous professional development.^{136,137} As part of this programme, clinicians are required to produce evidence of fitness to practice. The imminence of physician revalidation in addition to recent UK policy developments and the NHS Next Stage Review have resulted in widespread cultivation to improve quality of care through the measurement and reporting of quality indicators. The use of quality indicators is intended to enhance patient safety and experience, while also improving organisational efficiency and economic efficiency.¹ Such monitoring is essential to maintain and improve standards of care within organisations, allowing clinical units to understand variations in care, respond to these and subsequently evaluate their efforts.¹³⁸ Secondary plans to link payment of providers with clinical outcome performance suggest that monitoring of clinical outcomes will be imperative in the future.² To successfully monitor performance, valid outcome indicators are required.

Many professional organisations and governmental agencies are promoting the use of indicators, particularly in anaesthesia where perioperative mortality and morbidity is multifactorial.¹⁷ However, in anaesthesia, there are few verified quality indicators, and discrepancies exist regarding which outcome indicators should be measured and reported to meet targets.¹³⁷ Subsequently, there is considerable variation within the standards set.¹⁷ A systematic review conducted by Haller *et al.*¹⁷ identified 108 anaesthetic clinical indicators. Of those identified, 60% relied solely on expert opinion to determine validity. Sixty-two per cent of indicators had low levels of scientific evidence to determine 'how things should be done'. The challenges of developing sensitive and reliable quality indicators and patient satisfaction measures are well documented.^{6,17,139–141}

Hospitals have access to many sources of data that provide information on quality of care delivered.^{37,103,142,143} Such data include critical incidents, complaints, clinical outcomes, patient-reported outcomes and patient-experience data. Supporting personal professional development and continuous quality improvement at the individual and clinical unit levels, however, requires the effective design of quality monitoring systems capable of delivering accurate, timely and useful feedback to clinicians based on valid and reliable quality indicators.³⁷

Recent systematic reviews demonstrated that performance feedback to clinicians has a positive impact on behaviour and outcomes, resulting in small to moderate positive effects.^{107,144} Such effects have been displayed in terms of reduction of mortality rates and improved compliance with guidelines, among other outcomes.^{76,145–147} A number of important characteristics for effective feedback have been suggested by prior research, including the perceived validity and credibility of the data; their source and timeliness; the way units are benchmarked and the avoidance of individual profiling that could be misconstrued as punitive.⁹⁷ Tailoring feedback to the specific clinical setting has been shown to have a positive influence on its effectiveness along with ensuring that those issuing the feedback are perceived as experts.^{52,55,135,148} High intensity and frequency of feedback improves outcomes along with sustained monitoring.^{98,107,135,144,148,149} A number of strategies have been identified to support the effectiveness of performance feedback, including providing recipients with information on specific areas for improvement, action planning and educational components.^{52,55,98,107,135,144,148,149}

In addition to individual factors, the situation (contextual factors) in which feedback occurs influences its acceptance and effectiveness in improving performance. Assessment impels multiple tensions: on the individual, between individuals and involving the learning environment. Recognition of these factors, such as barriers and resistance to change, can provide an evidence basis for the success or failure of feedback interventions. Mann *et al.*¹⁵⁰ suggest that the cultural safety of seeking and receiving feedback needs to be increased, and that for success, interventions should be directed at various levels: the individual, the

collective and the institution. The departmental context in which feedback is administered may, therefore, be important, including factors such as local management processes, resource availability, staffing patterns, information infrastructure, organisational policies and local operating culture.^{108,132,151,152} However, previous work to reduce anaesthetic complications has hypothesised that individual-level feedback is critical for quality assurance.¹⁵³

The GMC Revalidation Framework consists of four domains: knowledge, skills and performance; safety and quality; communication; and partnership and teamwork. Many health-care systems are measuring clinical performance to improve quality. However, no absolute definition of 'quality' exists. The Joint Commission described quality of care through its different attributes: appropriateness, availability, continuity, effectiveness, efficacy, efficiency, prevention, respect and caring, safety and timeliness. To date, no gold standards exist for the development of quality indicators,¹⁵⁴ and currently only 40% of clinical indicators have been validated beyond face validity. This gives rise to concerns that important policy decisions stem from data with inadequate evidence base and experience.¹⁵⁵

It is only since 2001 that the potential for educational feedback at the level of the individual practitioner has come into focus.¹⁵⁶ Studies have identified important qualities of feedback; however, currently there is no tool to assess whether or not these characteristics are being utilised in current feedback interventions. At present, tools exist to measure performance with respect to clinicians (e.g. the Consultation Satisfaction Questionnaire, The Chronically Ill Patients Evaluate General Practice, The Patients Evaluate General/Family Practice and the Improving Practice Questionnaire) and hospitals (e.g. the World Health Organization PATH project). However, there is a need for methodological development to assess feedback, as valid and reliable performance feedback is required to assess fitness to practice.¹⁵⁶

To investigate the impact of the feedback initiative in achieving best practice in anaesthetics, we designed a survey, to be used pre and post intervention, to assess perceptions towards the feedback initiative and current quality indicators. It provides a basis on which to assess the amount of feedback being received on each quality dimension, allowing hospitals to prioritise areas of development, as well as comparing the level of feedback being received between multiple sites. In turn, this will enable multisite learning and strategic changes to be made.¹⁵⁷ This has similar objectives to the clinical risk management monitoring instrument developed by Briner *et al.*¹⁵⁸ Their primary aim was to develop an instrument that allows for continuous monitoring of current and planned developments within hospitals, allowing differentiation between various service qualities, repeated assessment and tracking of changes over time.

Understanding and monitoring the levels of feedback which clinicians receive and the influence of this on quality of care delivered will help in the tailoring of specific feedback programs and ascertain important factors for optimal results. The primary aim of this study was, therefore, to develop and validate the effectiveness of a questionnaire using psychometric methods to explore the underlying dimensionality in scale items. The secondary aims of the study were:

1. to investigate the characteristics of feedback that are perceived to be of most value to anaesthetists by analysing survey data to determine which factors predict perceptions of usefulness
2. to evaluate the effect of our feedback initiative on perceptions of local quality monitoring using pre and post measures.

Methods

Development of the measurement instrument

Design

The initial questionnaire items were developed based on a literature review of emerging theory in the area of data feedback for quality improvement. Potential survey items were drawn up, with measurement definitions and response scales. Iterations of the survey items were discussed and refined by a multidisciplinary research team, including three consultant anaesthetists and a social sciences researcher with experience in survey design. The survey was piloted using a cognitive walkthrough technique with two additional consultant anaesthetists (who were independent to the research team) in which presentation, item interpretation and wording was clarified.

Theoretical basis of the survey

Rationale for the included survey items was based on prior research and theory. Questions derived to determine quality indicator adequacy are supported by Wollersheim *et al.*,³⁷ who suggest that fundamental qualities of clinical indicators include relevance, validity, reliability and applicability. Moonsinghe *et al.*⁶ describe that anaesthetists should evaluate outcome measures against their potential to lead to improvements in standards of care and benchmarking. Hysong *et al.*¹⁵⁹ state that features of effective feedback involve the degree of perceived timeliness, data formatting and frequency of feedback delivery. Environment, individual opinion, knowledge and attitudes were also thought to be key. Overall quantification of clinicians' perceptions of quality indicator and reporting adequacy is important, as reported by Johnston *et al.*,¹⁶⁰ who described the need for awareness of professional attitudes to deliver positive benefits from audits.

The survey provides a basis on which to assess performance interventions: allowing hospitals to prioritise areas of development and make comparisons between sites. In turn, this will facilitate multisite learning and allow the formation of strategic changes.¹⁵⁷ This is of similar aim to the clinical risk management monitoring instrument developed by Briner *et al.*,¹⁵⁸ aiming to allow for continuous monitoring of current and planned developments within hospitals, enabling repeated assessment and allowing differentiation between various service qualities.

Instrument

The survey assessed feedback relating to standards of care delivered departmentally. The term feedback referred to quantitative data from measures and indicators, rather than anecdotal reports or conversations. A copy of the paper-based survey can be found in *Appendix 3*.

Forty standard items were grouped into four dimensions of quality (quality indicator adequacy, feedback adequacy, constructive use of data and workplace climate) and generic items which assessed whether or not clinicians received regular quantitative feedback on individual quality dimensions and overall perceptions on the level of feedback that clinicians receive.

Section A included the participant's personal details, which were used to obtain demographic data and to provide a means for follow-up measurement.

Section B comprised dichotomous questions, to obtain information on the dimensions of quality feedback received by participants (yes/no responses) and the level of feedback (true/false responses).

Section C assessed the effectiveness of the current feedback that participants were receiving. Participants were asked to rate the current feedback they received based on the quality indicators used (sections C1–4), the overall impression of the feedback (sections C5–9), the use of data for improvement (sections C10–15) and the attitudes to quality improvement within their working environment (section D).

Quantitative information for sections C and D was obtained using a scale rating from 1 to 8. For section C, 1 represented 'completely inadequate' and 8 represented 'excellent'. In section D, 1 represented 'strongly disagree' and 8 represented 'strongly agree'. Therefore, higher scores in sections C and D represent more favourable perceptions on effectiveness of current quality feedback and the working environment.

Item D12 was negatively phrased and, therefore, reverse-scored during the data analysis.

Survey components

The survey consisted of four parts:

- A. Personal details (name/hospital/specialty/grade/year of qualification/country of qualification).
- B. Focus of current quality feedback.
 - Dimensions of quality (do you receive regular quantitative feedback on each of the following dimensions of quality?):
 - clinical effectiveness of care
 - compliance with best practice guidelines
 - productivity and efficiency
 - financial performance
 - patient safety
 - patient experience.
 - Level of feedback (is the feedback you receive based on care delivered by the trust/department/individual?):
 - trust/department/personal.
- C. Effectiveness of current quality feedback.
 - Perceptions of current quality care indicators in your area:
 - comprehensiveness: the degree to which the data you receive are comprehensive and cover all important dimensions of care quality
 - relevance: the degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients
 - reliability: the degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of data over time
 - improvability: the degree to which indicators measure aspects of care that you and your unit can have a direct impact on through changing behaviour, the care process or local systems.
 - Perceptions of current feedback you receive:
 - level of analysis: the degree to which the data you receive are broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre, your patients)
 - timeliness: the degree to which the frequency of feedback you receive helps you to monitor how quality of care varies over time
 - means of communication: the degree to which the channel and method for dissemination (e.g. meetings, e-mail, reports, posters) are useful and engaging
 - data presentation: the degree to which the format in which data are presented (e.g. tables, graphs, scorecards) is clear and easy to use, with the right number of data presented
 - data credibility: the degree to which the data are viewed as credible and from a trustworthy, unbiased source.

- Perceptions of how quality-of-care data are used for improvement:
 - identifying problem areas and good practice: the degree to which data feedback helps us to rapidly detect problems and identify instances of excellent care
 - benchmarking: the degree to which the data feedback allows us to compare ourselves against similar units and/or national guidelines in a meaningful way
 - prioritising action: the degree to which data feedback supports prioritising where we put our efforts to improve care and which specific processes to focus on
 - setting measurable objectives: the degree to which data feedback supports setting quantifiable targets for improvement
 - monitoring progress: the degree to which data feedback supports evaluation of our progress towards targets over time and whether or not any gains are sustained
 - overall usefulness for improvement: the degree to which current data feedback is useful in monitoring variations and improving care.

- D. The environment in which you work.
 - In this department we are proactive in striving to continuously improve standards of care.
 - In this department we routinely review data on quality-of-care outcomes.
 - In this department it is clear as to what are acceptable standards of care.
 - In this department we monitor of compliance with best practice guidelines for clinical care.
 - In this department we respond to variation constructively, to improve care, rather than blaming and punishing individuals.
 - In this department we openly discuss minor failures in care to learn lessons.
 - In this department we review of critical incidents and serious failures to improve systems.
 - In this department we can demonstrate to senior levels of the organisation the quality of care that we are delivering.
 - In this department we are supported by the organisation in our efforts to collect and use data.
 - I have adequate knowledge and training on the statistics required to interpret quality-of-care data.
 - I have adequate knowledge and training in quality improvement methods.
 - In this department variations in the quality of care delivered to patients often go undetected.
 - In this department it is clear who is responsible for taking action to make changes and improve care processes.
 - In this department we use personal data on the quality of care that individuals deliver in a constructive way.
 - I am comfortable for the quality of care received by my patients to be monitored and fed back to me.
 - In this department we are effective in taking action based on audit and data feedback.

Data collection

Anaesthetists from two NHS trusts (Imperial College Healthcare NHS Trust and Royal Cornwall Hospitals NHS Trust) were recruited to complete the survey at baseline. Data were collected from participants using surveys administered at a single time point prior to any local development of quality monitoring programmes. The target sample was all practising anaesthetists at each site. Clinical leads at each site were tasked with identifying respondents and obtaining responses. A combination of paper-based and online survey administration was used for data collection.

The intervention was later evaluated using follow-up surveys at two further time points with anaesthetists from the primary site at Imperial College Healthcare NHS Trust. Post-intervention data were also collected at two additional sites at Imperial College Healthcare NHS Trust. At the primary site, time point 2 data were collected approximately 12 months after baseline data, when anaesthetists were receiving simple feedback. Time point 3 data were collected approximately 27 months after baseline data and 15 months after time point 2 data, when anaesthetists were receiving enhanced feedback. At the two additional sites,

time point 2 data were collected approximately 12 months after baseline data, when anaesthetists were receiving enhanced feedback. The target sample was all practising anaesthetists at the three sites. Clinical leads at each site were tasked with identifying respondents and obtaining responses. Online survey administration was used for data collection. *Table 10* provides an overview of the data collection time points for each site.

Analysis

Psychometric analysis of scale dimensionality

Internal consistency was assessed using Cronbach's alpha, and the values of 'alpha if item deleted' were given. To demonstrate acceptable internal consistency, survey scales should display Cronbach's alpha values > 0.7 , with no further increments in Cronbach's alpha value if items are deleted. The removal of any item from each domain should have no effect on internal consistency. Corrected item-total correlation within each domain refers to the correlation between one item and the rest of the scale (the total of other items in the same domain). Criterion-related validity was assessed by measuring the strength of the association between the aggregated scale of generic items that measured whether or not feedback was received and the other domains using Pearson's correlation.

The item discrimination was assessed by comparing the difference in mean scores of the two domains between those participants who did receive feedback and those who did not. Statistically different scores across each survey domain, reflecting different perspectives on indicator efficacy and departmental climate, suggest the survey's ability to distinguish between these two groups. Overall, two groups of *t*-tests were performed, assessing the relationship between section B of the survey (quantifying the levels of feedback) with perceptions on indicator efficacy (section C) and perceptions of departmental climate (section D).

Hypotheses for these *t*-tests were:

1. Clinicians who receive regular quantitative feedback will perceive quality indicators and feedback on quality improvement more positively than those who have not received regular feedback. Positive perceptions were indicated by higher ratings (1–8) on each domain item.
2. Clinicians who receive regular quantitative feedback will perceive that their department displays a more positive operating culture than those who do not receive regular feedback. Positivity in departmental operating culture is reflected by higher ratings (1–8) on each domain item.

Finally, exploratory factor analyses (with varimax rotation, for inter-reliability) were conducted to explore the factor structure. Variables were isolated into factors based on their factor loading (the degree of association of each variable with a factor; -1.00 to $+1.00$). The criterion for acceptance of latent variables

TABLE 10 Data collection time points by site and level of feedback

Time point	2010	2011	2012	2013
SMH time point 1 (no feedback)				
SMH time point 2 (simple feedback)				
CXH time point 1 (no feedback)				
HH time point 1 (no feedback)				
SMH time point 3 (enhanced feedback)				
CXH time point 2 (enhanced feedback)				
HH time point 2 (enhanced feedback)				

CXH, Charing Cross Hospital; HH, Hammersmith Hospital; SMH, St Mary's Hospital.
Shaded cells indicate time periods of data collection for each study component.

into common factors was a factor loading value of 0.6. Factor variables correlate substantially within dimensions but not between dimensions. Two separate analyses were performed to ascertain the latent factors affecting perceptions on quality indicators and feedback (section C), and to ascertain the underlying dimensionality of departmental workplace climate (section D). Factors were suppressed from the output if variables had loading factors of < 0.4 . In line with the structure of our survey measure, we would expect that section C will be divided into three main latent factors: quality indicators used to collect data, feedback of data and subsequent action from feedback to aid quality improvement.¹⁶¹ We hypothesised that there are readily identifiable, reliable, latent factors underlying the scale.

Multiple regression analysis

A statistical modelling approach was used to investigate a number of hypotheses, in line with prior research and theory, concerning the influence of individual factors, the departmental context, and a range of feedback content and design characteristics that have been suggested to be important determinants of feedback effectiveness.^{76,153}

As mentioned previously, the survey measure comprised four items assessing the comprehensiveness of local quality monitoring, that is whether or not clinicians received regular quantitative feedback on a number of individual quality dimensions. These items were summed into a scale representing comprehensiveness of monitoring for the purposes of the regression analysis. Responses to the 16 items evaluating the departmental climate for quality improvement were measured on an 8-point Likert agreement scale ranging from 1 (strongly disagree) to 8 (strongly agree) and aggregated into a single-scale score for the purposes of this analysis. *Box 3* provides examples of the variables included, along with an internal consistency metric (Cronbach's alpha) for the aggregated score. The dependent measure for the study was based on a single survey item, 'overall usefulness for improvement', defined as 'the degree to which current data feedback is useful in monitoring variations and improving care' and rated on an 8-point scale ranging from 1 (completely inadequate) to 8 (excellent).

BOX 3 Examples of the main areas covered by a 16-item scale designed to measure departmental climate for quality improvement. Individual items were rated on an 8-point Likert agreement scale

Processes of monitoring and quality improvement at departmental level.

Constructive response to observed variations in care.

Openness of professional climate for discussing failures.

Organisational support for departmental quality improvement initiatives.

Professional competency with quality improvement and statistics among clinicians.

Responsibility for acting on observed variations in care.

Willingness to disclose personal performance data to the department.

Cronbach's alpha of the combined items = 0.91.

Multiple linear regression analysis with hierarchical variable entry was performed to contribute to the evidence base for valid and reliable quality indicators for anaesthetic care, and to assess the perceived acceptability and utility of this information system for individual end-users. The following research hypotheses were tested using hierarchical entry of specific predictors in steps:

- Hypothesis 1: length of time since qualification (tenure) will influence perception of the degree to which current local data feedback is useful for monitoring variation and improving care in anaesthesia.
- Hypothesis 2: organisational membership will influence perception of the degree to which current local data feedback is useful for monitoring variation and improving care in anaesthesia.
- Hypothesis 3: the reported local departmental climate for quality improvement will influence perception of the degree to which current local data feedback is useful for monitoring variation and improving care in anaesthesia.
- Hypothesis 4: the reported scope of local quality monitoring will influence perception of the degree to which current local data feedback is useful for monitoring variation and improving care in anaesthesia.
- Hypothesis 5: the design and content characteristics of feedback will influence perception of the degree to which current local data feedback is useful for monitoring variation and improving care in anaesthesia.

The statistical significance of the additional proportion of variance in the dependent measure accounted for by each successive entry of variables was assessed in order to establish the role of each specific hypothesised predictor, having controlled for previously entered factors.¹⁵³ Forced entry regression was selected as the most appropriate method, as pre-existing research findings and theory are available to support the order of causal and temporal priority among the independent variables.

Longitudinal analysis

Descriptive statistics were run to compare the mean, standard deviation (SD), median and range across the time points. A one-way analysis of variance (ANOVA) (accompanied by Bonferroni post-hoc tests) was used to test for significant differences between time points for the primary site. Independent samples *t*-tests were used to test for significant differences between time points for the two additional hospital sites.

Results 1: psychometric analysis

Demographics

Eighty-nine respondents from two acute health-care organisations participated in the study. This represents a response rate of 59% (70% for acute health-care organisation A and 48% for acute health-care organisation B). Eighty-two (92.1%) participants were consultants, six (6.7%) were trainees and one (1.1%) was non-consultant faculty. Anaesthetists included were from a mixture of specialties typical of a large, urban, academic teaching hospital.

Internal consistency of scales

The level of internal consistency is displayed in *Table 11*. The values of Cronbach's alpha for each domain ranged from 0.91 to 0.97, indicating excellent internal consistency ($\alpha > 0.7$). There were significant results based on the value of 'alpha if item deleted'. The removal of any item from subsections (C01–04, C05–09, C10–15 and C01–15) of the domain evaluating the effectiveness and usefulness of quality indicators and feedback (section C) would not augment the internal consistency (Cronbach's alpha would remain at 0.953, 0.93, 0.966 and 0.968, respectively). However, the removal of two individual items from the overall domain assessing the local departmental climate for effective use of feedback (section D) revealed improvements in internal consistency – 'detecting variation' (D12) and 'comfortable with monitoring' (D15) – which augmented internal consistency from 0.912 to 0.917 and 0.916, respectively.

TABLE 11 Internal consistency of survey scales

Dimension	Number of items	Cronbach's alpha
C01–04	4	0.953
C05–09	5	0.930
C10–15	6	0.966
C1–15	15	0.968
D	16	0.912

Based on the corrected item-domain correlation, all items of each domain should be well correlated with the other items from the same domain. All evaluative subsections concerning the effectiveness and usefulness of quality indicators and feedback (section C) showed strong positive correlations (C01–04, 0.845–0.914; C05–09, 0.763–0.893; C10–15, 0.864–0.933; C1–15, 0.715–0.862). Perceptions of the local departmental climate for effective use of feedback showed mixed results, with a weak positive association with the item assessing 'comfortable with monitoring' (D15). All other items within each domain showed a moderate to very strong positive relationship ranging from 0.363 to 0.754.

There was a positive correlation between the aggregated scale for items identifying the components of quality monitored (section B) against both perceptions on the effectiveness of quality indicators and feedback (section C) ($0.404 \leq r \leq 0.484$; $p < 0.01$) and perceptions of local departmental climate for effective use of feedback (section D) ($r = 0.273$; $p < 0.05$).

Scale validation

We employed *t*-tests to investigate item discrimination. The survey should be able to distinguish between those who receive regular quantitative feedback and those who do not. The two experimental conditions were participants who received regular quantitative feedback on each of the four dimensions of quality and participants who did not receive regular quantitative feedback on the four dimensions of quality. If the scale displays good item discrimination, we would expect a statistically significant difference in scores throughout the survey between these two groups. Separate *t*-tests were conducted for each dimension with sections C and D as the dependent variables:

1. clinical effectiveness of care (B01)
2. compliance with best practice guidelines (B02)
3. patient safety (B05)
4. patient experience (B06).

Similarly, *t*-tests were conducted based on whether or not a participant receives feedback at each organisational level, with sections C and D as the dependent variables:

1. feedback based on the care delivered by the trust or hospital
2. feedback based on the care delivered by the department
3. feedback based on the care delivered by the individual.

Clinicians who received regular quantitative feedback on each dimension of quality and at each organisational level had more positive perceptions on the effectiveness of quality indicator data and use of feedback to aid quality improvement, for example feedback on compliance with best practice guidelines [$t(30.63) = -3.00$; $p < 0.01$, with a difference between the mean perceptions of overall effectiveness of 1.34].

Only clinicians who received regular quantitative feedback delivered by the trust [$t(83) = -2.10$; $p < 0.05$, with a difference between the means of 0.77] and care delivered by the department [$t(83) = -2.34$; $p < 0.05$,

with a difference between the means of 0.87] perceived strong departmental climate for effective use of feedback.

Exploratory factor analysis

Two exploratory factor analyses with varimax rotation were conducted to explore the underlying dimensionality of the survey.

Perceived effectiveness and usefulness of quality indicators and feedback (section C)

All items within this domain correlated well. Interitem correlations ranged from 0.38 to 0.91. Communalities effectively indicate the reliability of a common variable, by determining how well items load onto a dimension. The output of variables with communalities < 0.4 was suppressed. The range of communalities was 0.60–0.91. Following examination of scree plots and varimax factor rotation, two factors were retained. These factors represented the latent variables, explaining 80% of the total variance in perceived effectiveness and usefulness of quality indicators and feedback:

- factor 1: how the data are presented
- factor 2: usefulness of the data in achieving quality improvement targets.

The rotated component matrix is displayed in *Table 12*. Descriptions of each factor and its underlying components are displayed in *Table 13*.

TABLE 12 Rotated component matrix

Survey item	Component	
	1	2
C08 Data presentation	0.871	
C09 Data credibility	0.840	
C01 Comprehensiveness	0.816	
C03 Reliability	0.813	
C05 Level of analysis	0.806	
C04 Improvability	0.799	0.424
C02 Relevance	0.793	
C06 Timeliness	0.737	
C07 Communicability	0.724	
C11 Benchmarking		0.895
C13 Setting measurable objectives		0.885
C12 Prioritising		0.882
C14 Monitoring progress		0.867
C10 Identifying problems	0.492	0.753
C15 Usefulness	0.523	0.745

Extraction method: principal component analysis.

Rotation method: varimax with Kaiser normalisation.

Rotation converged in three iterations. The Kaiser–Meyer–Olkin (KMO) statistic was superb – 0.913 (KMO > 0.5) – and Bartlett’s Test of Sphericity was significant at < 0.0001, indicating that factor analysis was appropriate.

TABLE 13 Latent factors attributable to the variance in perceived effectiveness and usefulness of quality indicators and feedback

Factor title	How the data are presented	Usefulness of the data in achieving quality improvement targets
High loading components (top three)	<ul style="list-style-type: none"> Data presentation Data credibility Comprehensiveness 	<ul style="list-style-type: none"> Benchmarking Setting measurable objectives Prioritising
Description of concept	<ul style="list-style-type: none"> Methods of communicating and presenting the data that optimise perceptions of quality indicator feedback effectiveness PRESENTATION 	<ul style="list-style-type: none"> How the feedback of data allows clinicians to make quality improvement changes USE

Perceived local departmental climate for effective use of feedback (section D)

Most items within this domain correlated well. Inter-item correlations ranged from -0.08 to 0.85 . Communalities ranged between 0.51 and 0.87 . Following examination of scree plots and varimax factor rotation, three factors were retained. These factors represented the latent variables, explaining 66% of the total variance in departmental climate for effective use of feedback:

- factor 1: an open culture for proactive use of data
- factor 2: review and interpretation of data on quality of care
- factor 3: knowledge and resources for measurement-driven improvement.

The rotated component matrix is displayed in *Table 14*. Descriptions of each factor and its underlying components are displayed in *Table 15*.

TABLE 14 Rotated component matrix

Survey item	Component		
	1	2	3
D06 Open discussion of failures	0.844		
D07 Review critical incidents	0.769		
D16 Audits and action	0.759		
D15 Comfortable with monitoring	0.689	-0.400	
D05 Respond constructively	0.683		
D01 Proactivity in improving care	0.651		
D03 Clear standards	0.583		
D13 Responsibility for change	0.519	0.416	
D12 Detecting variation		0.740	
D02 Routine data review	0.465	0.648	
D14 Use of personal data		0.621	
D08 Demonstrability to seniors	0.419	0.596	
D04 Monitor compliance	0.438	0.543	0.442
D10 Statistical knowledge			0.930
D11 Knowledge of quality improvement			0.887
D09 Support from organisation		0.504	0.596

Extraction method: principal component analysis.
Rotation method: varimax with Kaiser normalisation.
Rotation converged in six iterations. The Kaiser–Meyer–Olkin (KMO) statistic was very good – 0.846 ($KMO > 0.5$) – and Bartlett’s Test of Sphericity was significant at < 0.0001 , indicating the appropriateness of factor analysis.

TABLE 15 Latent factors attributable to the variance in perceived positive departmental culture

Factor title	An open culture for the proactive use of data	Review and interpretation of data on quality of care	Knowledge and resources for measurement-driven improvement
High loading components (top three)	<ul style="list-style-type: none"> Open discussion of failures Review of critical incidents Audits and action 	<ul style="list-style-type: none"> Detecting variation Routine review of data Use of personal data 	<ul style="list-style-type: none"> Statistical knowledge Knowledge of quality improvement Support from the organisation
Description of concept	<ul style="list-style-type: none"> Factors in departmental culture which involve active use of data and actions to implement quality improvement changes ACTION 	<ul style="list-style-type: none"> Factors and methods which allow departments to detect and monitor the quality of care they are providing AUDIT 	<ul style="list-style-type: none"> Factors which are necessary for departments to participate in and sustain quality improvement initiatives SKILLS + RESOURCES

Discussion

Scale validation

Analysis of the survey demonstrated statistically robust evidence supporting its validity. Internal consistency was excellent, with all values of Cronbach's alpha exceeding 0.9. It particularly supported the grouping of items pertaining to the perceived effectiveness of quality indicators and feedback. In general, the scale reliability of section C of the survey, assessing perceptions of the effectiveness of feedback, was stronger than section D, which assessed positivity and proactivity to quality improvement in the working environment. The structure was further supported by values of Cronbach's alpha if deleted, which showed that the removal of any item from this domain would have no effect on the internal consistency of the survey.

In contrast, the removal of two items, 'detecting variation' (D12) and 'comfortable with monitoring' (D15), from the overall domain assessing perceptions of the local departmental climate for effective use of feedback (section D) would augment internal consistency. The potential confliction of these weaker components of section D may be due to their multifactorial nature. In particular, experience, seniority and personality may influence behaviour and comfort with the monitoring of personal performance. Certain individuals may be proactive in quality improvement and achieving best practice standards, but may not endorse being individually monitored. Moreover, workplace dynamics such as the incidence and management of adverse events and critical incidents will influence perceptions on monitoring. Current literature sanctions non-punitive, open environments, with active leadership as favourable workplace dynamics, influencing departmental climate for proactivity. Bradley *et al.*⁵⁸ highlighted that people may not be comfortable with personal data feedback reflecting their own clinical practice, which may be perceived as punitive. Such physician profiling, and its impact on improving personal performance, is further influenced by the use of personal data and interpretation/understanding of the data. Huppelschoten *et al.*¹⁶² concluded that we need to target physicians' desire to change and their ability to translate performance feedback into improvement strategies.

Results from the *t*-tests showed that the survey can make clear distinctions between those who have received more regular quantitative feedback and those who have not. These differences are also reflected in domain scores of the survey, assessing perceptions of quality indicator efficacy and departmental climate. The generic domain, which ascertained the presence of feedback in general and quantified the monitoring of different components of quality (B01–09), was significantly related to the perceived effectiveness and usefulness of quality indicators and feedback (section C). Clinicians who received regular quantitative feedback on each dimension of quality had higher scores, reflecting perceptions on the effectiveness of quality indicators and their feedback to aid quality improvement. Of note, only two items, 'trust feedback' (B07) and 'departmental feedback' (B08), of the generic domain were significantly related to perceptions of local departmental climate for effective use of feedback (section D). All levels of feedback, whether aimed at care delivered by the trust, the department or individually, showed significantly higher scores in its perceived effectiveness if received regularly.

Exploratory factor analysis

From the exploratory factor analysis, we can infer that the presentation and format of data that are fed back to clinicians as well as their usefulness in making quality improvement changes account for 80% of the variance in the perceived effectiveness of feedback. This supports two out of the three subdimensions in our survey structure. Three factors were identified to explain 66% of the variance in departmental climate for effective use of feedback. From this, we can infer that departmental climate is multidimensional, primarily comprising an open culture for the proactive use of data, the review and interpretation of data on quality of care and knowledge/resources for measurement-driven improvement.

In view of the recently published Francis Report,³ these findings are particularly relevant. There is a growing body of literature surrounding safety culture,^{163–173} however, little work has addressed which aspects of safety culture are critical. This analysis reveals the fundamental components of departmental safety climate which need to be accounted for to target this subset of organisational culture.^{3,165}

Emerging literature supports holistic organisational change. The first empirical evidence of a link between safety climate and organisational culture was seen by Singer *et al.*¹⁷⁰ Studies highlight the need for engagement at all levels of an organisation.¹⁶³ As well as the departmental level, this includes targeting hierarchy, management and leadership.^{166,168,170,172,173} To date, interventions have been designed to target hospital management and leadership. Frankel's intervention, Patient Safety Leadership Walkrounds, has been widely adopted.^{165,172}

There is consensus that local ownership is needed. Facilitating factors are teamwork, staff engagement and communication, as well as organisational characteristics such as openness and transparency.^{164,168,169,172,173} These results suggest that there is scope for modification of the survey to reflect these newly identified subdimensions – incorporating two subsections to evaluate perceptions of feedback effectiveness, and three subsections to assess departmental climate. Other amendments may include increasing the number of reverse scored items to maintain consistency and assess reliability of responses as well as reordering components of the questionnaire to limit clustering of similar items.⁵⁸

Results 2: predictors of perceptions of data feedback usability

Descriptives

Eighty-nine respondents from two acute health-care organisations participated in the study. This represents a response rate of 59% (70% for acute health-care organisation A and 48% for acute health-care organisation B). Eighty-two (92.1%) participants were consultants, six (6.7%) were trainees and one (1.1%) was non-consultant faculty. Anaesthetists included were from a mixture of specialties typical of a large, urban, academic teaching hospital. Following exclusion due to missing data, 78 survey responses were included in the regression analysis.

Seventy-six per cent of participants had been qualified (medical undergraduate degree) for between 11 and 30 years and the mean length of time since qualification was 20 years (SD = 8.1 years). Respondent characteristics are presented in *Table 16*.

The overall scope of local quality monitoring, with a mean value of 0.85 (SD = 1.20), was notably low (from a range of 0–4). This was reflected in the amount of feedback being received by participants on both levels of care (departmental and individual). The dependent variable, with a mean value of 2.83 (SD = 2.01), indicates that perceived usefulness of feedback for monitoring variations and improving care at these organisations was generally low. *Table 17* presents categorical items and their frequency of responses, while *Table 18* presents the mean scores and SDs of all scale items included in the regression model.

Regression analysis

The statistical model parameters of the different stages of the regression analysis examining the significance of the hypothesised predictors of usefulness of data feedback are given in *Table 19*. Regarding study hypotheses 1 and 2, neither tenure nor organisational membership significantly predicted perceived usefulness of current data feedback. The departmental climate for quality improvement (hypothesis 3) explained an additional 27.5% of the variance in the usefulness measure ($p < 0.0001$). The stronger the perception of a departmental climate for quality improvement, the greater the perception of the degree to which current local data feedback was viewed as useful for monitoring variations and improving care. In the third model in *Table 19*, partialling out the effects of all prior predictors resulted in departmental

TABLE 16 Respondents' characteristics: descriptive information about the clinicians that were included in the analysis

Characteristic	Cases, n (%)
Length of time since qualification (tenure) (years)	
0–10	10 (13)
11–20	31 (40)
21–30	28 (36)
31–40	8 (10)
41–50	1 (1)
Acute health-care organisation	
A	52 (67)
B	26 (33)

TABLE 17 Descriptive statistics of survey items: percentage responses to categorical variables, with number of responses shown in brackets, split by acute health-care organisation

Level of care focused on by quality monitoring	Acute health-care organisation A, % (n)	Acute health-care organisation B, % (n)	Total, % (N)
I receive monthly or more regular feedback concerning the care delivered by my department			
True	23 (12)	19 (5)	22 (17)
False	77 (40)	81 (21)	78 (61)
I receive monthly or more regular feedback concerning the care that I delivered personally			
True	15 (8)	8 (2)	13 (10)
False	85 (44)	92 (24)	87 (68)

TABLE 18 Descriptive statistics of survey items: mean responses to scale variables, with SD shown in brackets, split by acute health-care organisation

Survey item	Acute health-care organisation A, mean (SD)	Acute health-care organisation B, mean (SD)	Total, mean (SD)
Scope of local quality monitoring (0–4)			
Aggregated scale of categorical responses to the four quality dimensions (clinical effectiveness of care; compliance with best practice guidelines; patient safety; and patient experience)	0.83 (1.15)	0.88 (1.31)	0.85 (1.20)
Perceptions of the current quality indicators in your area (1–8)			
Comprehensiveness: the degree to which the data you receive are comprehensive and cover all important dimensions of care quality	2.48 (1.58)	2.46 (1.66)	2.47 (1.59)
Relevance: the degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients	2.63 (1.74)	2.42 (1.72)	2.56 (1.73)
Reliability: the degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of the data over time	2.77 (1.83)	2.54 (1.79)	2.69 (1.81)
Improvability: the degree to which indicators measure aspects of care that you and your unit can have a direct impact on through changing behaviour, the care process or local systems	3.21 (2.12)	2.58 (1.84)	3.00 (2.04)
Perceptions of the current feedback that you receive (1–8)			
Level of analysis: the degree to which the data that you receive are broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre or your patients)	2.46 (1.71)	2.35 (1.83)	2.42 (1.74)
Timeliness: the degree to which the frequency of feedback you receive helps you to monitor how care quality varies over time	2.50 (1.76)	2.50 (1.88)	2.50 (1.79)
Means of communication: the degree to which the channel and method for dissemination (e.g. meetings, e-mail, reports or posters) are useful and engaging	2.87 (1.86)	3.35 (2.10)	3.03 (1.94)
Data presentation: the degree to which the format in which data are presented (e.g. tables, graphs or scorecards) is clear and easy to use with the right number of data presented	2.56 (1.81)	3.04 (2.01)	2.72 (1.88)
Data credibility: the degree to which the data are viewed as credible and from a trustworthy, unbiased source	2.56 (2.07)	2.50 (1.77)	2.54 (1.97)
Local departmental climate for quality improvement (1–8)			
Aggregated scale of all 16 scale items	4.67 (1.25)	4.77 (1.36)	4.71 (1.28)
Dependent variable: usefulness of current local data feedback (1–8)			
Overall usefulness for improvement: the degree to which current data feedback is useful in monitoring variations and improving care	2.83 (2.09)	2.85 (1.87)	2.83 (2.01)

TABLE 19 Model summary: overview of statistics illustrating model fit for each of the five study hypotheses

Model sequence and description	R	R ²	R ² change	F change	Significance of F change
Tenure (hypothesis 1)	0.14	0.02	0.02	1.60	0.21
Tenure + trust (hypothesis 2)	0.14	0.02	0.00	0.002	0.97
Tenure + trust + departmental climate for quality improvement (hypothesis 3)	0.54	0.30	0.28	28.94	< 0.0001
Tenure + trust + departmental climate for quality improvement + scope of local quality monitoring/level of care focused on (hypothesis 4)	0.64	0.41	0.11	4.48	0.006
Tenure + trust + departmental climate for quality improvement + scope of local quality monitoring/level of care focused on + generic characteristics of feedback (hypothesis 5)	0.82	0.67	0.26	5.55	< 0.0001

climate for quality improvement making a significant positive contribution to the dependent variable ($\beta = 0.83$; $p < 0.0001$). When hypothesis 4 was investigated, the scope of local quality monitoring explained a further significant 11.2% of the variance in local usefulness of data feedback ($p = 0.006$). In this model, both comprehensiveness of feedback received ($\beta = 0.45$; $p = 0.02$) and provision of feedback at the level of the individual clinician ($\beta = 1.19$; $p = 0.049$), as opposed to department-level feedback, were significant predictors of local usefulness, once prior factors had been controlled for.

In the final fitted model (hypothesis 5), a number of variables representing feedback design and content characteristics were entered in the model, after controlling for all prior entered factors, including tenure, organisational membership, local contextual factors and the scope of any local quality monitoring initiatives. Feedback characteristics explained a further 26.4% of the variance in perceived local usefulness ($p < 0.0001$). The final model demonstrated that with the effects of all other factors held constant, two feedback content characteristics were significant predictors of usefulness (Table 20). These were the perceived relevance of the quality indicators to the specific service area ($\beta = 0.64$; $p = 0.01$) and the perceived credibility of the data as coming from a trustworthy, unbiased source ($\beta = 0.55$; $p = 0.01$).

Discussion

This study aimed to investigate the role of a range of demographic, contextual and feedback content/design characteristics in predicting the perception of usefulness of data feedback on quality of care. In investigating these aims we sought to answer the question of what are the important features of data feedback and quality monitoring initiatives that facilitate effective local monitoring and reduction of variations in care.

Neither tenure (hypothesis 1) nor organisational membership (hypothesis 2) significantly influenced perceptions of usefulness, demonstrating that there were no significant differences in perceptions of current local feedback attributable to professional experience or owing to location at either of the study sites. Variations in perceptions of the local departmental climate for quality improvement (hypothesis 3), however, was a significant predictor accounting for a large proportion of the variance in the dependent measure (27.5%). Although there was no significant difference in perceptions of the effectiveness of feedback between study sites, how individuals rated their local departmental context in terms of whether or not it was supportive of quality improvement was strongly associated with how instrumental they believed local feedback to be in improving care. This finding is interesting as it suggests that without a supportive local context, providing information on variations in care may not result in actual improvement, as support for acting on the data and implementing changes may be constrained. It is additionally compatible with prior research, which suggests that having a local operating culture that is conducive to quality improvement is an important contextual factor influencing the success of local initiatives.¹⁵¹

TABLE 20 Coefficients: model parameters for final fitted model with all variables entered

Final model with all variables entered	Unstandardised coefficients		Standardised coefficients			95% confidence interval for β	
	β	Standard error	β	t	Significant	Lower bound	Upper bound
Constant	-0.40	0.72		-0.56	0.58	-1.85	1.04
Tenure	-0.01	0.02	-0.03	-0.33	0.74	-0.05	0.03
Trust	0.01	0.34	0.001	0.01	0.99	-0.68	0.68
Departmental climate for quality improvement	0.30	0.16	0.19	1.82	0.07	-0.03	0.62
Comprehensiveness of dimensions of feedback received	0.12	0.17	0.07	0.70	0.49	-0.22	0.46
Departmental feedback	-0.52	0.45	-0.11	-1.16	0.25	-1.41	0.37
Personal feedback	-0.43	0.60	-0.07	-0.72	0.47	-1.64	0.77
The degree to which data are comprehensive	-0.19	0.22	-0.15	-0.86	0.39	-0.62	0.24
The degree to which indicators are relevant to the specific service area	0.64	0.25	0.55	2.57	0.01	0.14	1.13
The degree to which indicators are reliable and accurate	-0.05	0.24	-0.04	-0.20	0.84	-0.52	0.42
The degree to which indicators measure aspects of care that can be improved	0.02	0.20	0.02	0.08	0.94	-0.38	0.41
The degree to which data analysis is at a level which is relevant to you	-0.36	0.19	-0.31	-1.94	0.06	-0.73	0.01
The degree to which frequency of feedback helps monitor trends	0.07	0.16	0.06	0.41	0.69	-0.26	0.39
The degree to which feedback is communicated effectively	0.001	0.13	0.001	0.01	0.996	-0.26	0.27
The degree to which data presentation is adequate for effective use	0.09	0.22	0.089	0.42	0.67	-0.34	0.53
The degree to which the source of the data is credible	0.55	0.22	0.54	2.57	0.01	0.12	0.98

Hypotheses 4 and 5 investigated the role of various feedback design and content characteristics in order to understand which factors were most important to clinical end-users. Both the scope of local quality monitoring and the level of feedback (i.e. whether it was specific to performance at the departmental level or performance at the level of the individual clinician) were significant predictors of perceived usefulness (hypothesis 4). This suggests that the more dimensions of care on which an individual receives feedback (e.g. patient safety, clinical effectiveness, patient experience, compliance with best practice), the more useful information they have to interpret and from which to learn. Higher overall intensity of feedback has been shown to increase its effectiveness.^{107,144} A particularly relevant finding concerns the level of feedback provided within a health service. In our model, receiving feedback on care delivered by the individual practising clinician was a strong positive predictor of perceived usefulness. This finding reinforces the notion that personal professional feedback is important in learning and improving clinical practice.^{45,48,52,105}

The final regression model investigated the role of feedback content characteristics, which were found collectively to explain a large proportion (26.4%) of the remaining variance in perceived usefulness, once all prior factors had been controlled for. As expected, the content and design of feedback, in terms of factors such as relevance, reliability, timeliness, presentation and credibility, is clearly the most important predictor of its utility. When all content characteristics were entered into the model simultaneously, only two factors were found to have a significant unique effect on perceived usefulness of feedback. These were the relevance of the quality indicators to the specific service area and the credibility of the data as coming from a trustworthy, unbiased source. Prior research has highlighted the importance of the perceived credibility of data from quality indicators and the extent to which they originate from a trusted source.^{37,52,104,107} Qualitative findings highlight the importance of investing time to establish the credibility of performance data and of involving respected members of senior staff to achieve this in quality improvement programmes.⁹⁷ Two systematic reviews support this concept further by identifying feedback provided by experts as resulting in more effective feedback.^{135,148} The local relevance of quality indicators has additionally been highlighted as an important characteristic of effective feedback. Two systematic reviews concluded that tailoring a feedback intervention to the local setting augmented its effectiveness.^{52,55}

Results 3: longitudinal evaluation based on perceptions of data feedback

Primary site (St Mary's Hospital)

Demographics

Table 21 shows the number of participants that completed a survey at each time. The number of participants decreased across the three time points. This is to be expected for longitudinal studies of this nature. The table also demonstrates a record of the number of participants who participated repeatedly in each time point and the number of new participants at each time point. Eighty-two per cent of time point 2 participants had also participated at time point 1, and 54% of time point 3 participants had also participated at time points 1 and 2. All new participants at time points 2 and 3 had had full exposure to the intervention.

Descriptive statistics

Tables 22–26 show the descriptive statistics for each of the survey sections across the three time points. The general pattern for all sections is for both the mean and the median to increase between time points 1 and 2 and then decrease very slightly between time points 2 and 3. Simultaneously, the SD generally decreases between time points 1 and 2 and then rises slightly between time points 2 and 3. It is interesting to note that the final section on workplace climate was relatively stable across the three time points when compared with the other evaluative sections.

TABLE 21 Overview of participants at each time point

Time point	Total participants, <i>N</i>	Repeated participants, <i>n</i>	New participants, <i>n</i>	Response rate, %
1	28	N/A	N/A	82
2	22	18	4	56
3	13	7	6	40

N/A, not applicable.

TABLE 22 Items C1–4 (quality indicators)

Time point	Mean	SD	Median	Range
1	2.43	1.37	2.25	4.00
2	6.08	1.00	6.25	4.00
3	5.83	1.02	5.88	3.00

TABLE 23 Items C5–9 (feedback)

Time point	Mean	SD	Median	Range
1	2.27	1.24	2.00	4.00
2	6.50	1.10	6.60	5.00
3	6.43	1.18	6.60	4.80

TABLE 24 Items C10–15 (data use)

Time point	Mean	SD	Median	Range
1	3.06	1.96	3.17	6.17
2	5.49	1.04	5.80	4.50
3	5.12	1.25	5.33	3.50

TABLE 25 Items C1–15 (overall effectiveness)

Time point	Mean	SD	Median	Range
1	2.64	1.33	2.53	3.93
2	5.98	0.95	6.20	4.27
3	5.71	0.95	5.92	3.69

TABLE 26 Items D1–16 (workplace climate)

Time point	Mean	SD	Median	Range
1	4.53	1.23	4.31	4.75
2	5.19	0.86	5.50	3.25
3	5.08	1.39	4.88	4.87

Inferential statistics

Figure 2 depicts the mean scores from each survey subsection by study time point for St Mary's Hospital.

Items C1–4 (quality indicators)

A one-way between-subjects ANOVA was conducted to compare the effect of time point on responses to items C1–4 (quality indicators) at time point 1, time point 2 and time point 3.

There was a significant effect of time point on responses to items C1–4 (quality indicators), $F(2) = 69.33$; $p < 0.0001$.

Post-hoc comparisons using the Bonferroni test indicated that the mean score for time point 1 (mean = 2.43, SD = 1.37) was significantly different from the mean score for time point 2 (mean = 6.08, SD = 1.00) and the mean score for time point 3 (mean = 5.83, SD = 1.02). However, the mean score for time point 2 did not significantly differ from the mean score for time point 3.

Items C5–9 (feedback)

A one-way between-subjects ANOVA was conducted to compare the effect of time point on responses to items C5–9 (feedback) at time point 1, time point 2 and time point 3.

There was a significant effect of time point on responses to items C5–9 (feedback), $F(2) = 93.89$; $p < 0.0001$.

Post-hoc comparisons using the Bonferroni test indicated that the mean score for time point 1 (mean = 2.27, SD = 1.24) was significantly different from the mean score for time point 2 (mean = 6.50, SD = 1.10) and the

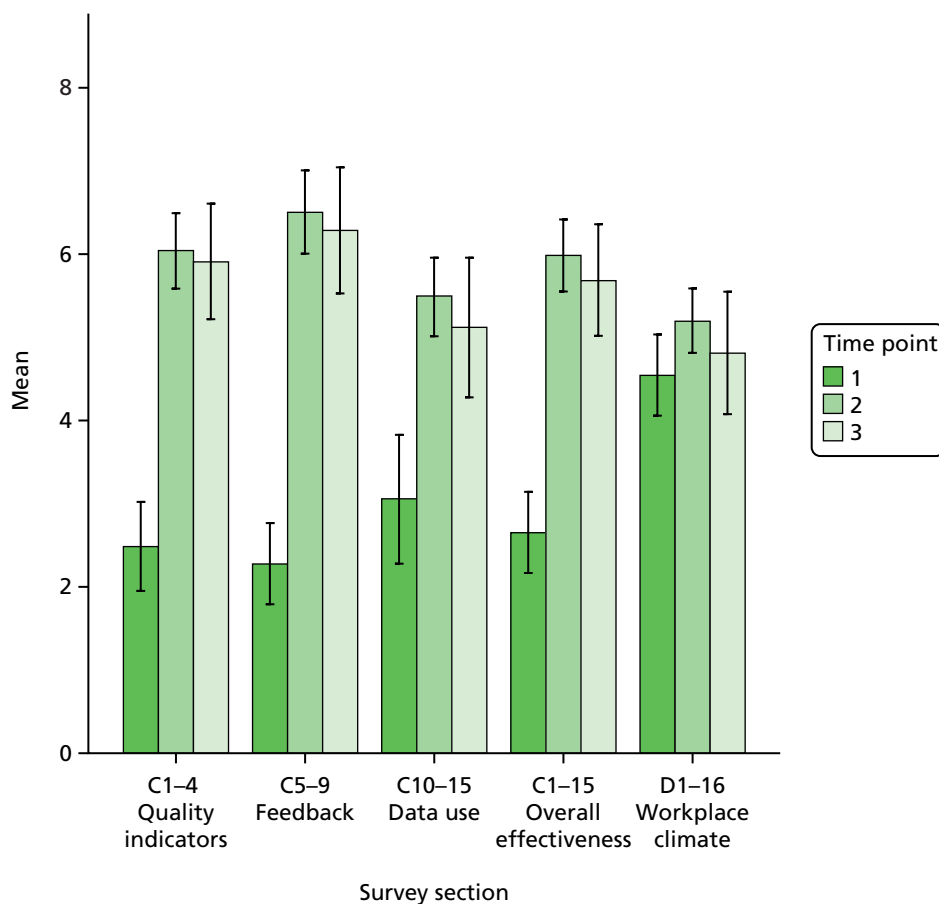


FIGURE 2 Mean scores by survey section and time point: St Mary's Hospital. Error bars: 95% confidence intervals.

mean score for time point 3 (mean = 6.43, SD = 1.18). However, the mean score for time point 2 did not significantly differ from the mean score for time point 3.

Items C10–15 (data use)

A one-way between-subjects ANOVA was conducted to compare the effect of time point on responses to items C10–15 (data use) at time point 1, time point 2 and time point 3.

There was a significant effect of time point on responses to items C10–15 (data use), $F(2) = 39.47$, $p < 0.0001$.

Post-hoc comparisons using the Bonferroni test indicated that the mean score for time point 1 (mean = 3.06, SD = 1.96) was significantly different from the mean score for time point 2 (mean = 5.49, SD = 1.04) and the mean score for time point 3 (mean = 5.12, SD = 1.25). However, the mean score for time point 2 did not significantly differ from the mean score for time point 3.

Items C1–15 (overall effectiveness)

A one-way between-subjects ANOVA was conducted to compare the effect of time point on responses to items C1–15 (overall effectiveness) at time point 1, time point 2 and time point 3.

There was a significant effect of time point on responses to items C1–15 (overall effectiveness), $F(2) = 78.34$; $p < 0.0001$.

Post-hoc comparisons using the Bonferroni test indicated that the mean score for time point 1 (mean = 2.64, SD = 1.33) was significantly different from the mean score for time point 2 (mean = 5.98, SD = 0.95) and the mean score for time point 3 (mean = 5.71, SD = 0.95). However, the mean score for time point 2 did not significantly differ from the mean score for time point 3.

Items D1–16 (workplace climate)

A one-way between-subjects ANOVA was conducted to compare the effect of time point on responses to items D1–16 (workplace climate) at time point 1, time point 2 and time point 3.

There was not a significant effect of time point on responses to items D1–16 (workplace climate), $F(2) = 2.17$; $p = 0.12$.

Charing Cross Hospital

Demographics

Table 27 shows the number of participants that completed a survey at each time point. The number of participants remained the same across the two time points. The table also demonstrates a record of the number of participants who participated repeatedly in each time point and the number of new participants at the second time point. Fifty-eight per cent of time point 2 participants had also participated at time point 1. All new participants at time point 2 had had full exposure to the intervention.

TABLE 27 Overview of participants at each time point

Time point	Total participants, <i>N</i>	Repeated participants, <i>n</i>	New participants, <i>n</i>	Response rate, %
1	12	N/A	N/A	44
2	12	7	5	48
N/A, not applicable.				

Descriptive statistics

Tables 28–32 show the descriptive statistics for each of the survey sections across the two time points. The general pattern for all sections is for both the mean and the median to increase between time points 1 and 2. Simultaneously, the SD generally decreases between time points 1 and 2. Again, it is interesting to note that the final section on workplace climate was relatively stable across the two time points when compared with the other evaluative sections.

TABLE 28 Items C1–4 (quality indicators)

Time point	Mean	SD	Median	Range
1	3.57	2.17	3.00	5.50
2	5.16	0.94	5.00	3.25

TABLE 29 Items C5–9 (feedback)

Time point	Mean	SD	Median	Range
1	2.95	1.95	2.00	5.40
2	6.11	1.06	6.60	3.40

TABLE 30 Items C10–15 (data use)

Time point	Mean	SD	Median	Range
1	3.03	2.10	2.00	5.67
2	5.24	0.93	5.33	2.90

TABLE 31 Items C1–15 (overall effectiveness)

Time point	Mean	SD	Median	Range
1	3.15	1.89	2.93	5.40
2	5.52	0.85	5.73	2.80

TABLE 32 Items D1–16 (workplace climate)

Time point	Mean	SD	Median	Range
1	4.89	1.43	4.38	4.43
2	5.41	0.77	5.44	2.73

Inferential statistics

Figure 3 depicts the mean scores from each survey subsection by study time point for Charing Cross Hospital.

Items C1–4 (quality indicators)

An independent samples *t*-test was conducted to compare mean responses to items C1–4 (quality indicators) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 3.57, SD = 2.18) and time point 2 (mean = 5.16, SD = 0.94); $t(13.58) = -2.23$; $p = 0.04$.

Items C5–9 (feedback)

An independent samples *t*-test was conducted to compare mean responses to items C5–9 (feedback) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 2.95, SD = 1.95) and time point 2 (mean = 6.11, SD = 1.06); $t(15.42) = -4.74$; $p < 0.0001$.

Items C10–15 (data use)

An independent samples *t*-test was conducted to compare mean responses to items C10–15 (data use) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 3.03, SD = 2.10) and time point 2 (mean = 5.24, SD = 0.93); $t(13.78) = -3.19$; $p = 0.01$.

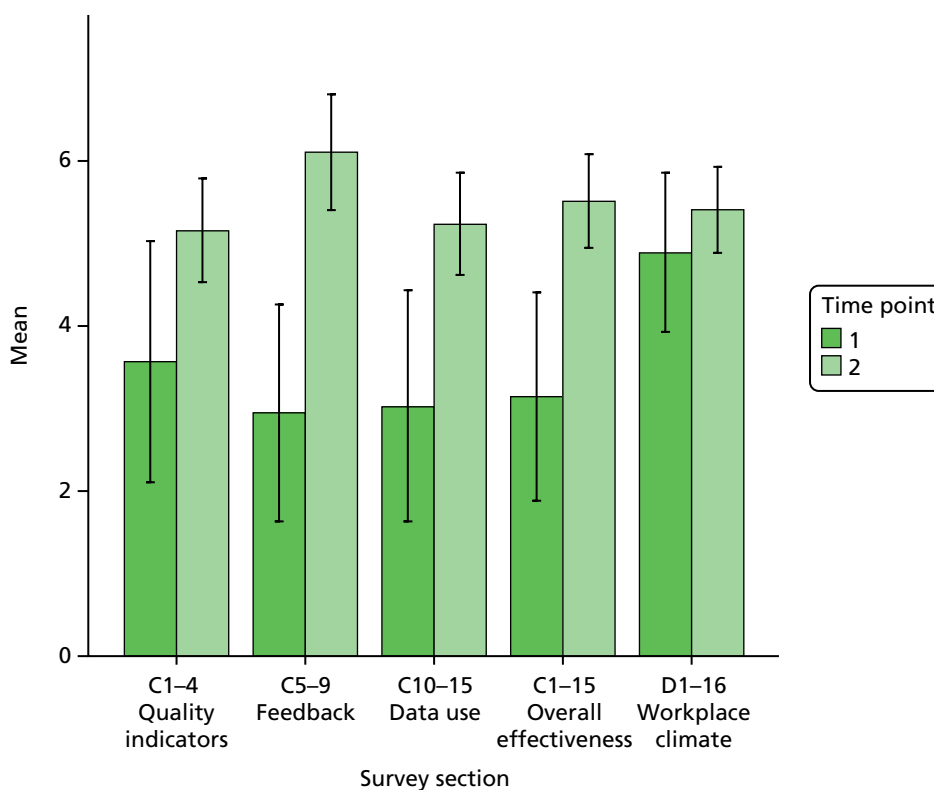


FIGURE 3 Mean scores by survey section and time point: Charing Cross Hospital. Error bars: 95% confidence intervals.

Items C1–15 (overall effectiveness)

An independent samples *t*-test was conducted to compare mean responses to items C1–15 (overall effectiveness) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 3.15, SD = 1.89) and time point 2 (mean = 5.52, SD = 0.85); $t(13.92) = -3.79$; $p = 0.002$.

Items D1–16 (workplace climate)

An independent samples *t*-test was conducted to compare mean responses to items D1–16 (workplace climate) at time point 1 and time point 2.

There was not a significant difference in the scores for time point 1 (mean = 4.89, SD = 1.43) and time point 2 (mean = 5.41, SD = 0.77); $t(15.33) = -1.06$; $p = 0.31$.

Hammersmith Hospital**Demographics**

Table 33 shows the number of participants that completed a survey at each time point. The number of participants decreased by 33% from time point 1 to time point 2. The table also demonstrates a record of the number of participants who participated repeatedly in each time point and the number of new participants at the second time point. One hundred per cent of time point 2 participants had also participated at time point 1.

TABLE 33 Overview of participants at each time point

Time point	Total participants, <i>N</i>	Repeated participants, <i>n</i>	New participants, <i>n</i>	Response rate, %
1	15	N/A	N/A	88
2	10	10	0	60
N/A, not applicable.				

Descriptive statistics

Tables 34–38 show the descriptive statistics for each of the survey sections across the two time points. The general pattern for all sections is for both the mean and the median to increase between time points 1 and 2. Simultaneously, the SD generally decreases between time points 1 and 2. Again, it is interesting to note that the final section on workplace climate was relatively stable across the two time points when compared with the other evaluative sections.

TABLE 34 Items C1–4 (quality indicators)

Time point	Mean	SD	Median	Range
1	2.60	1.79	2.00	6.25
2	3.94	1.56	4.25	4.75

TABLE 35 Items C5–9 (feedback)

Time point	Mean	SD	Median	Range
1	2.79	1.89	2.20	6.40
2	5.05	1.71	5.30	4.67

TABLE 36 Items C10–15 (data use)

Time point	Mean	SD	Median	Range
1	2.70	1.65	2.17	5.00
2	4.58	1.53	5.00	4.33

TABLE 37 Items C1–15 (overall effectiveness)

Time point	Mean	SD	Median	Range
1	2.70	1.67	2.20	5.53
2	4.58	1.41	4.70	4.27

TABLE 38 Items D1–16 (workplace climate)

Time point	Mean	SD	Median	Range
1	4.74	1.16	5.06	4.56
2	4.93	0.94	4.83	2.54

Inferential statistics

Figure 4 depicts the mean scores from each survey subsection by study time point for Hammersmith Hospital.

Items C1–4 (quality indicators)

An independent samples *t*-test was conducted to compare mean responses to items C1–4 (quality indicators) at time point 1 and time point 2.

There was not a significant difference in the scores for time point 1 (mean = 2.60, SD = 1.79) and time point 2 (mean = 3.94, SD = 1.56); $t(21) = -1.78$; $p = 0.09$.

Items C5–9 (feedback)

An independent samples *t*-test was conducted to compare mean responses to items C5–9 (feedback) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 2.79, SD = 1.89) and time point 2 (mean = 5.05, SD = 1.71); $t(21) = -2.83$; $p < 0.01$.

Items C10–15 (data use)

An independent samples *t*-test was conducted to compare mean responses to items C10–15 (data use) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 2.70, SD = 1.65) and time point 2 (mean = 4.58, SD = 1.53); $t(21) = -2.67$; $p = 0.01$.

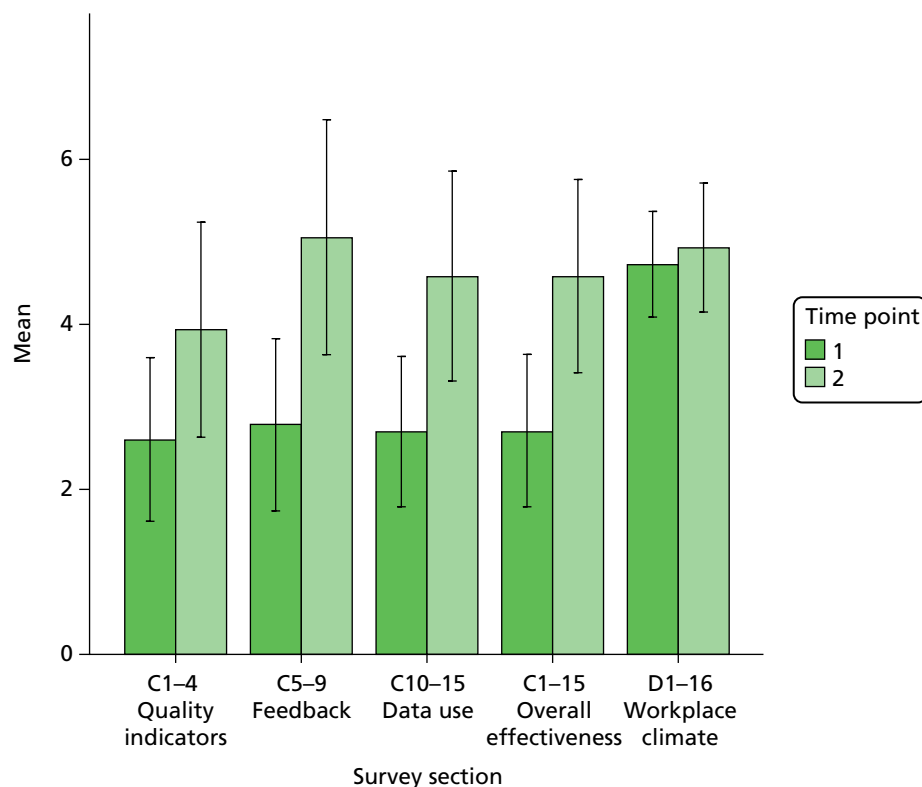


FIGURE 4 Mean scores by survey section and time point: Hammersmith Hospital. Error bars: 95% confidence intervals.

Items C1–15 (overall effectiveness)

An independent samples *t*-test was conducted to compare mean responses to items C1–15 (overall effectiveness) at time point 1 and time point 2.

There was a significant difference in the scores for time point 1 (mean = 2.70, SD = 1.67) and time point 2 (mean = 4.58, SD = 1.41); $t(21) = -2.70$; $p = 0.01$.

Items D1–16 (workplace climate)

An independent samples *t*-test was conducted to compare mean responses to items D1–16 (workplace climate) at time point 1 and time point 2.

There was not a significant difference in the scores for time point 1 (mean = 4.74, SD = 1.16) and time point 2 (mean = 4.93, SD = 0.94); $t(21) = -0.41$; $p = 0.69$.

Change in mean per item

In order to investigate what might be driving the observed effects between study time points, analysis against raw scores at the item level were conducted.

Tables 39–41 show changes in mean scores across the time points for each item. Graphs showing individual items by time point for each site are included in *Appendix 4*.

Primary site (St Mary's Hospital)

At the primary site the items that saw the greatest mean increase between time points 1 and 2 were level of analysis (+4.66), timeliness (+4.39), data credibility (+4.33) and data presentation (+4.29) (see *Table 39*). The items that saw a mean decrease between time points 1 and 2 were having a department that reviews critical incidents and serious failures to improve systems (−0.68), that openly discusses minor failures (−0.46) and that is supported by the organisation in its efforts to collect and use data (−0.23). Feeling comfortable for quality of care received by your patients to be monitored and fed back also saw a mean decrease (−0.17).

The items that saw the greatest mean increase between time points 2 and 3 were having a department that is supported by the organisation in its efforts to collect and use data (+1.22), having adequate knowledge and training in quality improvement methods (+0.87) and timeliness of feedback (+0.68). The items that saw the greatest mean decrease were data credibility (−1.14) and having a department that reviews critical incidents and serious failures to improve systems (−1.07).

The items that saw the greatest mean increase between time points 1 and 3 were timeliness (+5.07), level of analysis (+4.47), data presentation (+3.87) and relevance of quality indicators (+3.75). The items that saw the greatest mean decrease were having a department that reviews critical incidents and serious failures to improve systems (−1.75) and that openly discusses minor failures (−1.20).

Charing Cross Hospital

At additional site 1 the items that saw the greatest mean increase between time points 1 and 2 were timeliness (+3.64), means of communication (+3.36), level of analysis (+3.18) and data presentation (+3.00) (see *Table 40*). The items that saw a mean decrease were having a department that responds to variation constructively (−0.87) and that is clear about what are acceptable standards of care (−0.37).

TABLE 39 Descriptives by item and time point

Survey item	T1 mean (SD)	T2 mean (SD)	T3 mean (SD)	Change in mean T1-2	Change in mean T2-3	Change in mean T1-3
Quality indicators						
Comprehensiveness: the degree to which the data you receive are comprehensive and cover all important dimensions of care quality	2.11 (1.20)	6.00 (1.31)	5.55 (1.21)	+ 3.89	-0.45	+ 3.44
Relevance: the degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients	2.25 (1.27)	6.27 (1.35)	6.00 (0.85)	+ 4.02	-0.27	+ 3.75
Reliability: the degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of the data over time	2.54 (1.62)	5.64 (1.29)	5.33 (1.44)	+ 3.10	-0.31	+ 2.79
Improability: the degree to which indicators measure aspects of care that you and your unit can have a direct impact on through changing behaviour, the care process or local systems	2.82 (1.95)	6.41 (1.05)	6.33 (0.99)	+ 3.59	-0.08	+ 3.51
Feedback						
Level of analysis: the degree to which the data that you receive are broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre, your patients)	1.86 (1.04)	6.52 (1.08)	6.33 (1.30)	+ 4.66	-0.19	+ 4.47
Timeliness: the degree to which the frequency of feedback you receive helps you to monitor how care quality varies over time	1.85 (1.03)	6.24 (1.26)	6.92 (1.38)	+ 4.39	+ 0.68	+ 5.07
Means of communication: the degree to which the channel and method for dissemination (e.g. meetings, e-mail, reports, posters) are useful and engaging	2.81 (1.84)	6.33 (1.35)	6.10 (1.29)	+ 3.52	-0.23	+ 3.29
Data presentation: the degree to which the format in which data are presented (e.g. tables, graphs, scorecards) is clear and easy to use with the right number of data presented	2.33 (1.47)	6.62 (1.28)	6.20 (1.30)	+ 4.29	-0.42	+ 3.87
Data credibility: the degree to which the data are viewed as credible and from a trustworthy, unbiased source	2.48 (1.83)	6.81 (1.47)	5.67 (1.41)	+ 4.33	-1.14	+ 3.19

continued

TABLE 39 Descriptives by item and time point (continued)

Survey item	T1 mean (SD)	T2 mean (SD)	T3 mean (SD)	Change in mean T1-2	Change in mean T2-3	Change in mean T1-3
Data use						
Identifying problem areas and good practice: the degree to which data feedback helps us to rapidly detect problems and identify instances of excellent care	3.00 (2.06)	5.40 (1.31)	5.00 (1.05)	+2.40	-0.40	+2.00
Benchmarking: the degree to which the data feedback allows us to compare ourselves against similar units and/or national guidelines in a meaningful way	2.89 (1.93)	4.80 (1.67)	5.10 (1.37)	+1.91	+0.30	+2.21
Prioritising action: the degree to which data feedback supports prioritising where we put our efforts to improve care and which specific processes to focus on	3.15 (2.23)	5.24 (1.45)	5.20 (1.23)	+2.09	-0.04	+2.05
Setting measurable objectives: the degree to which data feedback supports setting quantifiable targets for improvement	3.26 (2.01)	5.76 (0.89)	4.80 (1.55)	+2.50	-0.96	+1.54
Monitoring progress: the degree to which data feedback supports evaluation of our progress towards targets over time and whether or not any gains are sustained	3.00 (2.11)	5.67 (1.24)	4.70 (1.64)	+2.67	-0.97	+1.70
Overall usefulness for improvement: the degree to which current data feedback is useful in monitoring variations and improving care	3.04 (2.18)	6.05 (1.24)	5.50 (1.43)	+3.01	-0.55	+2.46
Workplace climate						
In this department we are proactive in striving to continuously improve standards of care	5.15 (1.46)	5.67 (1.49)	5.67 (1.83)	+0.52	0	+0.52
In this department we routinely review data on quality-of-care outcomes	4.00 (1.94)	5.62 (1.63)	5.27 (1.19)	+1.62	-0.35	+1.27
In this department it is clear as to what are acceptable standards of care	4.74 (1.95)	5.48 (1.37)	5.00 (1.41)	+0.74	-0.48	+0.26
In this department we monitor compliance with best practice guidelines for clinical care	3.89 (1.97)	5.10 (1.64)	4.58 (1.88)	+1.21	-0.52	+0.69
In this department we respond to variation constructively, to improve care, rather than blaming and punishing individuals	4.63 (2.32)	5.14 (1.77)	4.17 (2.04)	+0.51	-0.97	-0.46

Survey item	T1 mean (SD)	T2 mean (SD)	T3 mean (SD)	Change in mean T1-2	Change in mean T2-3	Change in mean T1-3
In this department we openly discuss minor failures in care to learn lessons	5.70 (1.77)	5.24 (1.97)	4.50 (1.78)	-0.46	-0.74	-1.20
In this department we review critical incidents and serious failures to improve systems	6.25 (1.56)	5.57 (1.78)	4.50 (1.78)	-0.68	-1.07	-1.75
In this department we can demonstrate to senior levels of the organisation the quality of care we are delivering	3.96 (1.99)	6.10 (1.18)	5.36 (1.57)	+2.14	-0.74	+1.40
In this department we are supported by the organisation in our efforts to collect and use data	3.37 (2.22)	3.14 (1.49)	4.36 (1.57)	-0.23	+1.22	+0.99
I have adequate knowledge and training on the statistics required to interpret quality-of-care data	4.07 (2.29)	5.24 (1.34)	5.83 (1.90)	+1.17	+0.59	+1.76
I have adequate knowledge and training in quality improvement methods	4.00 (2.06)	4.71 (1.59)	5.58 (1.93)	+0.71	+0.87	+1.58
In this department variations in the quality of care delivered to patients often go undetected	3.07 (1.69)	3.70 (1.30)	4.09 (1.58)	+0.63	+0.39	+1.02
In this department it is clear who is responsible for taking action to make changes and improve care processes	4.11 (1.93)	4.48 (1.50)	4.00 (2.21)	+0.37	-0.48	-0.11
In this department we use personal data on the quality of care that individuals deliver in a constructive way	3.37 (2.27)	5.33 (1.59)	4.91 (1.81)	+1.96	-0.42	+1.54
I am comfortable for the quality of care received by my patients to be monitored and fed back to me	7.41 (1.25)	7.24 (0.83)	6.75 (1.36)	-0.17	-0.49	-0.66
In this department we are effective in taking action based on audit and data feedback	4.67 (1.86)	5.24 (1.30)	5.00 (1.67)	+0.57	-0.24	+0.33
T, time point.						

TABLE 40 Descriptives by item and time point

Survey item	T1 mean (SD)	T2 mean (SD)	Change in mean T1–2 (SD)
Quality indicators			
Comprehensiveness: the degree to which the data you receive are comprehensive and cover all important dimensions of care quality	3.36 (2.11)	5.27 (0.91)	+1.91
Relevance: the degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients	3.55 (2.30)	5.09 (1.22)	+1.54
Reliability: the degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of the data over time	3.27 (2.15)	4.45 (1.51)	+1.18
Improvability: the degree to which indicators measure aspects of care that you and your unit can have a direct impact on through changing behaviour, the care process or local systems	4.09 (2.43)	5.82 (1.40)	+1.73
Feedback			
Level of analysis: the degree to which the data that you receive are broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre, your patients)	3.09 (2.21)	6.27 (1.42)	+3.18
Timeliness: the degree to which the frequency of feedback you receive helps you to monitor how care quality varies over time	3.09 (2.30)	6.73 (1.35)	+3.64
Means of communication: the degree to which the channel and method for dissemination (e.g. meetings, e-mail, reports, posters) are useful and engaging	2.73 (1.62)	6.09 (1.22)	+3.36
Data presentation: the degree to which the format in which data are presented (e.g. tables, graphs, scorecards) is clear and easy to use with the right number of data presented	2.91 (2.12)	5.91 (1.22)	+3.00
Data credibility: the degree to which the data are viewed as credible and from a trustworthy, unbiased source	2.91 (2.55)	5.55 (1.37)	+2.64
Data use			
Identifying problem areas and good practice: the degree to which data feedback helps us to rapidly detect problems and identify instances of excellent care	2.91 (2.02)	5.64 (0.81)	+2.73
Benchmarking: the degree to which the data feedback allows us to compare ourselves against similar units and/or national guidelines in a meaningful way	3.00 (2.28)	4.91 (1.76)	+1.91
Prioritising action: the degree to which data feedback supports prioritising where we put our efforts to improve care and which specific processes to focus on	3.00 (2.32)	4.55 (1.44)	+1.55
Setting measurable objectives: the degree to which data feedback supports setting quantifiable targets for improvement	3.09 (2.34)	5.09 (0.94)	+2.00
Monitoring progress: the degree to which data feedback supports evaluation of our progress towards targets over time and whether or not any gains are sustained	3.09 (2.26)	5.70 (0.82)	+2.61
Overall usefulness for improvement: the degree to which current data feedback is useful in monitoring variations and improving care	3.09 (2.26)	6.00 (1.32)	+2.91

TABLE 40 Descriptives by item and time point (continued)

Survey item	T1 mean (SD)	T2 mean (SD)	Change in mean T1–2 (SD)
Workplace climate			
In this department we are proactive in striving to continuously improve standards of care	5.27 (1.95)	5.91 (1.30)	+0.64
In this department we routinely review data on quality-of-care outcomes	4.55 (1.81)	5.27 (1.49)	+0.72
In this department it is clear as to what are acceptable standards of care	5.64 (1.57)	5.27 (1.49)	–0.37
In this department we monitor compliance with best practice guidelines for clinical care	4.27 (2.05)	4.91 (1.58)	+0.64
In this department we respond to variation constructively, to improve care, rather than blaming and punishing individuals	5.60 (1.84)	4.73 (1.79)	–0.87
In this department we openly discuss minor failures in care to learn lessons	6.36 (1.21)	6.70 (0.95)	+0.34
In this department we review critical incidents and serious failures to improve systems	6.70 (1.34)	7.27 (0.79)	+0.57
In this department we can demonstrate to senior levels of the organisation the quality of care we are delivering	4.27 (2.05)	5.55 (1.04)	+1.28
In this department we are supported by the organisation in our efforts to collect and use data	3.55 (2.54)	5.18 (1.47)	+1.63
I have adequate knowledge and training on the statistics required to interpret quality-of-care data	4.55 (2.51)	5.27 (1.56)	+0.72
I have adequate knowledge and training in quality improvement methods	4.45 (2.16)	5.10 (2.13)	+0.65
In this department variations in the quality of care delivered to patients often go undetected	2.82 (1.94)	3.40 (1.58)	+0.58
In this department it is clear who is responsible for taking action to make changes and improve care processes	4.27 (2.05)	4.55 (2.21)	+0.28
In this department we use personal data on the quality of care that individuals deliver in a constructive way	4.73 (2.33)	4.73 (1.27)	0
I am comfortable for the quality of care received by my patients to be monitored and fed back to me	7.09 (2.07)	7.18 (1.17)	+0.09
In this department we are effective in taking action based on audit and data feedback	4.80 (1.87)	5.36 (0.92)	+0.56
T, time point.			

Hammersmith Hospital

At additional site 2 the items that saw the greatest mean increase between time points 1 and 2 were data presentation (+3.23), overall usefulness for improvement (+3.00) and data credibility (+2.96) (see *Table 41*). The items that saw the greatest mean decrease were feeling comfortable for quality of care received by your patients to be monitored and fed back (-2.67) and having a department that reviews critical incidents and serious failures to improve systems (-1.96).

TABLE 41 Descriptives by item and time point

Survey item	T1 mean (SD)	T2 mean (SD)	Change in mean T1-2 (SD)
Quality indicators			
Comprehensiveness: the degree to which the data you receive are comprehensive and cover all important dimensions of care quality	2.33 (1.59)	4.29 (1.25)	+1.96
Relevance: the degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients	2.47 (1.89)	3.86 (1.46)	+1.39
Reliability: the degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of the data over time	2.60 (1.99)	4.14 (1.95)	+1.54
Improvability: the degree to which indicators measure aspects of care that you and your unit can have a direct impact on through changing behaviour, the care process or local systems	3.00 (2.14)	4.63 (2.07)	+1.63
Feedback			
Level of analysis: the degree to which the data that you receive are broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre, your patients)	2.93 (2.02)	3.75 (1.83)	+0.82
Timeliness: the degree to which the frequency of feedback you receive helps you to monitor how care quality varies over time	3.13 (2.03)	5.88 (1.64)	+2.75
Means of communication: the degree to which the channel and method for dissemination (e.g. meetings, e-mail, reports, posters) are useful and engaging	2.93 (2.15)	5.25 (2.05)	+2.32
Data presentation: the degree to which the format in which data are presented (e.g. tables, graphs, scorecards) is clear and easy to use with the right number of data presented	2.60 (2.13)	5.83 (1.60)	+3.23
Data credibility: the degree to which the data are viewed as credible and from a trustworthy, unbiased source	2.33 (2.19)	5.29 (1.70)	+2.96
Data use			
Identifying problem areas and good practice: the degree to which data feedback helps us to rapidly detect problems and identify instances of excellent care	2.80 (2.08)	4.38 (1.85)	+1.58
Benchmarking: the degree to which the data feedback allows us to compare ourselves against similar units and/or national guidelines in a meaningful way	2.53 (1.89)	4.57 (1.40)	+2.04
Prioritising action: the degree to which data feedback supports prioritising where we put our efforts to improve care and which specific processes to focus on	2.73 (2.09)	5.43 (1.81)	+2.70
Setting measurable objectives: the degree to which data feedback supports setting quantifiable targets for improvement	2.93 (2.12)	5.39 (0.95)	+2.46
Monitoring progress: the degree to which data feedback supports evaluation of our progress towards targets over time and whether or not any gains are sustained	3.07 (1.67)	5.00 (1.29)	+1.93
Overall usefulness for improvement: the degree to which current data feedback is useful in monitoring variations and improving care	2.13 (1.77)	5.13 (2.32)	+3.00

TABLE 41 Descriptives by item and time point (continued)

Survey item	T1 mean (SD)	T2 mean (SD)	Change in mean T1–2 (SD)
Workplace climate			
In this department we are proactive in striving to continuously improve standards of care	6.20 (1.74)	4.88 (1.73)	–1.32
In this department we routinely review data on quality-of-care outcomes	4.13 (2.30)	3.71 (1.80)	–0.42
In this department it is clear as to what are acceptable standards of care	5.20 (2.21)	4.50 (1.93)	–0.70
In this department we monitor compliance with best practice guidelines for clinical care	4.07 (2.22)	5.25 (1.91)	+1.18
In this department we respond to variation constructively, to improve care, rather than blaming and punishing individuals	5.67 (1.92)	6.25 (1.98)	+0.58
In this department we openly discuss minor failures in care to learn lessons	6.27 (2.28)	6.13 (2.23)	–0.14
In this department we review critical incidents and serious failures to improve systems	6.53 (1.46)	4.57 (1.62)	–1.96
In this department we can demonstrate to senior levels of the organisation the quality of care we are delivering	4.33 (2.32)	2.86 (1.46)	–1.47
In this department we are supported by the organisation in our efforts to collect and use data	2.40 (1.45)	4.63 (1.77)	+2.23
I have adequate knowledge and training on the statistics required to interpret quality-of-care data	3.40 (1.55)	4.50 (1.60)	+1.10
I have adequate knowledge and training in quality improvement methods	3.60 (1.77)	4.13 (1.36)	+0.53
In this department variations in the quality of care delivered to patients often go undetected	4.40 (2.23)	4.50 (1.60)	+0.10
In this department it is clear who is responsible for taking action to make changes and improve care processes	3.73 (2.02)	4.13 (1.36)	+0.40
In this department we use personal data on the quality of care that individuals deliver in a constructive way	3.80 (2.11)	4.50 (2.07)	+0.70
I am comfortable for the quality of care received by my patients to be monitored and fed back to me	6.67 (1.84)	4.00 (1.83)	–2.67
In this department we are effective in taking action based on audit and data feedback	5.40 (1.72)	7.63 (0.52)	+2.23
T, time point.			

Discussion of longitudinal findings

This longitudinal analysis demonstrates the various influences of no feedback, basic feedback and enhanced feedback on perceptions of local quality monitoring and workplace climate.

At the primary site there was a significant improvement in perceptions of quality indicators, feedback, data use and overall effectiveness of quality monitoring from time point 1 when no feedback was being received to time point 2 when basic feedback was being received. However, there was actually a decrease in perceptions (although this did not reach statistical significance) between time point 2 when basic feedback was being received and time point 3 when enhanced feedback was being received. For additional site 1 there was a significant improvement in perceptions of quality indicators, feedback, data use and overall effectiveness from time point 1 when no feedback was being received to time point 2 when enhanced feedback was being received. Finally, for additional site 2 there was a significant improvement in perceptions of feedback, data use and overall effectiveness from time point 1 when no feedback was being received to time point 2 when enhanced feedback was being received. The improvement in perceptions of quality indicators did not reach statistical significance.

It appears that the novelty of receiving any type of feedback after receiving none at all is important here. This could explain why we saw similar effects between anaesthetists at the primary site going from no feedback to basic feedback and anaesthetists at the additional sites going from no feedback to enhanced feedback. Receiving knowledge of performance for the first time may have been particularly stimulating for clinicians resulting in a stronger cognitive and affective reaction. Implementation factors associated with quality improvement interventions have been shown to be particularly powerful.⁷⁶

The reasons for the decrease in perceptions at the primary site between basic and enhanced feedback may be more complex. It may be the case that anaesthetists developed higher expectations from the feedback as the project developed and, therefore, rated it more critically when completing the survey for the third time. Their involvement in the project, combined with the ever-growing focus on quality monitoring in the NHS, may have influenced the way that the anaesthetists conceptualised and understood effective feedback. This explanation is supported by the fact that new participants at each time point had been exposed to the initiative in its entirety. On the other hand these findings could also be indicative of an attenuation effect experienced by those who were repeatedly participating in research activity associated with this project. Too much focus on a project and its potential for 'transformation' can pose negative effects for quality improvement.⁹⁶

The intervention did not influence perceptions of local workplace climate in the same way that it did with the evaluative items. For all three sites there were no significant differences between perceptions of workplace climate across the time points. Small increases were noted between no feedback and basic feedback at the primary site, and no feedback and enhanced feedback at the additional sites. There was a decrease in perceptions of workplace climate between basic feedback and enhanced feedback at the primary site.

This finding comments on the mechanisms through which the initiative may have been working. It seems to be the case that the feedback intervention worked through providing individuals with accurate and usable performance data to work with, and helping them to identify the benefits of this for them as a health-care professional. Prior literature demonstrates the importance of key characteristics such as timeliness, feedback format, level of feedback, frequency and data quality.^{52,97,98,107,149} The provision of feedback does not seem to have worked through the stimulation of local climate for quality improvement across the department in the case of our intervention.

Analysis against raw scores revealed which specific items of the survey were most sensitive to the intervention and, therefore, driving the observed effects across study time points. Items associated with making the data more personal to the individual seemed to be most sensitive to improvement across the time points. Perceptions of data presentation was the only item to have notable mean increases across time points 1 and 2 for all sites and time point 1 and 3 for the primary site. Level of analysis and timeliness saw notable mean increases at both the primary site and additional site 1.

The items that saw the greatest decreases were all associated with workplace climate. Having a department that reviews critical incidents and serious failures to improve systems decreased across all time points at the primary site and between time points 1 and 2 at additional site 2. Feeling comfortable for quality of care received by your patients to be monitored and fed back decreased between time points 1 and 2 at both the primary site and additional site 2. Having a department that openly discusses minor failures decreased across all three time points at the primary site. Effectiveness has been associated with having a culture that promotes non-punitive discussion of local practice and quality improvement.⁹⁷

It seems that the intervention may have benefited from more time and resources devoted to improving the local workplace climate for quality improvement. The extent to which anaesthetists were able to engage with feedback in an optimal way may have been restricted by the culture within which they were working. Studies have demonstrated that organisational/departmental culture can be a barrier to effective quality improvement.^{96,97} However, it should be noted that some of the workplace climate items appeared to be more sensitive to positive effects of the initiative than others. For example, at both the primary site and additional site 1, mean increases were noted for having a department that is able to demonstrate to senior levels of the organisation the quality of care that they are delivering. The intervention clearly made individuals feel more equipped as a department to evidence the effectiveness of their professional practice. This is particularly interesting in light of the current focus on clinician revalidation. In providing clinicians with performance data the initiative was supporting them in the process of demonstrating their fitness to practice on a wider level. This can be seen as a form of reward to the individuals for participating in the project.

At the primary site and additional site 2, the intervention had a positive effect on perceptions of having adequate knowledge and training on the statistics required to interpret quality-of-care data. In this sense, receiving feedback may have had a secondary effect on an individual's ability to extract meaningful information from raw data. Effective feedback has been shown to influence recipient behaviour and decision-making.¹⁷³ Having a strong understanding of associated methodology has been demonstrated as a key characteristic of successful quality improvement interventions.¹⁷⁴

At both additional sites, mean increases were also noted for having a department that is supported by the organisation in its efforts to collect and use data. This could be an effect of the initiative being up and running for a relatively long period of time without any organisational level barriers. Clinicians experience many quality improvement initiatives being introduced and disregarded, and the consistency of the feedback intervention may have indicated support from the organisation at a higher level. Organisational readiness has been shown to be of importance to the success of quality improvement interventions.¹⁶⁶

Conclusions from the survey study

This chapter demonstrates that it is possible to develop a valid and reliable survey to capture end-user perceptions on local quality monitoring and workplace climate. Psychometric analysis supports the validity and reliability of the survey construct. It shows that it is possible to measure multidimensional concepts such as perceptions on performance feedback using multiscale items. Exploratory factor analysis suggests that workplace climate for effective use of feedback is also multidimensional, comprising open culture for proactive data use, data interpretation and knowledge for measurement-driven improvement. This

highlights potential scope for future modification of the survey structure, reflecting these newly identified subdimensions. This was not followed up as part of the current study owing to a reliance on having the same survey items in order to effectively fulfil the plans for longitudinal data collection and analysis.

The existing survey has been used to identify which characteristics of feedback predict end-user perceptions of usefulness. Given the current enthusiasm for using quality monitoring and improvement to drive improvements in practice, there is surprisingly little evidence to inform the development of effective feedback from quality indicators. The findings from this study suggest that anaesthetists perceive a range of factors as important in determining the usefulness of feedback. Specifically, the local departmental context and its support of quality improvement is an important determinant of how instrumental feedback from monitoring quality indicators is likely to be. Furthermore, feedback that is tailored to be relevant to the personal professional practice of the individual clinician is an important predictor of usefulness. In terms of the feedback content and design characteristics that anaesthetists value most, the perceived credibility of the data and the local relevance of the quality indicators are paramount.

Use of the survey to evaluate the effects of basic and enhanced feedback on perceptions of local quality monitoring and workplace climate revealed a stronger influence on individual interaction with the data than on local culture to support quality improvement. The fact that perceptions of workplace climate did not significantly improve as a result of our initiative may indicate that usability was not optimal. This is in the light of regression findings that local departmental context and its support of quality improvement is an important determinant of how instrumental feedback from monitoring quality indicators is likely to be. Targeting the local context may increase the success of this type of quality improvement initiative. Further work is needed to evaluate processes and identify the mechanisms by which feedback may have an effect, particularly the interactions with local departmental climate. Our findings should thus serve as a basis for further research which might, for example, clarify the potentially reciprocal relationship between the development of quality monitoring processes and local departmental climate for open and effective use of data to improve care. Does a more supportive local climate lead to enhanced feedback or vice versa? Our study certainly suggests that where feedback is limited, the local climate is particularly important in determining its usefulness.

Although the findings from this chapter add to the limited evidence base for effective feedback from quality monitoring initiatives, there are some limitations with the study designs. The baseline sample of respondents was relatively small and based on two acute health-care organisations only. However, these organisations were both large teaching hospitals and the overall response rate was good. Longitudinal data presented a small sample size at time point 3 for the primary site and across both time points for the two additional sites. A decline in response rate is expected, however, when conducting a longitudinal study using the same target respondents at each time point.

Survey and other forms of self-report measures are open to a number of respondent biases. For example, there is potential for participants to demonstrate demand characteristics when they are taking part in a pre and post survey (i.e. they may presume that the researchers are expecting an improvement and therefore give socially desirable answers). Subjectivity might be considered a strength, however, where the aim is to extract an 'end-user' or 'stakeholder' perspective on usability. We chose not to modify the survey based on the findings of the factor analysis, which may have influenced its overall effectiveness.

It is noteworthy that the mean scores against survey items were generally low at baseline, indicating that participants were receiving little systematic feedback. This might have challenged them when they were asked to rate effectiveness characteristics. Although we have described the available prior theory related to this area of research and hence justify the form and causal priority of the variables in our regression model, the lack of previous data imposes limits on interpretation. The statistical approach we adopted, however, was chosen to provide a preliminary investigation of the factors underpinning perceptions of effective quality monitoring and feedback in perioperative care, and we believe that the resulting findings are both intuitively plausible and compatible with the prior theory that does exist.

Finally, the large number of predictors within the final iteration of the regression model may give rise to concerns regarding overfitting. We addressed this issue through a number of measures undertaken in our design. First, the primary purpose of our analysis was to test specific a priori hypotheses linked to theory, rather than to find the most parsimonious predictive model from applying a large number of variables in a single analytic step. This was achieved through hierarchical entry of variables in sets (sometimes known as 'blocked regression'), and our primary test statistic was the additional proportion of variance explained by a set of predictors, corresponding to a specific hypothesis, rather than the beta coefficient of any single predictor in the model. While we have reported the most important predictors from the final model, the majority of our analysis is based on the aggregated effect of a small number of groups of variables at different iterations of the model. Finally, before reporting the strongest predictors within the final iteration of the regression model, we ran statistical regression in which variables were automatically excluded based on statistical criteria in order to identify the most parsimonious predictive model. The results from statistical regression should be interpreted with great caution as statistical criteria drives model specification rather than prior theory, but we employed this step as a check on the validity of the main predictors that emerged from our final predictive model, which was supported. While these steps should enhance the generalisability of the predictive results from the regression model, as with any investigative work in a novel area, further replication of the study within additional contexts would strengthen the conclusions drawn.

Chapter 8 Quasi-experimental evaluation of effect on anaesthetic quality and perioperative outcome indicators

Introduction

This section of the report details the interrupted time series (ITS) study that represented the main quasi-experimental component of the work. *Box 4* includes a summary of the research aims from the original protocol, indicating in bold which of these will be addressed by the current section of the research.

The central causal theory underlying the anaesthetics quality indicator programme under investigation in this project, and hence the core hypothesis for research and evaluation, is the premise that performance feedback in the form of data from anaesthetic quality indicators may stimulate improvements in both professional practice and clinical care. It is important to conceptualise and make explicit the hypothesised causal mechanism presumed to underpin the effects of any clinical practice or quality improvement intervention. In this case, the hypothesised causal mechanism draws on a broad range of prior research and theory in the areas of psychological learning theory,¹⁷⁵ professional behaviour change,⁶³ the effects of performance feedback in health care, in the form of both audit and feedback,^{54,176} and feedback from medical registries,⁵² along with evolving theory and practical knowledge in improvement science concerning the role of measurement and evaluation in quality improvement.^{58,68} In addition to these areas

BOX 4 Primary and secondary aims from original protocol

Research aims

To evaluate the impact of a departmental continuous quality monitoring and multilevel feedback initiative on the quality of anaesthetic care and efficiency of perioperative workflow within a London teaching hospital over a 2-year period. Data will be analysed at multiple time points over the course of the project in order to generate both formative and summative information to support development and evaluation of the initiative.

To employ a quasi-experimental time series design to provide robust evidence concerning the impact of a serial data feedback intervention on anaesthetic quality indicators while controlling for baseline variance.

To contribute to the evidence base for valid and reliable quality indicators for anaesthetic care including effective patient experience measures.

To document the main features of the data feedback intervention as it develops through the CLAHRC programme for replication at other sites, including definition of metrics, data processes, feedback format and action mechanisms.

To assess the perceived acceptability and utility of this information system for individual end-users, the clinical department and other organisational stakeholders, using a formative, mixed-methods design.

of prior research and theory, described in more detail elsewhere in this report, and supporting data feedback as a viable quality improvement intervention, we note that the content of the intervention as conceived in this study fulfils a more simplistic informational need, to provide a means of visualising variation in perioperative parameters that have not hitherto been the routine focus of quality monitoring initiatives. We note that considerable recent policy and specialty interest has emphasised the need for both development of quality indicators that are specifically relevant to anaesthetics practice (e.g. postoperative pain, nausea and functional recovery) and that are capable of capturing the patient's experience of perioperative care, most notably patient satisfaction with anaesthesia.^{17,137,139}

One further element of scientific rationale serves as introduction to the mode of analysis and evaluation employed in this section of the report, in addition to the nature of the intervention being tested. There is a broad-reaching literature on 'complex interventions' in health services research, often invoked to account for the complex effects of policy-level interventions, organisational change, large-scale national programmes and more localised, but potentially organic, quality improvement initiatives within clinical microsystems or other subunits of the health-care system.^{118,177,178} In the field of quality improvement, which perhaps finds most resonance with the intervention implemented in this project, a number of features have been identified as giving rise to complexity, including the presence of social in addition to technical intervention components, the serial and cumulative effects over time of iterative improvement programmes, and the presence of interactions with context in all its forms (e.g. cultural, organisational).^{92,179-181} While we have implemented a mixed-methods evaluative design to account for causal complexity and sociotechnical processes in the intervention under investigation, the evolving, iterative nature of the departmental quality monitoring and data feedback initiative that we have studied necessitates further consideration in the analysis of its effects on the quantitative variables and perioperative indicators that have been observed. Specifically, our evaluative design must account for the multiple, serial nature of the intervention as successive iterations of the measurement and feedback component were developed and implemented, coupled with escalating engagement, active discussion and dissemination within the perioperative service, that ultimately spread to all hospital sites and anaesthetists within the institution by the initiative's third year. We therefore selected a longitudinal, quasi-experimental design for our evaluation, ITSA, as recommended by the methodological literature on evaluation of quality improvement initiatives.^{182,183} This design is both statistically robust and a powerful form of quasi-experiment, deriving its rigour from the ability to isolate and control for temporal effects on an outcome.¹⁸⁴ In ITSA, the statistical model of effect parameters implemented represents a true longitudinal perspective, sensitive to the time series ordering of data points and baseline trends within a data set, rather than merely aggregating data at a number of discrete temporal study epochs for comparisons of observed levels of a dependent variable. We are, therefore, able to model the cumulative and serial effects of a phased or escalating intervention as a series of changes in slope and level, while controlling for prior trends. Few statistical methods are as useful for evaluating effects in time-series data, with the possible exception of some forms of statistical process control charts developed in non-health-care industries.

Methods

Research design

Evaluation of the impact of the anaesthetics quality monitoring and feedback initiative on anaesthetic quality indicators and perioperative outcomes utilised a single-group longitudinal design, with multiple study epochs. ITSA was used as the primary evaluative model, with interrupts representing multiple intervention time points. ITSA was chosen in order to model changes in time-series level and slope associated with implementation and escalation of the intervention, representing the stepwise impact of the intervention and effect on rate of change in a parameter over time, whilst controlling for prior temporal trends.

Research aims

In order to fulfil the primary evaluative aims of the project, the following main hypotheses in *Box 5* were formulated to investigate the role of feedback at different levels of intensity on perioperative process and outcome indicators.

Table 42 provides further details of how the two main hypotheses relate to specific areas of anaesthetics practice and relevant perioperative process and outcome indicators for evaluation. In addition to the specific and cumulative effects of the implementation of basic feedback (hypothesis 1) and enhanced feedback (hypothesis 2), the distinction is made between hypothesised primary effects of the initiative on perioperative process and patient-reported QoR indicators, and secondary effects of the initiative on downstream perioperative outcome indicators such as surgical site infection (SSI) and 30-day postoperative mortality.

Further to the main evaluative hypotheses concerning the overall effect of feedback as an intervention within the cohort of anaesthetists participating in the study, a third related hypothesis was investigated relating to the impact of feedback on anaesthetists with comparatively low scores on specific quality indicators at baseline. Provision of performance feedback to improve professional practice is likely to benefit those individuals with low comparative rankings on specified performance measures most, relative to the distribution of performance scores within a comparable cohort or peer group. 'Low'-performing groups may experience stronger gains from the same intensity of feedback than their comparatively 'high'-performance colleagues, who possess less 'room for improvement', or experience diminished returns from their improvement efforts. There may be ceiling effects limiting possible variation in the measurement scales used to assess performance or real limitations in terms of simply how much additional variance is possible (e.g. patient core temperature cannot be increased indefinitely, nor can the total absence of postoperative nausea or pain be improved on). The effects of the feedback intervention should, therefore, be strongest in the low-performance group. In order to investigate how the lower-ranked anaesthetists responded to the feedback intervention, a third hypothesis may be specified (*Box 6*).

BOX 5 Research hypotheses formulated to investigate the role of feedback at different levels of intensity on perioperative process and outcome indicators

H1: Implementation of routine departmental anaesthetic quality monitoring and basic feedback at the level of the individual consultant anaesthetist will improve professional practice and improve anaesthetic care as evidenced by positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality.

H2: Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve professional practice and improve anaesthetic care as evidenced by positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality.

H, hypothesis; SSI, surgical site infection.

TABLE 42 Table linking anaesthetic/perioperative practice area with main evaluative hypotheses, causal mechanisms and outcomes

Anaesthetic/perioperative practice area	Causal mechanism linked to implementation of data feedback (H1)	Causal mechanism linked to implementation of enhanced feedback (H2)	Hypothesised effects of feedback	Dependent measures
Perioperative normothermia: the regulation of patient core temperature during the perioperative pathway and intraoperative phase to prevent inadvertent perioperative hypothermia which may result in increased risk of blood loss, wound infections and cardiac problems	Feedback of data concerning variation in the performance of perioperative normothermia processes will cause improved compliance with best practice guidelines in this area (NICE guidelines), including enhanced monitoring and compensatory actions	Provision of enhanced feedback including detailed case category breakdown, speciality-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination will facilitate personal professional learning from case experience and motivation to improve or maintain high standards of practice	Primary effect: implementation of feedback and escalation of feedback intensity will result in patients arriving in recovery with warmer core temperature overall and there will be fewer instances of patients experiencing inadvertent perioperative hypothermia Secondary effect: improvements in perioperative normothermia regime compliance will reduce the instance of SSI and 30-day surgical mortality, through reduced suppression of the immune system and decreased risk of postoperative complications	<ol style="list-style-type: none"> 1. Average patient temperature on arrival in recovery 2. Proportion of patients arriving in recovery with temperature < 36°C 3. Postoperative QoR score 4. SSI rate 5. 30-day surgical mortality
Postoperative pain management: the implementation of appropriate analgesic therapy that minimises patient discomfort, patient dissatisfaction and adverse side effects, while enhancing the possibility of rapid postoperative mobilisation and recovery	Feedback of data concerning variation in patient-reported pain will cause improved analgesic therapy including enhanced monitoring and compensatory actions	Provision of enhanced feedback including detailed case category breakdown, speciality-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination will facilitate personal professional learning from case experience and motivation to improve or maintain high standards of practice	Primary effect: implementation of feedback and escalation of feedback intensity will result in patients reporting enhanced freedom from moderate and severe pain during the recovery period and an enhanced overall QoR experience Secondary effect: improvements in perioperative analgesia will reduce the instance of 30-day surgical mortality, through reduced risk of postoperative complications	<ol style="list-style-type: none"> 1. Pain on admission (first conscious pain score) 2. Patient-reported freedom from severe or constant moderate pain during the recovery period 3. Postoperative QoR score 4. 30-day surgical mortality

Anaesthetic/perioperative practice area	Causal mechanism linked to implementation of data feedback (H1)	Causal mechanism linked to implementation of enhanced feedback (H2)	Hypothesised effects of feedback	Dependent measures
Postoperative nausea: the use of appropriate antiemetics and analgesic that provides adequate pain control while minimising PONV side effects	Feedback of data concerning variation in patient-reported nausea will cause improved analgesic and antiemetic therapy including enhanced monitoring and compensatory actions	Provision of enhanced feedback including detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination will facilitate personal professional learning from case experience and motivation to improve or maintain high standards of practice	Primary effect: implementation of feedback and escalation of feedback intensity will result in patients reporting enhanced freedom from moderate and severe pain during the recovery period and an enhanced overall QoR experience	<ol style="list-style-type: none"> 1. Pain on admission (first conscious pain score) 2. Patient-reported freedom from nausea, dry-retching and vomiting during the recovery period 3. Postoperative QoR score

H, hypothesis.

BOX 6 Research hypothesis specified to investigate how lower-ranked anaesthetists responded to the feedback intervention

H3: Implementation and escalation of performance feedback at the level of the individual consultant anaesthetist will improve professional practice in anaesthetists ranked lowest at baseline. Any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort.

H, hypothesis.

Intervention timeline

The development and features of the feedback intervention implemented within this project have been described in detail elsewhere in this report and the current section is, therefore, confined to a summary overview to serve as context for the proposed statistical analysis and interpretation of the findings. During the course of the project, many minor iterations were made to the feedback that consultant anaesthetists received, both owing to development in the available quality indicator data set and in response to specific requests and feedback from the anaesthetist group. The project developed following the CLAHRC programme quality improvement model, which draws on established improvement science in health care and advocates rapid-cycle iterative development of quality improvement solutions.

In order to pursue the aims for the evaluation, maximise learning in terms of the effects of specific modes of feedback intervention, it was necessary to take a rationalised view of the intervention timeline that best represented the main phases of development in the project. *Figure 5* provides an overview of the project timeline and illustrates the relationship between both of the main feedback iterations executed within the project representing the main milestones in the development of the local feedback programme (four in total, themselves a simplification of what was in reality a more organic developmental sequence), mapped to a number of rationalised intervention timelines based on functional analysis of the resulting feedback that anaesthetists received.

A two-stage intervention timeline was selected a priori for the primary statistical analysis as representing the main functional iterations of the feedback intervention and illustrated by the shaded timeline in *Figure 6*. The two-stage intervention model distinguished between basic and enhanced feedback (which was most closely related to the distinction between passive and active feedback, as suggested by previous literature in medical education). Baseline data collection began in March 2010, with basic feedback (version 1 of the report) introduced 6 months later in October 2010. Enhanced feedback (version 3 of the feedback report) was introduced in July 2012 and ran until the end of the project in November 2013.

Basic feedback consisted of the provision of monthly personal data summaries in tabled form for a limited number of summary quality metrics, compared with group-level averages without adjustment. Limited longitudinal and normative comparisons were included in graphical format. In contrast, the enhanced feedback phase of the intervention employed a design rationale that was driven by more active engagement with users and with a view towards providing specific data and statistical perspectives geared towards supporting personal learning from case experience. Provision of enhanced feedback included, in addition to the basic feedback content, monthly detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination. During the enhanced feedback phase, engagement with the anaesthetist group was much more active, involving regular presentation of statistical results at meetings, consultative interviews by the research team for formative evaluation of the preferred features of the feedback for end-users, and more focused engagement and facilitated peer interaction on specific specialty areas in which potential quality issues were identified (e.g. pain management after gynaecological surgery). During the enhanced feedback phase, the scope of data collection was increased to include multiple sites, which increased the prominence of the quality monitoring and feedback activities within the broader perioperative department across the trust as a whole.

Calendar year	2010 (weeks 9–53)		2011 (weeks 1–53)		2012 (weeks 1–53)		2013 (weeks 1–44)	
Week number	Weeks 201,009–201,039		Weeks 201,040–201,209		Weeks 201,210–201,226		Weeks 201,227–201,330	
Main feedback iterations	Baseline (no feedback)		Feedback v1 (basic)		Feedback v2 (transitional)		Feedback v3 (enhanced)	
Two-stage rationalised intervention timeline for evaluation	Baseline (no feedback)		Feedback v1 (basic)					
One-stage simple intervention timeline for sensitivity analysis	Baseline (no feedback)		Feedback					
Three-stage complex intervention timeline for sensitivity analysis	Baseline (no feedback)		Feedback v1 (basic)		Feedback v2 (transitional)		Feedback v3 (enhanced)	

FIGURE 5 Intervention timeline variants depicting relationship between the intervention timelines used in the evaluative ITSA and the main observed iterations in the development of the feedback intervention that took place within the project. v, version.

Calendar year	2010 (weeks 9 – 53)	2011 (weeks 1 – 53)	2012 (weeks 1 – 53)	2013 (weeks 1 – 44)
Intervention timeline	Baseline (no feedback)	Feedback v1 (basic)	Feedback v2 (transitional)	Feedback v3 (enhanced)
Week number	Weeks 201,009 – 201,039	Weeks 201,040 – 201,209	Weeks 201,210 – 201,226	Weeks 201,227 – 201,330
Admission type, ASA, age and gender	Shaded	Green	Green	Green
Anaesthetic type	Shaded	Green	Green	Green
Anaesthetist identifier	Shaded	Green	Green	Green
Hospital site identifier	Shaded	Green	Green	Green
Multisite case data	Shaded	Shaded	Green	Green
PACU pain and nausea scores	Shaded	Green	Green	Green
Patient temperature	Shaded	Green	Green	Green
OoR Scale (including pain and nausea data)	Shaded	Green	Green	Green
30-day mortality	Shaded	Green	Green	Green
SSIs	Shaded	Green	Green	Green

FIGURE 6 Timeline of key data items present in case-level data set (research quality indicators data set merged with trust patient administration system and theatre administration data). Shaded areas of the timeline denote the presence of data, while empty cells indicate weeks for which data are missing. Note that in several instances baseline (pre-feedback) data are incomplete and it is, therefore, only possible to compare basic with enhanced feedback conditions. v, version.

In addition to the primary two-stage intervention model, two further intervention timelines were specified, also prior to examination of the final data set, in order to perform further sensitivity analysis. Specifically, a simple single-phase intervention timeline was defined to assess the overall impact of feedback versus no feedback while controlling for temporal trends, in order to investigate whether or not this model better fitted trends in the data. A further 'complex' three-stage intervention model was additionally specified to account for the roll-in period immediately prior to the enhanced feedback intervention (a period of 4 months from March 2012), in which a series of rapid development cycles were undertaken to implement and pilot new features within the feedback report, before the major iteration represented by version 3 was adopted in July 2012. The three-stage intervention timeline was specified to investigate any early effects of new features which might have attenuated the immediate impact of implementation of the enhanced feedback phase.

Sample and inclusion criteria

The target patient population for the anaesthetic feedback intervention was all perioperative cases undergoing a surgical procedure requiring an anaesthetic intervention and subsequent stay in recovery (subsequently referred to as the PACU). As the data feedback intervention was administered at the level of the individual participating anaesthetist and was hypothesised to operate in part through supporting personal professional learning and behaviour change, the caseload of participating anaesthetists was selected as an inclusion criterion (i.e. perioperative cases performed by locums or anaesthetists who had not been long-term participants in the programme were excluded). Thus, it was hypothesised that the intervention operated at the level of the individual's professional practice and the effects of enhanced perioperative normothermia, analgesic and antiemetic therapy would be observable at the level of the original or pilot anaesthetist group's caseload.

Over the course of the programme, the feedback intervention was implemented beyond the pilot site and consultant group at St Mary's Hospital, and incrementally implemented in the broader trust perioperative service at two additional sites, each with a substantial case volume. In order to ensure longitudinal homogeneity in the perioperative cases sampled for analysis and to exert an acceptable degree of experimental control in the evaluation, a variety of contextual data for each case describing patient demographics, theatre site and procedural information were obtained from local administrative systems, to permit further disaggregation of cases and sensitivity analysis. The primary inclusion criteria, however, was that the case had been performed by one of the anaesthetists in the pilot group for the programme. The pilot group comprised 44 consultant anaesthetists who had the following characteristics:

1. had participated in the programme owing to the fact that they were predominantly based at (and conducted the majority of their cases at) St Mary's Hospital in one or more phases of the programme
2. had received regular feedback from anaesthetic quality indicators measured for their cases over a prolonged period of time (i.e. appeared in each study epoch or both of the latter two study epochs representing basic and enhanced feedback)
3. had contributed a substantial volume of cases to the research data set (each anaesthetist had contributed between 194 and 3292 individual surgical cases over the course of the study).

By controlling for a select number of attending anaesthetists who had been exposed to the intervention, this ensures that the effects of the intervention are not diluted by staff movement or late induction to the programme and ensures a degree of stability in case mix over time. Where the intervention timeline assessed was truncated owing to missing data (described in *Research measures and data collection*), this permitted the impact of the programme on a different case volume to be examined in specific instances, for example when the included timeline could draw on multisite data for analysis (i.e. broader impact) or when complete patient administrative data permitted a narrower focus on elective GA cases undertaken by the pilot consultant group. Similarly, within the pilot anaesthetist group, subsamples were drawn for sensitivity analysis (described in *Statistical analysis*) based on anaesthetists with a high number of baseline data and anaesthetists who fell in the lower 50th percentile of ranking on specific quality indicators during the baseline or early phases of the programme.

Research measures and data collection

A range of data items representing anaesthetic quality indicator and perioperative outcome endpoints for the evaluation were collected between March 2010 and November 2013, along with a range of additional data items compiled from additional sources. A complete description of the measures and their source and data range is included in *Table 43*. Data items collected in the primary study site at St Mary's Hospital used a standardised data collection form ('PACU form') which is reproduced in *Appendix 5*. In addition to the research measures detailed in the table, various time series variables and intervention descriptors were generated for the purposes of ITSA, and these are detailed in *Statistical analysis*. In addition to those detailed in *Table 41*, the following variables were compiled based on administrative data for the purposes of covariate analysis:

- patient age
- patient ASA score
- patient gender.

In order to select appropriate subgroups of cases for sensitivity analysis, the following metrics were additionally compiled from the theatre administration system data set:

- responsible anaesthetist (throughout the study, all participating anaesthetists were assigned a numeric code to preserve anonymity)
- admission type (elective or non-elective)
- type of anaesthesia (GA or other).

The research measures were derived from multiple data sources, including continuous collection of a core set of patient-reported quality indicators in the PACU of the primary study site at St Mary's Hospital for the duration of the project and in additional Imperial Healthcare sites for the later phases of the project. Multisite data collection was necessary in order to provide participating anaesthetists with comprehensive feedback on their surgical cases, as the majority of consultant anaesthetists at Imperial Healthcare practised anaesthesia at multiple sites within the institution. PACU data were concurrently supplemented by surgical case administrative data from the theatre administration system used in Imperial Healthcare NHS and retrospectively from the local patient administration system which included Hospital Episode Statistics and diagnostic coding. Both supplementary data sources accumulated patient data from all Imperial hospital sites.

By the end of the project in November 2013, the master database, which serves as the basis of both the monthly feedback reports and the case-level data set for the evaluation detailed in this report, had accumulated over 100,000 individual surgical records since the onset of data collection in March 2010. Over 12,000 manually completed PACU data collection forms from St Mary's Hospital, including QoR score data, were entered into the data set. Across all Imperial sites, over 63,000 pain and nausea scores were recorded and over 65,000 core temperature readings taken during the period of the project. For the ITSA, the rate of data collection supported weekly aggregation with the resulting available weekly time series data points varying dependent on source as illustrated in *Figure 6*.

The effects of variation in the available case-level data are missing time points in the intervention timeline for certain variables, particularly at the start of the baseline period. This was because the theatre administration system database maintained by perioperative staff and the source of information concerning case-level admission type, patient identification number, ASA score, patient demographics and procedure type was implemented in July 2010 after the onset of the project. Once theatre administrative data were available, these were used to supplement PACU-collected case-level data. The patient identifiers included within the theatre administration data additionally enabled case-level matching with hospital episode data, which yielded diagnostic coding and mortality data. Thus, SSI rates and 30-day surgical mortality could be calculated only from July 2010 for the evaluation. Subsequent to the development of the original PACU data collection form based on a QoR Scale, two additional indicators for postoperative

TABLE 43 Table defining dependent variables used as research end points within the evaluation

Metric	Definition	Source	Type	Data level	Timeline
Mean weekly patient temperature	Weekly average patient core temperature taken on arrival in PACU	Manually collected data recorded on PACU form at St Mary's and in theatre administration system at all sites	Continuous scale variable	Aggregated by project week for time series analysis. Original case level metric: patient core temperature	Full: March 2010 to November 2013
Weekly proportion of patients with temperature < 36 °C	Weekly proportion of patients with core temperature < 36 °C on arrival in recovery	Manually collected data recorded on PACU form at St Mary's and in theatre administration system at all sites	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case level metric: patient core temperature < 36 °C (binary)	Full: March 2010 to November 2013
Mean weekly QoR Scale score	Weekly average QoR Scale score. ²⁶ Aggregated scale recorded by PACU nurses at patient bedside prior to discharge from recovery. Scale based on 8 items each scored 0–2 representing functional aspects of recovery, freedom from pain and nausea, provision of support and general psychological well-being following surgery	Manually collected data recorded on PACU form at St Mary's site only	Continuous scale variable (ranging from 0 to 16)	Aggregated by project week for time series analysis. Original case level metric: patient QoR Scale score	Full: March 2010 to November 2013
Weekly proportion of patients with QoR Scale score over 14	Weekly proportion of patients with a QoR score in the range 15–16 at discharge from recovery unit, indicating high-quality recovery	Manually collected data recorded on PACU form at St Mary's site only	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case level metric: patient QoR Scale score above 14 (binary)	Full: March 2010 to November 2013
Weekly proportion of patients mostly free from PONV	Weekly proportion of patients responding at level 2 'most of the time' for QoR Scale item: 'Have you been free from nausea, dry retching or vomiting' prior to discharge from recovery	Manually collected data recorded on PACU form at St Mary's site only	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case level metric: patient QoR PONV item (ordinal 0 to 2)	Full: March 2010 to November 2013

continued

TABLE 43 Table defining dependent variables used as research end points within the evaluation (continued)

Metric	Definition	Source	Type	Data level	Timeline
Weekly proportion of patients mostly free from moderate or severe postoperative pain	Weekly proportion of patients responding at level 2 'most of the time' for QoR Scale item: 'Have you been free from any severe pain or constant moderate pain' prior to discharge from recovery	Manually collected data recorded on PACU form at St Mary's site only	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: patient QoR pain item (ordinal 0 to 2)	Full: March 2010 to November 2013
Weekly proportion of patients with first conscious pain score recorded as 'none' or 'mild'	Weekly proportion of patients whose first recorded pain scale score on arrival in recovery is 0 (no pain) or 1 (mild), as opposed to 2 (moderate), 3 (severe) or 4 (unbearable)	Manually collected data recorded on arrival in recovery at all Imperial sites	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: patient pain score (scale 0 to 4)	Partial owing to late introduction of this metric: September 2010 to November 2013
Weekly proportion of patients experiencing retching or vomiting	Weekly proportion of patients experiencing retching or vomiting during stay in recovery as indicated by a score of 2 in the following categories: 0 (nil), 1 (nausea) and 2 (vomiting/retching). Nurse-observed measure not dependent on input from the patient	Manually collected data recorded during stay in recovery at all Imperial sites	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: patient nausea score (ordinal 0 to 2)	Partial owing to late introduction of this metric: September 2010 to November 2013
Weekly proportion of patients free from nausea	Weekly proportion of patients experiencing no nausea, retching or vomiting during stay in recovery as indicated by a score of 0 in the following categories: 0 (nil), 1 (nausea) and 2 (vomiting/retching). Nurse-observed measure not dependent on input from the patient	Manually collected data recorded during stay in recovery at all Imperial sites	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: patient nausea score (ordinal 0 to 2)	Partial owing to late introduction of this metric: September 2010 to November 2013

Metric	Definition	Source	Type	Data level	Timeline
Weekly SSI rate	Weekly proportion of patients with one or more SSI-specific ICD diagnostic codes in their patient administration system record for the current hospital episode in which the stay in PACU occurred, including T81.4 'infection following a procedure' or any of a number of additional codes indicating infection following a specific surgical procedure (T82.6, T82.7, T83.5, T83.6, T84.5, T84.6, T84.7, T85.7, T87.4)	Retrospective case match between theatre administration system patient identifier and hospital records	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: presence of SSI (binary)	Partial owing to late introduction of theatre administration system records which are necessary for patient administration system record matching: July 2010 to November 2013
Weekly 30-day surgical mortality rate	Weekly proportion of patients who subsequently experienced mortality within 30 days of the date of surgery. Calculated from patient administration system records of 'date of death' and 'days until death'	Retrospective case match between theatre administration system patient identifier and hospital patient administration system records	Continuous scale variable (proportion ranging from 0 to 1)	Aggregated by project week for time series analysis. Original case-level metric: presence of 30-day mortality (binary)	Partial owing to late introduction of theatre administration system records which are necessary for PAS record matching: July 2010 to November 2013

ICD, *International Classification of Diseases*; PAS, patient administration system.

pain and nausea were added to the data set in September 2010 (these were included in addition to the pain and nausea indicators included as part of the QoR Scale, for which data are available for the full duration of the project timeline). While the additional pain and nausea metrics may be used to assess the effects of development of basic feedback within the intervention timeline (i.e. hypothesis 2), they are not available for the baseline period and, therefore, cannot be used to investigate the effects of implementation of basic feedback (hypothesis 1). Variance in availability of data across individual measures, subsamples and specific analyses meant that an analytic strategy had to be adopted that maximised available cases (and hence statistical power) for the primary analyses. Secondary or sensitivity analysis was subsequently performed to investigate the effects of introduction of further quasi-experimental controls through variant case selection criteria but invariably at a loss of the available length of time series for analysis and with further implications for statistical power.

Statistical analysis

For the sake of clarity and fidelity to our evaluative aims, the plan and sequence of statistical analysis was split into two phases: (1) primary ITSA and (2) secondary sensitivity analysis. The primary analysis was designed to directly and deductively address the principal research hypotheses specified for the evaluation, while making best use of the available data to maximise statistical power in the models employed. The secondary analysis, by necessity, was supplementary to the primary analysis and was designed to investigate how the primary results behaved under variable statistical criteria (i.e. a sensitivity analysis). The questions addressed by the secondary analysis invariably employed subsets of data, at a loss of statistical power, and employed a more exploratory or inductive mode of enquiry. Specifying our primary analytic aims a priori and relegating additional analyses to a secondary or supportive function allows minimisation of escalation in the family-wise error rate due to multiple testing. Consequently, the usual caveats should be applied to interpretation of the findings from the secondary analysis, which involve fitting a larger number of statistical models.

Primary interrupted time series analysis

The statistical approach used for ITSA was segmented regression analysis¹⁸⁵ in which parameters representing interrupt (onset of each intervention phase), baseline trend and post-intervention trend are entered into a linear regression model, along with covariates, to account for variance in each dependent measure. Least squares estimation was used for statistical modelling and the Durbin Watson statistic to test for autocorrelation. The models were run in Statistical Product and Service Solutions (SPSS) version 21 (IBM SPSS Statistics, Armonk, NY, USA) for Windows (Windows® operating system, Microsoft Corporation, Redmond, WA, USA). The general form of the segmented regression equation, along with explanation of its components, is described below:

$$Y_t = \beta_0 + \beta_1 \text{time}_t + \beta_2 \text{intervention}_t + \beta_3 \text{time after intervention}_t + e_t. \quad (1)$$

In this equation, the following parameters and associated statistical estimates are defined:

- Y for time t is the dependent variable expressed as an aggregated weekly value (e.g. proportion of patients with temperature $< 36^\circ\text{C}$ in week t).
- time is a continuous variable indicating time in consecutive weeks at time t from the start of the measurement period.
- intervention is a nominal variable indicating whether time t occurred before the onset of the intervention phase or after it, taking the value 0 for the former and 1 for the latter.
- $\text{time after intervention}$ is a continuous variable expressing the number of consecutive weeks that time t occurred post onset of the intervention phase. If time t occurred prior to the onset of the intervention phase then $\text{time after intervention}$ takes the value 0.
- β_0 estimates the baseline level of the dependent measure at $\text{time} = 0$.
- β_1 estimates the change in the dependent measure associated with each consecutive increment of time in the baseline or pre-intervention phase (i.e. the pre-intervention 'slope' or 'trend').

- β_2 estimates the change in level of the dependent measure immediately following onset of the intervention phase.
- β_3 estimates the change in the slope of the dependent measure associated with each consecutive increment of *time* following onset of the intervention phase (the post-intervention slope is, therefore, given by $\beta_1 + \beta_3$).
- e_0 represents the error term in the fitted model.

For the sake of clarity, in this mode of analysis, a desirable effect in the dependent measure associated with the onset of the intervention (or a phase of the intervention), is indicated by:

- a significant stepwise shift in the level of *Y* in a desirable direction and corresponding to the interruption of the time series by onset of the intervention phase (as indicated by a significant β_2 coefficient), or
- a significant change in the post intervention slope of *Y* in a desirable direction and corresponding to the interruption of the time series by onset of the intervention phase (as indicated by a significant β_3 coefficient), or
- both of the above.

The model described above is applicable to a simple, single-phase intervention, but the principle extends to multiphase interventions through the addition of further parameters expressing the onset and subsequent trend associated with further intervention phases. In the current analysis, the majority of models fitted to the data set took the latter form. In this sense the serial effects of a phased intervention, such as escalating feedback, may be estimated and the additive impact of the components of a complex intervention that are separated in time can be isolated.

In the current analysis, parameters for the two-phase feedback intervention model were estimated as the primary approach to answering the research question (hypothesis 2). In this model, three segments in the time series are specified, representing (1) the pre-intervention baseline phase, (2) the basic data feedback phase, and (3) the enhanced feedback phase. In addition to this primary three-stage ITS model, missing data in the pre-intervention phase of the project for the analogue pain scale, the nurse-reported PONV indicator and a range of admission type and anaesthetic-type descriptors necessitated fitting two-stage models that focused on the change in the dependent measures between basic and enhanced feedback conditions, without pre-intervention data.

Secondary sensitivity analysis

Following the primary analysis of the effects of the feedback intervention using ITSA, a series of sensitivity analyses were performed for each analysed quality indicator by running a number of ITS model variants, as described below. Sensitivity analysis was performed to investigate the effects of implementing a range of additional quasi-experimental controls through varying the case inclusion criteria, to investigate the fit of alternative intervention models to the observed data and to explore the robustness of findings from the preliminary models, made necessary by the presence of missing data in the research data set.

In the methodological descriptions of the secondary analyses that follow, any specific case inclusion criteria or subgroup selections are applied in addition to those that define the primary research data set used in the primary analysis. For the sake of clarity, the primary inclusion criteria were:

1. cases performed by the pilot group of 44 anaesthetists, based predominantly at St Mary's Hospital, who received the first version of the feedback intervention, and
2. cases performed in St Mary's main theatres and recovery suite, as opposed to any other hospital site or theatre suite.

Effect on lower 50th percentile anaesthetist subgroup

In order to investigate hypothesis 3, that the feedback intervention would have a positive effect on the lower-ranked subgroup of anaesthetists who fell in the lower end of the rank distribution of their peer group, the lower 50th percentile of anaesthetists (or those ranked below the median rank in the group) was selected for subgroup analysis.

Throughout the study, in the research data set, all anaesthetists were assigned a numeric code to preserve anonymity. The lowest 50th percentile subgroup was selected on the basis of rank-ordering anaesthetist codes according to average or aggregated scores on each dependent measure during both the baseline and the initial (basic) feedback phase. In this manner, two unique subgroups were generated for each quality measure, based on performance on the measure: one for performance during the baseline period for analysis of the effects of the full intervention timeline, and one based on performance in the initial feedback phase for analysis of the effects of implementation of enhanced feedback based on a truncated timeline.

To attenuate regression to the mean effects associated with selection of experimental cohort based on baseline scores, only anaesthetists with large data sets during the assessment period were included in the rank ordering, to reduce measurement error in the test statistic used to calculate rank. This meant that of the 44 anaesthetists in the pilot group, 30 were entered into the ranking for the baseline period and 40 were entered into the ranking for the first feedback intervention phase. The resulting lower 50th percentile groups for the analysis comprised 15 and 20 anaesthetists, respectively. Time series data sets based on aggregated weekly figures were calculated as for the primary analysis, but limited to surgical cases in which the recorded responsible anaesthetist fell within the target subgroup.

Statistical control of case age, gender and American Society of Anesthesiologists score

As described previously, prior research has highlighted a number of factors affecting patient satisfaction with surgery and ratings of postoperative pain, nausea and QoR.^{28,29} The severity of the patient's condition prior to surgery, as indicated by the ASA score, is likely to result in increased surgical duration and complexity of procedure, corresponding to more complex and prolonged anaesthetic requirements with increased risk of postoperative side effects affecting the comfort of the patient during the recovery period. Similarly, older patients as a surgical group are likely to present more complex requirements with multiple comorbidities requiring management during perioperative care. Research suggests that there are gender and age differences in perceptions of satisfaction with surgery and anaesthesia, as well as in expectations regarding acceptable postoperative pain and discomfort.

Given these factors, age, gender and preoperative condition severity are all likely to be valid confounders in any evaluation of an intervention designed to impact on patient-reported satisfaction and perioperative outcome indicators, of the type used in this study. In order to address this in the analysis, secondary sensitivity analysis was performed by fitting further ITS models for each of the dependent measures used in the study, with prior entry of age, gender and ASA score metrics in the regression model, to control for any associated effects of these parameters.

For the time series models, mean weekly patient age was calculated as the research covariate. Similarly, ASA score was calculated as the weekly aggregated mean. Patient gender was expressed as weekly percentage of male cases. The covariates were entered into the regression model as a single block to partial out variance associated with these factors prior to estimation of the unique contribution of the intervention time series parameters. The results from the nested covariate model were compared with the results from the original model without covariates in order to determine whether or not longitudinal variation in age, gender and severity within the surgical case group attenuated any temporal effects of the feedback intervention.

Owing to the absence of theatre administrative data for the earliest phase of the intervention timeline, it should be noted that entering age, gender and ASA score as covariates with listwise exclusion effectively truncates the length of timeline that can be evaluated in the analysis, meaning that the comparative results based on the original full-time series model must be interpreted carefully.

Intervention timeline variants

As described in detail above in description of the intervention timeline, the primary analysis involved fitting ITS models for a two-stage intervention model in which three study epochs were delineated a priori: (1) baseline (no feedback), (2) basic feedback (version 1) and (3) enhanced feedback (version 3). For the reasons discussed, sensitivity analysis based on variations in this timeline suggested by retrospective analysis of the developmental process that took place within the anaesthetics quality improvement programme (but prior to examination of the data) might yield further insight into how the observed perioperative parameters responded to the feedback initiative longitudinally. The specific aims of conducting sensitivity analysis of the intervention timeline variants was, therefore, twofold:

1. to investigate whether a more complex three-stage intervention model that accounted for the intermediary development period before the implementation of enhanced feedback could better account for the observed longitudinal variance in patient temperature data, as an elaboration of hypothesis 2
2. to investigate whether or not a more simple intervention model (absence vs. presence of feedback) could better account for the observed longitudinal variance in patient temperature data, as a direct test of hypothesis 1.

Variant time series models were generated by inclusion or reduction of time series parameters in the segmented regression model to account for more or less interrupt time points (and subsequent changes in slope) in the fitted model.

Variations in case inclusion according to site and admission/anaesthetic type

Owing to the complexity of potential contextual and case-level factors that might influence observed scores on the dependent measures used for evaluation in this study, further sensitivity analysis of the primary model results were run based on varying the case inclusion criteria in order to investigate the effects of both broader and more stringent criteria on the analysis of intervention effects. Two specific variants were investigated:

1. Limiting case inclusion using two criteria: (a) limiting to cases recorded in hospital administrative data as 'elective' and excluding 'non-elective', and (b) limiting to cases recorded in the theatre administration system as requiring a GA and excluding any other anaesthetic type. These criteria were specified to create a more homogenous subgroup of surgical cases in order to see if this affected estimation of the effects of the intervention. Owing to the resultant reduction in number of cases used to calculate each time series data point, the precision of these estimates would inevitably be affected, however, requiring care in interpretation. Additionally, owing to missing admission and procedural data in the research data set, implementing these criteria truncated the length of time-series data available to the analysis at both ends of the time series, at a loss of statistical power.
2. Broadening the case inclusion criteria beyond the pilot intervention site (St Mary's main theatres), to include surgical cases performed by the pilot anaesthetist group at other Imperial sites. Multisite data for patient temperature, pain on arrival in recovery, nurse-reported nausea, SSI rate and 30-day mortality rate became available retrospectively from June 2011 (i.e. half way through the first intervention phase) and was subsequently implemented in the feedback that anaesthetists received from iteration version 3 (i.e. 'enhanced' feedback). QoR Scale data were not collected at the secondary study sites. In principle, analysis of surgical case data at the multisite level should provide a more complete view of the longitudinal performance of anaesthetists in the pilot group and their response to escalation in the intensity of feedback. In order to avoid the onset of multisite data collection acting as a confounding factor in any analysis of intervention effects, however, the study baseline and early intervention phase 1 data must be excluded from this analysis, truncating the available time series data for evaluative purposes.

Similar to the procedure for sensitivity analysis outlined above, for hospital site, admission type and anaesthetic type, alternative time series data sets were created for each dependent measure based on varying the case inclusion criteria. The results from fitting ITS models were then compared with the results from the primary analysis to determine if variable case inclusion criteria yielded further insight into the intervention effects.

Results

Following summary of the statistical features of the observed macro-level data set, the statistical results from inferential analyses are presented below by outcome measure.

Observed sample

Over the course of the project (March 2010 to November 2013) and excluding locums, 112 consultant anaesthetists conducted anaesthesia at Imperial sites for 102,034 recorded surgical cases appearing in the master data set. The subgroup of cases utilised in the majority of analyses for the evaluation was defined as all surgical cases performed by consultant anaesthetists who were primarily based at the pilot intervention site, St Mary's Hospital, and who had received the feedback intervention and registered cases in both the main phases of the feedback programme. The pilot anaesthetist group's data set will hitherto be referred to as the 'research' or 'evaluation' data set and consisted of the cases of 44 consultant anaesthetists, each returning between 194 and 3293 case records over the course of the project. *Figure 7* depicts the frequency of cases returned by anaesthetist (anonymised) and by study intervention phase. Fourteen of the 44 anaesthetists registered no case data during the baseline period (March to September 2010), having joined the group during the initial intervention phase, and these anaesthetists were subsequently excluded from any longitudinal analysis that involved establishing a subgroup based on baseline rankings.

The research data set (pilot anaesthetist group) included 50,235 cases in total over the duration of the project. Of these, 22,670 cases were performed at the primary intervention site (St Mary's Hospital) and the remaining 27,565 cases were performed by members of the pilot anaesthetist group at other Imperial sites (the anaesthetists received feedback on all their cases regardless of site). The additional multisite cases were added to the data set and included in the feedback that anaesthetists received from June 2011. *Figure 8* provides an overview of the rate of case accumulation over time and by site for the duration of the project, within the research data set. To account for the onset of data collection at parallel sites being a confounder in any analysis of longitudinal trend, in the subsequently reported analyses, the time series models were fitted first for the St Mary's main theatres data and subsequently for a data set comprising all cases performed by the pilot anaesthetist group, as part of the sensitivity analysis.

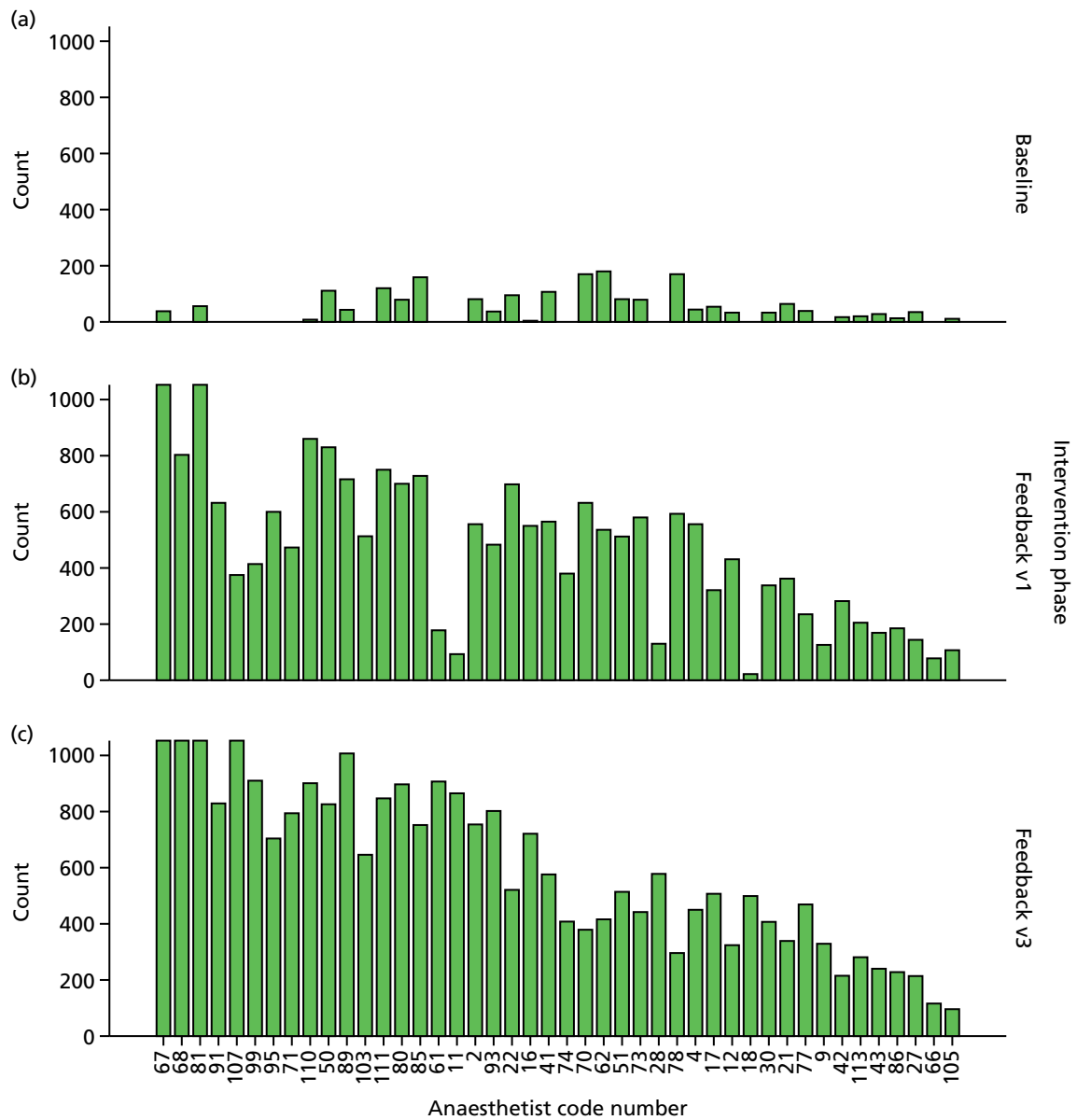


FIGURE 7 Case frequency by anaesthetist (pilot group only) and main intervention phase (study epoch). v, version.

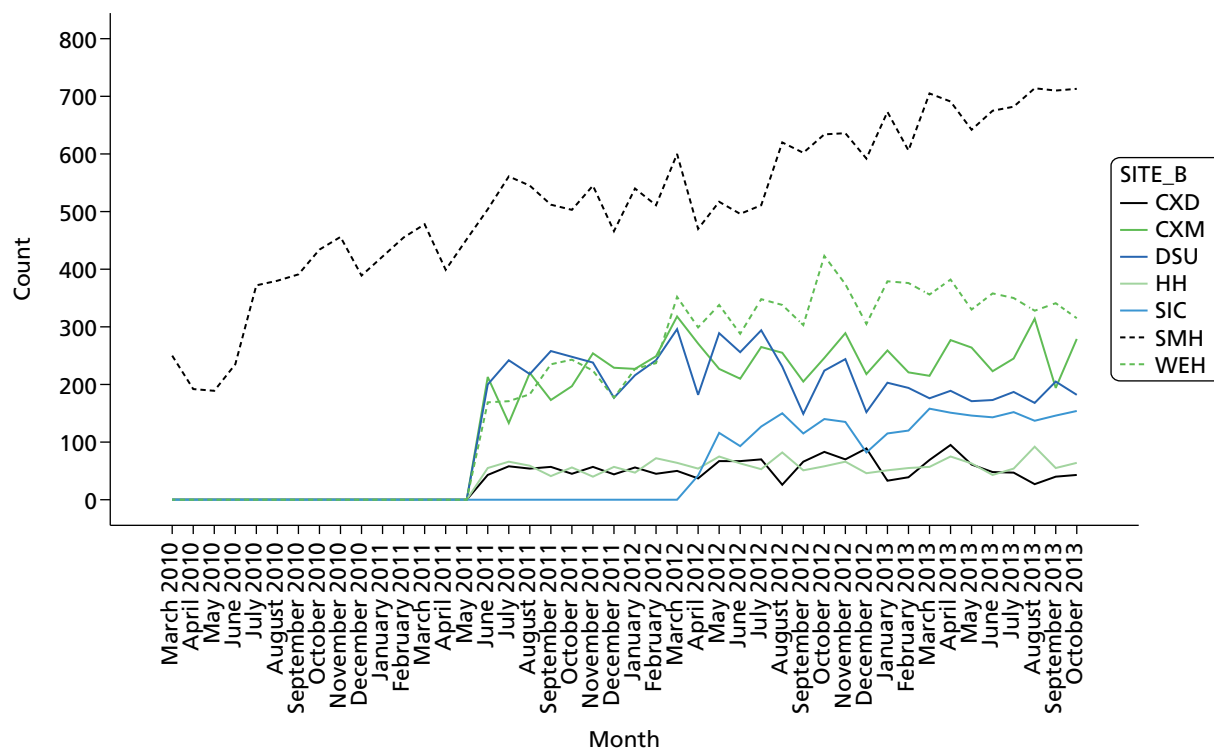


FIGURE 8 Monthly rate of accumulation of perioperative cases within the research data set by contributing Imperial site. The 'SMH' category represents cases performed at St Mary's Hospital main theatres and this group was selected as the primary research sample for the evaluation, representing a homogeneous group of patients over the duration of the intervention programme. CXD, Charing Cross Day Surgery Unit; CXM, Charing Cross Main Theatres; DSU, Day Surgery Unit at St Mary's; HH, Hammersmith Hospital; SIC, Surgical Innovation Centre at St Mary's; SMH, St Mary's Hospital; WEH, Western Eye Hospital.

In terms of the frequencies of different types of surgical cases performed by the pilot anaesthetist group at St Mary's Hospital main theatres, the majority of cases were classified as falling within the following specialties, according to local administrative data: orthopaedic (3248), gynaecology (2469), general surgery (2452), ear, nose and throat (2340), and vascular surgery (2318). In terms of anaesthetic type, 28,567 patients received a GA (80.7%) and 6823 patients received a local anaesthetic, regional block or sedation (19.3%). The vast majority of patients were elective (32,316; 84.8%) as opposed to non-elective (5797; 15.2%). *Table 44* presents a breakdown of the case frequencies by specialty and for both levels of the data set, along with overall case counts.

In terms of patient demographics in the pilot anaesthetist group's case data, for cases performed at St Mary's Hospital, 21,737 valid gender descriptors were recorded and representation of male and female patients was well balanced (49.3% female). Across all Imperial sites, including St Mary's main theatres, 49,302 gender descriptors were recorded, slightly biased towards female patients (54% female). The completeness of gender data within the full research data set was 98.1%. *Table 45* includes summary gender data for both levels of the research data set.

In St Mary's main theatres, the mean age of patients in cases performed by the pilot anaesthetist group was 48.83 years (median = 50; SD = 22.77 years), compared with 41.61 years (median = 41; SD = 25.24 years) for patients within the multisite data set. Overall, St Mary's Hospital dealt with a slightly older surgical patient demographic. The distributions of patient age for both levels of data are depicted in *Figure 9*. Both distributions show a peak in case frequency below age 6, indicating the presence of paediatric patients, with St Mary's theatres dealing with a larger proportion of paediatric patients than all of the Imperial theatre suites combined. The completeness of patient age at time of admission data within the full research data set was 98.1% of recorded cases.

TABLE 44 Table of recorded case frequencies at St Mary's Hospital main theatres and for all Imperial sites (including St Mary's Hospital main theatres) by surgical speciality

Surgical speciality	SMH main theatres		Full multisite data	
	Count, <i>n</i>	Row total, %	Count, <i>n</i>	Row total, %
Orthopaedic surgery	3248	58.5	5554	100.0
Gynaecology	2469	33.8	7303	100.0
General surgery	2452	45.4	5403	100.0
ENT	2340	49.5	4728	100.0
Vascular surgery	2318	92.6	2502	100.0
Paediatric surgery	1657	99.9	1659	100.0
Trauma surgery	1566	98.0	1598	100.0
Colorectal surgery	1336	67.1	1991	100.0
Anaesthetics/ICU	836	74.2	1126	100.0
Breast surgery	610	60.3	1012	100.0
Urology	476	24.0	1987	100.0
Haematology/oncology	395	99.7	396	100.0
Plastic surgery	370	41.2	897	100.0
Upper gastrointestinal surgery	338	98.5	343	100.0
Neurosurgery	316	23.7	1331	100.0
Obstetrics	279	98.6	283	100.0
Gynaecology oncology: surgery	201	73.6	273	100.0
Paediatric dentistry	103	100.0	103	100.0
Cardiac surgery	65	6.4	1014	100.0
Multiple	64	100.0	64	100.0
Cardiology	54	74.0	73	100.0
Emergency	53	96.4	55	100.0
Endocrine surgery	53	69.7	76	100.0
Ophthalmology	50	0.6	8800	100.0
Gastroenterology	25	80.6	31	100.0
Maxillofacial surgery	17	100.0	17	100.0
Neurosurgery – deep-brain stimulator	3	100.0	3	100.0
General renal failure/nephrology	2	1.7	120	100.0
Oral surgery	2	66.7	3	100.0
Respiratory medicine	2	100.0	2	100.0
Thoracic surgery	1	0.5	222	100.0
Genital reconstructive surgery	0	0.0	194	100.0
Gynaecology infertility surgery	0	0.0	12	100.0
Hepatobiliary and pancreatic surgery	0	0.0	40	100.0
Podiatry	0	0.0	50	100.0
Transplantation surgery	0	0.0	1	100.0

ENT, ear, nose and throat; ICU, intensive care unit; SMH, St Mary's Hospital.

TABLE 45 Table depicting summary gender statistics for cases performed by the pilot anaesthetist group at St Mary’s Hospital main theatres and across all Imperial sites, including St Mary’s Hospital main theatres

Gender	SMH main theatres, frequency (%)	Full multisite data, frequency (%)
Female	10,724 (49.3)	26,641 (54.0)
Male	11,013 (50.7)	22,661 (46.0)
Total	21,737 (100.0)	49,302 (100.0)

SMH, St Mary’s Hospital.

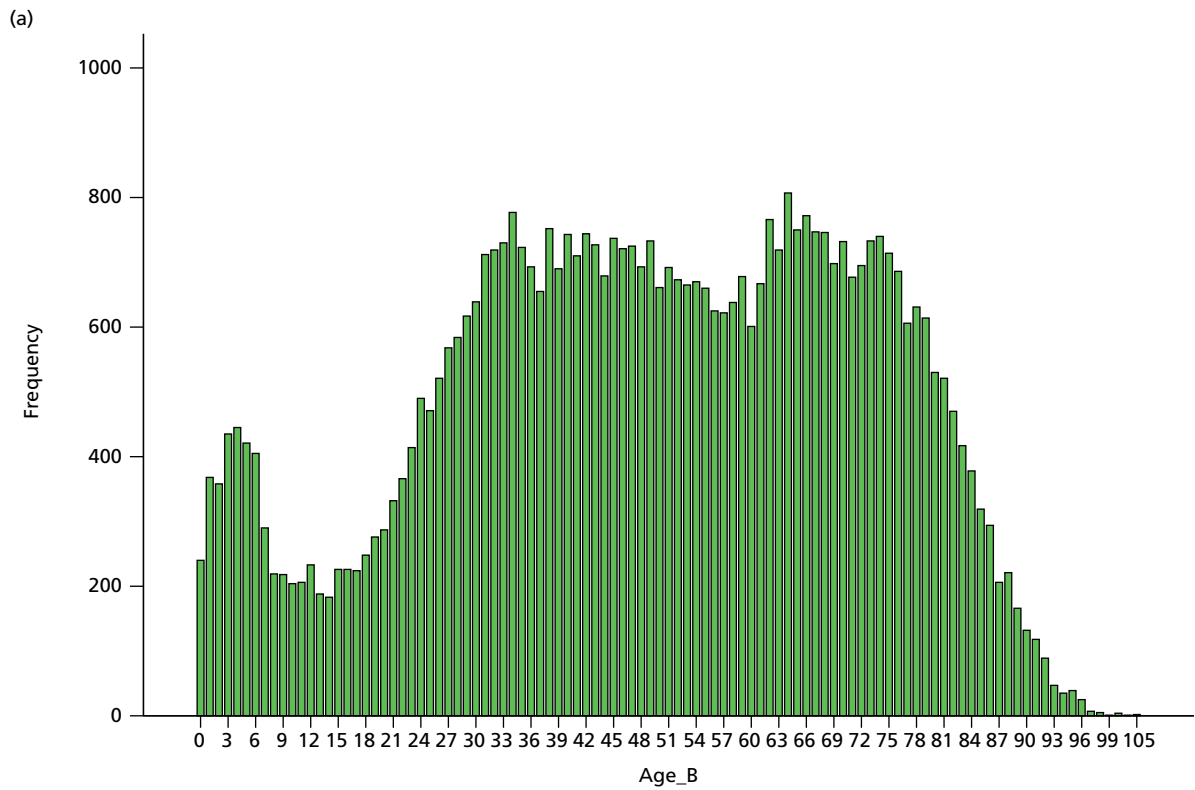


FIGURE 9 Frequency histograms depicting volume of cases by age (years) for (a) full multisite research data set and (b) St Mary’s Hospital main theatres. (continued)

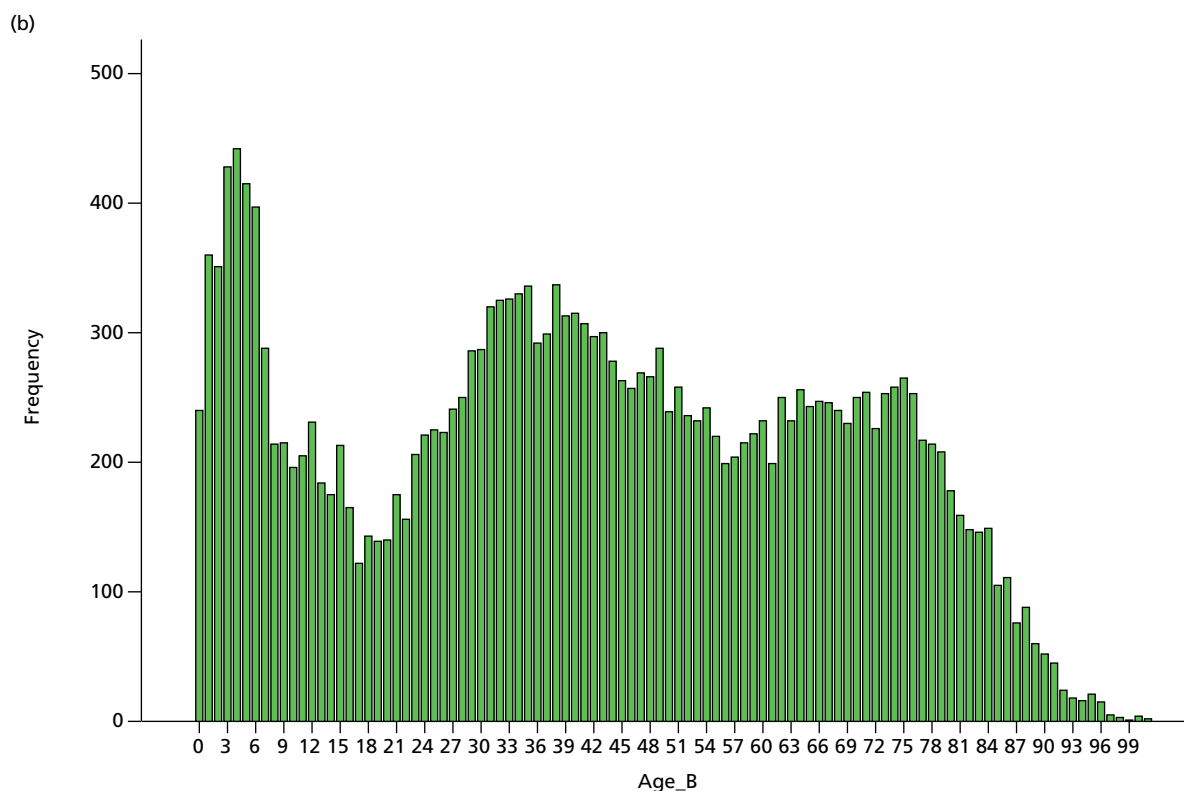


FIGURE 9 Frequency histograms depicting volume of cases by age (years) for (a) full multisite research data set and (b) St Mary's Hospital main theatres.

In terms of the disease severity and fitness of cases before surgery, the ASA five-category system was used for classification. The ASA system assigns a score of 1 to healthy patients, with subsequent indices indicating increasing severity up to a classification of 5, which describes moribund patients who are not expected to survive without the operation. Within the full multisite research data set, 80.1% of cases had a valid ASA classification and the mean ASA score was 1.80 (median = 2.00; SD = 0.86). Among cases performed at St Mary's Hospital, 10.7% missing ASA data were observed, with valid scores yielding a mean of 1.82 (median = 2.0; SD = 0.939). ASA score was, therefore, comparable across both levels of the data set. *Table 46* depicts frequency of ASA score assignments for cases within both levels of the research data set. Although mean ASA scores are comparable between both levels of case data, it should be noted that St Mary's main theatres dealt with proportionally more severe cases with ASA scores in the 3–5 range, compared with cases performed across all Imperial theatres.

TABLE 46 Frequency and valid proportions of ASA score assignments for surgical cases performed both at St Mary's Hospital main theatres and across the whole trust

ASA score	SMH main theatres, frequency (valid %)	Full multisite data, frequency (valid %)
1	9754 (48.2)	18,119 (45.0)
2	5587 (27.6)	13,563 (33.7)
3	3782 (18.7)	7102 (17.7)
4	1009 (5.0)	1329 (3.3)
5	104 (0.5)	124 (0.3)
Total	20,236 (100.0)	40,237 (100.0)

SMH, St Mary's Hospital.

Patient temperature on arrival in recovery

Summary statistics

Over the course of the project and for the pilot anaesthetist group, 19,407 valid surgical cases with patient temperature data were recorded at St Mary's main theatres (representing 85.6% of recorded cases within the research data set). Temperature data were recorded in degrees celsius (metric: 'Temp_B'). From the temperature data, the proportion of patients arriving in recovery with a core temperature below the guideline level of 36 °C was calculated (metric: 'TempU36').

Overall, the mean patient temperature on arrival in recovery was 36.53 °C (median = 36.50 °C; SD = 0.388 °C). Observed patient temperatures varied between 33.4 °C and 39.9 °C. The proportion of valid cases with temperature < 36 °C was 3.5% (672 cases).

Table 47 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, mean patient temperature increased both between the baseline and basic feedback condition and between the basic and enhanced feedback condition. In terms of the proportion of patients with temperature < 36 °C, the proportion increased from 3.88% to 4.53% between baseline and first intervention phase, and then decreased between first and second intervention phase (from 4.53% to 2.24% of cases).

Interrupted time series analysis

The time series charts for both the mean patient temperature and the proportion of patients with temperature < 36 °C, by project week, is depicted in Figures 10 and 11, respectively. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

As described previously, the primary ITS model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

TABLE 47 Longitudinal summary of mean patient temperature on arrival in recovery and proportion of patients with temperature < 36 °C by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)

Metric	Study epoch		
	Baseline	Feedback v1	Feedback v3
Temperature			
Mean	36.420	36.517	36.552
SD	0.378	0.417	0.352
Count	2009	10,255	10,406
Temperature < 36 °C			
Count	61	418	193
Proportion	3.88%	4.53%	2.24%

v, version.

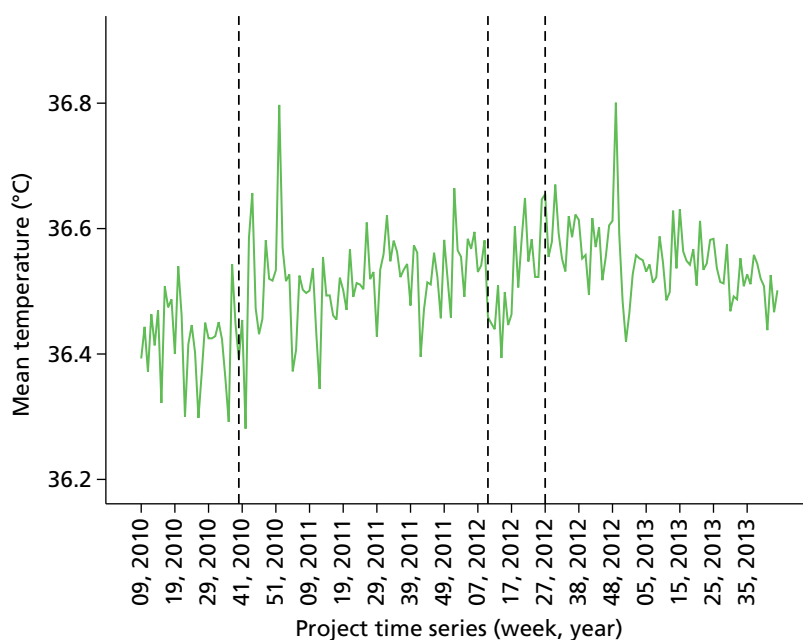


FIGURE 10 Time series for mean patient temperature recorded on arrival in recovery following surgery. Each data point represents an average of all patient temperature data observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback initiative.

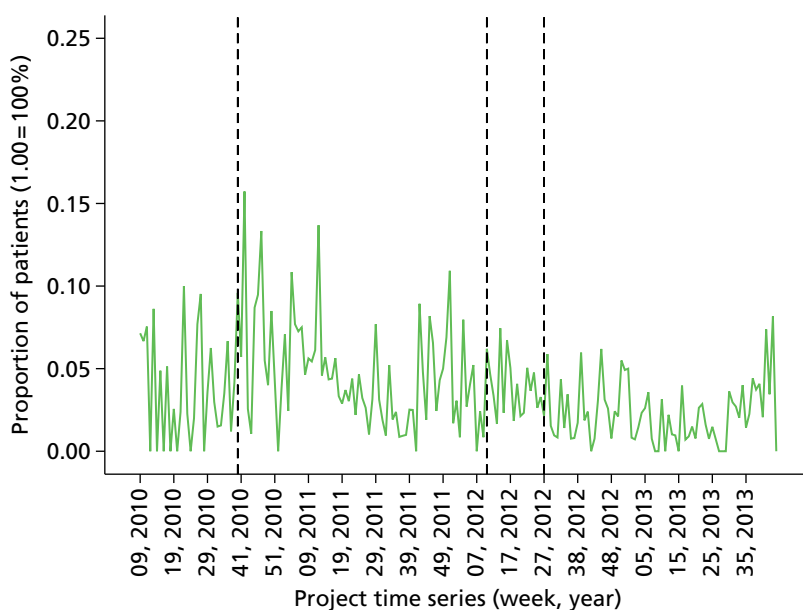


FIGURE 11 Time series for proportion of patients arriving in recovery following surgery with temperature $< 36^{\circ}\text{C}$. Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

The association between longitudinal trends in patient temperature and the specific feedback interventions was estimated using ITS models as described in the statistical methods section above. For the patient temperature data, it was possible to evaluate both main evaluative hypotheses as stated in *Table 48* (the formulation of causal theory underpinning these hypotheses is described at length earlier in this section and elsewhere in the report).

In order to evaluate both hypotheses, ITS models were fitted with two interrupts, representing the onset of basic and enhanced feedback, for both mean patient temperature and proportion of patients with temperature < 36 °C as dependent variables. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary’s main theatres.

For mean patient temperature, the fitted time series model explained 35.2% of the variance ($R = 0.593$; $F = 19.967$; $p < 0.001$). The model parameter estimates are reproduced in *Table 49*. From the model output it can be determined that the onset of both basic (version 1) and enhanced (version 3) feedback was associated with a stepwise increase in mean weekly patient temperature. Controlling for all other temporal trends, implementation of basic feedback after the baseline period was associated with an increase in patient temperature of 0.082 °C ($p < 0.005$) and subsequent implementation of enhanced feedback was associated with an increase of 0.064 °C ($p < 0.005$). Implementation of enhanced feedback was, however, additionally associated with a slight decrease in the rate of change (slope) in patient temperature over time ($\beta = -0.002$ °C; $p < 0.001$).

TABLE 48 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in average temperature of patients arriving in recovery 2. decrease in the proportion of patients arriving in recovery < 36 °C
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in average temperature of patients arriving in recovery 2. decrease in the proportion of patients arriving in recovery < 36 °C

H, hypothesis.

TABLE 49 Model parameter estimates for mean patient temperature, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Parameter	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	36.430	0.024		1511.231	0.000
Baseline trend	-0.001	0.001	-0.376	-0.386	0.700
Feedback (basic v1)	0.082	0.028	0.372	2.899	0.004
Trend after feedback v1	0.001	0.001	0.659	0.709	0.479
Feedback (enhanced v3)	0.064	0.021	0.390	3.080	0.002
Trend after feedback v3	-0.002	0.000	-0.493	-3.947	0.000

v, version.

For the proportion of patients with a temperature $< 36^{\circ}\text{C}$, the fitted time series model explained 19.3% of the variance ($R = 0.440$; $F = 8.812$; $p < 0.001$). The model parameter estimates are reproduced in *Table 50*. From the model output it can be determined that the onset of basic feedback (version 1) was associated with a significant detrimental increase in the proportion of patients arriving in recovery cold. Holding all other parameters constant, the onset of basic feedback was associated with an increase of 3.3% in the proportion of patients arriving in recovery cold ($p < 0.005$). No significant stepwise effect of enhanced feedback was detected, though the onset of enhanced feedback was associated with a very small ($< 0.1\%$) positive change in slope of the trend line ($p < 0.05$).

Sensitivity analyses

Effect on lower 50th percentile anaesthetist subgroup

In order to investigate the effect of the quality indicator feedback programme on cases performed by the lower 50th percentile of ranked anaesthetists, ITS models were fitted to a time series data set comprising weekly aggregated scores for the lowest 50th percentile ranked anaesthetists. The period used to calculate rankings for the patient temperature data was the full study baseline period. The metric used was the proportion of patients with temperature $< 36^{\circ}\text{C}$ on arrival in recovery. Of the 44 anaesthetists in the pilot group, 30 had sufficient data across the study phases, including the assessment period, to provide rank estimates. The resulting time series sample was, therefore, based on the surgical cases of 15 anaesthetists in the lower 50th percentile. The formal hypothesis for this analysis may be stated as follows in *Table 51*.

TABLE 50 Model parameter estimates for proportion of patients with temperature $< 36^{\circ}\text{C}$, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.043	0.010		4.328	0.000
Baseline trend	0.000	0.001	-0.701	-0.645	0.520
Feedback (basic v1)	0.033	0.012	0.406	2.836	0.005
Trend after feedback v1	-3.355E-05	0.001	-0.061	-0.059	0.953
Feedback (enhanced v3)	-0.006	0.008	-0.107	-0.756	0.451
Trend after feedback v3	0.000	0.000	0.327	2.348	0.020

v, version.

TABLE 51 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup

Main effect under investigation	Hypothesis
H3: effect of performance feedback on lower-ranked anaesthetists	Implementation and escalation of performance feedback will improve quality of care delivered by the subgroup of anaesthetists ranked in the lower half of the distribution of patient temperature performance measures, and any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort. Improved quality of care is detectable by a decrease in the proportion of patients arriving in recovery $< 36^{\circ}\text{C}$

H, hypothesis.

The fitted time series model explained a significant 16.8% of the variance in the dependent measure ($F = 7.5$; $p < 0.001$). Although the overall model fit was significant, none of the beta weight coefficients for the intervention time points reached significance and hypothesis 3 is, therefore, rejected. Compared with the primary analysis of the full pilot group of anaesthetists, however, the significant increase in proportion of patients arriving in recovery cold between the baseline and initial feedback condition was not reproduced in the lower 50th percentile data set. The direction of effect of feedback (version 1) was still positive but to a lesser extent ($\beta = 0.012$; not significant).

Statistical control of case age, gender and American Society of Anesthesiologists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*, above. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

For the mean weekly patient temperature on arrival in recovery measure, the full model including covariates explained 27.4% of the available variance ($R = 0.523$; $F = 7.674$; $p < 0.001$). In the covariate-only model and holding the effects of the other covariates constant, ASA score was positively associated with patient temperature ($\beta = 0.171$; $p < 0.005$) and age was negatively associated ($\beta = -0.008$; $p < 0.001$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, none of the covariates retained a statistically significant unique effect. In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of enhanced feedback being associated with a significant increase in patient temperature ($\beta = 0.062$; $p < 0.005$), but a small detrimental effect on the rate of change in patient temperature over time ($\beta = -0.002$; $p < 0.001$). In the full covariate model, the increase in mean patient temperature following onset of basic feedback did not reach statistical significance.

For the weekly proportion of patients arriving in recovery with temperature below the recommended 36 °C, the full model including covariates explained 23.5% of the available variance ($R = 0.484$; $F = 6.243$; $p < 0.001$). In the covariate-only model and holding the effects of the other covariates constant, age had a small significant positive association with the proportion of patients arriving cold ($\beta = 0.003$; $p < 0.001$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, none of the covariates retained a statistically significant unique effect. In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of basic feedback being associated with a significant increase in proportion of patients arriving cold ($\beta = 0.034$; $p < 0.05$). The onset of enhanced feedback was conversely associated with a small improvement in the rate of change in proportion over time ($\beta < 0.001$; $p < 0.05$).

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project lifecycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for mean patient temperature, the model accounted for 27.9% of the available variance in the dependent measure ($R = 0.528$; $F = 23.988$; $p < 0.001$). Similar to the result from the primary analysis, the onset of feedback was associated with a positive shift in mean patient temperature ($\beta = 0.095$; $p < 0.001$). When the single-stage intervention model was fitted for the proportion of patients with temperature < 36 °C, the model accounted for 16.9% of the available variance in the dependent measure ($R = 0.411$; $F = 12.589$; $p < 0.001$). Similar to the result from the primary analysis, the onset of feedback was associated with a positive shift in proportion of patients with temperature < 36 °C ($\beta = 0.027$; $p < 0.05$).

In terms of the three-stage intervention model, when fitted to mean patient temperature, the model accounted for 37.4% of the available variance ($R = 0.611$; $F = 15.513$; $p < 0.001$). With the roll-in period for enhanced feedback accounted for in the model, both the onset of basic feedback and the onset of the roll-in period were associated with significant stepwise shifts in the dependent variable, the former positively and the second negatively. Enhanced feedback no longer demonstrated a significant positive stepwise effect but the negative change in trend observed in the primary analysis was preserved. For the proportion of patients with temperature $< 36^{\circ}\text{C}$, the three-stage model accounted for 20.5% of the available variance ($R = 0.453$; $F = 6.714$; $p < 0.001$). With the roll-in period for enhanced feedback accounted for in the model, only the positive effect of onset of basic feedback reached significance, in concordance with the primary analysis findings. The negative effect of enhanced feedback was not replicated in the three-stage model.

Variations in case inclusion according to site and admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, two further ITS models were fitted, the first based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site, and the second based on time series data calculated from including all perioperative cases across all study sites within the trust. The former model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained owing to missing administrative data. The second model was based on the broader multisite data set that was initiated during the baseline feedback epoch. Although baseline data exist at the primary study site, no comparable multisite data exist for the secondary sites and, therefore, in the interests of a fair test using a homogenous data set, the baseline period is excluded from the multisite time series models. The methodological rationale and procedural details of these secondary analyses are outlined in full in *Methods*, above.

When the time series model based on elective GA cases was fitted for mean weekly patient temperature, the model explained 30.6% of the available variance ($R = 0.553$; $F = 11.625$; $p < 0.001$). Although the overall model fit was significant, however, in contrast to the primary model results, none of the model parameters reached statistical significance. When the time series model based on elective GA cases was fitted for the proportion of patients with temperature $< 36^{\circ}\text{C}$, the model explained 23.9% of the available variance ($R = 0.489$; $F = 8.313$; $p < 0.001$). In terms of the time series parameters, only the onset of baseline feedback was a significant predictor ($\beta = 0.048$; $p < 0.05$) demonstrating a positive association with the proportion of patients arriving cold and similar to the findings from the primary model. The negative change in slope observed in response to enhanced feedback onset in the primary model was not reproduced in the case-limited model.

The multisite case inclusion model accounted for 16.4% of the available variance in mean patient temperature ($R = 0.405$; $F = 8.114$; $p < 0.001$). In terms of the time series parameters, a significant negative change in slope was observed in response to the onset of enhanced feedback ($\beta = -0.002$; $p < 0.001$), similar to the result from the primary model. The positive effects on level of mean patient temperature in response to enhanced feedback observed in the primary analysis were not replicated in the multisite data set. For the second temperature measure, the multisite case inclusion model accounted for 20.3% of the available variance in the proportion of patients with temperature $< 36^{\circ}\text{C}$ ($R = 0.450$; $F = 10.520$; $p < 0.001$). In terms of the time series parameters, a significant positive change in slope was observed in response to the onset of enhanced feedback ($\beta = 0.001$; $p < 0.05$), similar to the result from the primary model.

Patient-reported quality of recovery

Summary statistics

Over the course of the project and for the pilot anaesthetist group, 8281 valid surgical cases with QoR score data were recorded at St Mary's main recovery suite (representing 36.5% of recorded cases within the research data set). QoR was recorded on a 16-point scale based on an established research-validated scale²⁶ described earlier in *Methods*. The resulting data demonstrated a strong ceiling effect evident from negative skew in the data set. For this reason, a further metric, proportion of patients with a QoR score above 14 (QoRA14), was generated for use in the feedback reports and in the evaluation.

Overall, across the duration of the project, the mean QoR score was 15.01 (median = 16.00; SD = 1.536). Observed QoR scores varied between 0 and 16. The proportion of valid cases with a QoR score above 14 was 77.7% (6438 cases).

Table 52 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, mean QoR score decreased between the baseline and basic feedback conditions, and subsequently increased between the basic and enhanced feedback conditions. In terms of the proportion of patients with QoR score above 14, the proportion followed a similar pattern, decreasing between baseline and first intervention phase, and then increasing between basic and enhanced feedback conditions.

Interrupted time series analysis

The time series charts for both the mean QoR score and the proportion of patients with QoR score above 14, by project week, is depicted in *Figures 12* and *13*, respectively. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

As described previously, the primary ITS model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

TABLE 52 Longitudinal summary of mean QoR score and proportion of patients with QoR score above 14 by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)

Metric	Study epoch		
	Baseline	Feedback v1	Feedback v3
QoR			
Mean	15.195	14.941	15.081
SD	2	2	1
Count	2009	10,255	10,406
QoR score above 14			
Count	782	3634	2022
Proportion	83.37%	76.09%	78.77%

v, version.

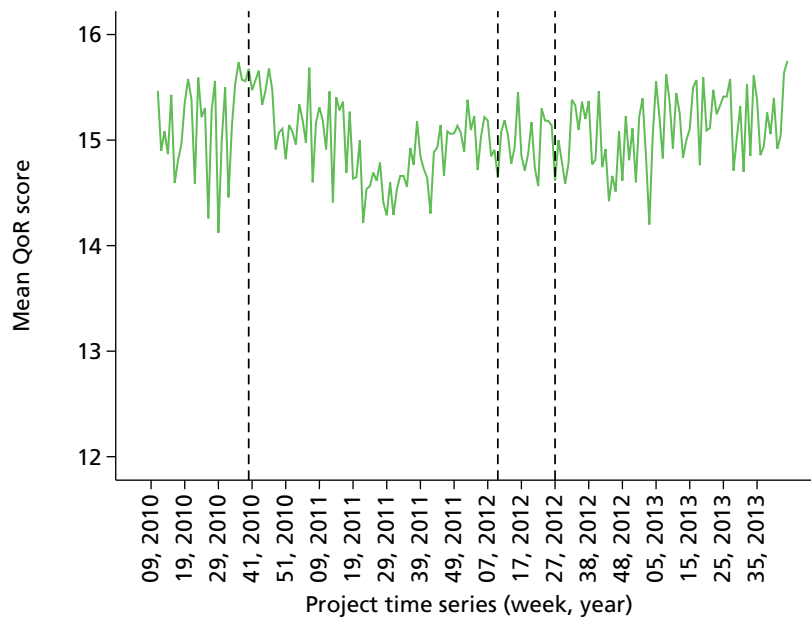


FIGURE 12 Time series for mean QoR Scale score completed prior to discharge from PACU. Each data point represents an average of all QOR data observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

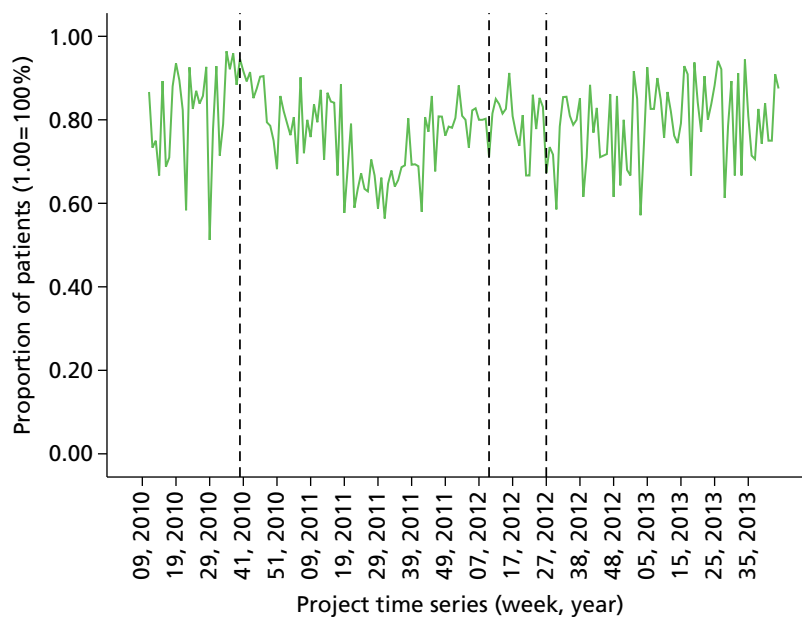


FIGURE 13 Time series for proportion of patients with QoR Scale scores above 14. Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

The association between longitudinal trends in QoR Scale data and the specific feedback interventions was estimated using ITS models as described in the statistical methods section above. For the QoR data, it was possible to evaluate both main evaluative hypotheses as stated in *Table 53* (the formulation of causal theory underpinning these hypotheses is described at length earlier in this section and elsewhere in the report).

In order to evaluate both hypotheses, ITS models were fitted with two interrupts, representing the onset of basic and enhanced feedback, for both mean QoR scores and proportion of QoR scores above 14 as dependent variables. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary's main theatres.

For mean QoR scores, the fitted time series model explained 11.1% of the variance ($R = 0.333$; $F = 4.535$; $p < 0.005$). The model parameter estimates are reproduced in *Table 54*. In the model output, only the post-enhanced feedback slope reached significance. Controlling for all other temporal trends, implementation of enhanced feedback was associated with a small positive increase in the rate of change in mean QoR score over time, trending towards improvement ($\beta = 0.009$; $p < 0.001$).

TABLE 53 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in mean QoR Scale scores 2. increase in the proportion of patients with QoR scores above 14
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in mean QoR Scale scores 2. increase in the proportion of patients with QoR scores above 14

H, hypothesis.

TABLE 54 Model parameter estimates for mean QoR, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	14.959	0.149		100.558	0.000
Baseline trend	0.011	0.008	1.723	1.382	0.169
Feedback (basic v1)	-0.177	0.156	-0.171	-1.134	0.258
Trend after feedback v1	-0.015	0.008	-2.102	-1.760	0.080
Feedback (enhanced v3)	0.070	0.112	0.093	0.624	0.534
Trend after feedback v3	0.009	0.002	0.533	3.632	0.000

v, version.

For the proportion of patients with QoR score above 14, the fitted time series model explained 7.3% of the variance ($R = 0.270$; $F = 2.864$; $p < 0.05$). The model parameter estimates are reproduced in *Table 55*. From the model output it can be determined that the onset of basic feedback (version 1) was associated with a significant decrease in the proportion of patients with QoR scores above 14. Holding all other parameters constant, the onset of basic feedback was associated with a decrease of 9.4% in the proportion of patients with high-quality recovery ($p < 0.05$). No significant stepwise effect of enhanced feedback was detected, though the onset of enhanced feedback was associated with a small positive change in slope of the trend line ($\beta = 0.001$; $p < 0.05$).

Sensitivity analyses

Effect on lower 50th percentile anaesthetist subgroup

In order to investigate the effect of the quality indicator feedback programme on cases performed by the lower 50th percentile of ranked anaesthetists, ITS models were fitted to a time series data set comprising weekly aggregated scores for the lowest 50th percentile ranked anaesthetists. The period used to calculate rankings for the QoR score data was the full study baseline period. The metric used was the proportion of patients with a QoR score above 14. Of the 44 anaesthetists in the pilot group, 30 had sufficient data across the study phases, including the assessment period, to provide rank estimates. The resulting time series sample was, therefore, based on the surgical cases of 15 anaesthetists in the lower 50th percentile. The formal hypothesis for this analysis may be stated as follows in *Table 56*.

The fitted time series model explained 8% of the variance in the dependent measure ($R = 0.284$; $F = 7.5$; $p < 0.01$). Of the time series parameters, the change in slope following implementation of enhanced feedback was a significant predictor of the proportion of QoR scores above 14 ($\beta = 0.002$; $p < 0.01$). In the primary analysis there was an apparent negative effect of the first feedback intervention on the proportion of QoR scores above 14. In the lower 50th percentile group, the apparent negative effect, as observed in the primary analysis, did not reach significance, and the positive change in slope associated with enhanced feedback was slightly larger.

TABLE 55 Model parameter estimates for proportion of patients with QoR scores above 14, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.753	0.041		18.414	0.000
Baseline trend	0.004	0.002	2.312	1.816	0.071
Feedback (basic v1)	-0.094	0.043	-0.335	-2.181	0.030
Trend after feedback v1	-0.004	0.002	-2.381	-1.952	0.053
Feedback (enhanced v3)	0.003	0.031	0.013	0.083	0.934
Trend after feedback v3	0.001	0.001	0.313	2.090	0.038

v, version.

TABLE 56 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup

Main effect under investigation	Hypothesis
H3: effect of performance feedback on lower-ranked anaesthetists	Implementation and escalation of performance feedback will improve quality of care delivered by the subgroup of anaesthetists ranked in the lower half of the distribution of QoR Scale scores and any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort. Improved quality of care is detectable by an increase in the proportion of patients with QoR scores above 14

H, hypothesis.

Statistical control of case age, gender and American Society of Anesthesiologists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*, above. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

For the mean weekly QoR Scale score, the full model including covariates explained 16.6% of the available variance ($R = 0.407$; $F = 4.048$; $p < 0.001$). In the covariate-only model, and holding the effects of the other covariates constant, ASA score was negatively associated with QoR score (beta = -0.816 ; $p < 0.005$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, none of the covariates retained a statistically significant unique effect. In terms of the time series parameters and in contrast to the primary analysis results, the onset of basic feedback was associated with a significant detrimental trend in the rate of change in QoR score (beta = -0.060 ; $p < 0.05$). Similar to the primary model results, the onset of enhanced feedback was associated with improvement in the rate of change of QoR Scale scores (beta = 0.007 ; $p < 0.005$).

For the proportion of patients with QoR scores exceeding 14, the full model including covariates explained 12.4% of the available variance ($R = 0.352$; $F = 2.878$; $p < 0.01$). In the covariate-only model and holding the effects of the other covariates constant, ASA was negatively associated with proportion of high QoR scores (beta = -0.230 ; $p < 0.005$). In the full model, with covariates and time series parameters, holding the effects of all other predictors constant, ASA score retained its significant unique effect (beta = -0.193 ; $p < 0.05$). In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of basic feedback associated with both a significant stepwise reduction in the proportion of patients reporting high QoR scores (beta = -0.126 ; $p < 0.05$) and a detrimental shift in the rate of change in proportion over time (beta = -0.014 ; $p < 0.05$). The positive change in trend following implementation of enhanced feedback observed in the primary model, however, was not reproduced in the covariate model.

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project life cycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for mean QoR score, the model accounted for 3.6% of the available variance in the dependent measure ($R = 0.189$; $F = 2.271$; not significant). In contrast to the primary model, the onset of feedback was associated with a significant negative shift in the mean QoR Scale score (beta = -0.360 ; $p < 0.05$). When the single-stage intervention model was fitted for proportion of patients with QoR scores above 14, the model accounted for 4.9% of the available variance in the dependent measure ($R = 0.221$; $F = 3.164$; $p < 0.05$). Similar to the primary model, the onset of feedback was associated with a significant negative shift in the proportion of patients with high QoR Scale scores (beta = -0.120 ; $p < 0.005$).

In terms of the three-stage intervention model, when fitted to mean QoR score, the model accounted for 13.7% of the available variance ($R = 0.370$; $F = 4.089$; $p < 0.001$). With the roll-in period for enhanced feedback accounted for in the model, only the negative change in trend post implementation of basic feedback reached significance (beta = -0.018 ; $p < 0.05$), in contrast to the results from the primary analysis, which detected a significant positive improvement in slope following implementation of the enhanced feedback. When the three-stage model was fitted to the proportion of patients with a QoR

score exceeding 14, it accounted for 10.9% of the available variance ($R = 0.330$; $F = 3.138$; $p < 0.005$). With the roll-in period for enhanced feedback accounted for in the model, only the negative change in trend post implementation of basic feedback reached significance ($\beta = -0.005$; $p < 0.05$), in contrast to the results from the primary analysis, which detected a significant positive improvement in slope following implementation of the enhanced feedback.

Variations in case inclusion according to admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, a further ITS model was fitted based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site. The model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained owing to missing administrative data. The methodological rationale and procedural details of this secondary analysis are outlined in full in *Methods*.

When the time series model based on elective GA cases was fitted for mean weekly QoR score, the model explained 8.9% of the available variance ($R = 0.298$; $F = 2.572$; $p < 0.05$). In terms of the time series parameters and in contrast to the results from the primary analysis, the onset of basic feedback was associated with a negative effect on both the level ($\beta = -0.510$; $p < 0.05$) and the trend ($\beta = -0.072$; $p < 0.05$) of mean QoR score. Additionally, the positive change in slope, observed in the primary model in response to onset of enhanced feedback, was not reproduced here. When the time series model based on elective GA cases was fitted for the proportion of patients with high QoR scores (above 14), the model explained 6.9% of the available variance ($R = 0.262$; $F = 1.949$; not significant). In terms of the time series parameters and similar to the primary analysis results, the onset of baseline feedback had a significant negative association with the proportion of high QoR scores ($\beta = -0.158$; $p < 0.05$). The significant positive change in slope found in response to enhanced feedback in the primary analysis was not reproduced in the limited case inclusion model.

Postoperative pain

Summary statistics

Postoperative pain was quantified using two metrics for the evaluation of the quality improvement initiative, as described in the methodological description of research measures earlier in this section. The two metrics used in this analysis are:

- QORPain2: freedom from severe pain or constant moderate pain during the recovery period (a subitem from the QoR Scale analysed in the preceding section). Calculated as the proportion of patients reporting being mostly free from severe pain or constant moderate pain (the lowest pain category of a three-category ordinal response scale).
- PainU2: pain scale score on admission to recovery (first conscious pain score). Calculated as the proportion of patients with pain scale scores in the range 0–1 (corresponding to 'none' or 'mild' pain) from a scale ranging from 0 to 4.

Over the course of the project and for the pilot anaesthetist group, 8773 valid surgical cases with QoR pain data were recorded at St Mary's Hospital main recovery suite (representing 38.7% of recorded cases within the research data set). For the second pain measure, 17,738 valid surgical cases with pain-on-arrival data were recorded at St Mary's Hospital main recovery suite (representing 78.2% of recorded cases in the research data set).

Overall, across the duration of the project, the proportion of patients reporting freedom from moderate or severe pain during the recovery period on the QoR pain scale was 66.4% (5829 cases). In terms of the pain-on-arrival scale score, the mean observed value was 0.62 (median = 0; SD = 1.067). The proportion of valid cases with pain on arrival in the range 0–1 was 77.6% (13,769 cases).

Table 57 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, the proportion of patients reporting freedom from severe pain during the recovery period (QoR Scale) decreased between the baseline and basic feedback conditions and subsequently increased between the basic and enhanced feedback conditions. The second measure, pain on arrival in recovery, was introduced concurrently with the basic feedback intervention, meaning that no baseline data exist for this measure. The overall proportion of patients reporting no or mild pain on arrival in recovery increased by 0.47% between the basic and enhanced feedback conditions.

Interrupted time series analysis

The time series charts for both postoperative pain measures, by project week, are included in Figures 14 and 15, respectively. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

As described previously, the primary time series analysis model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

TABLE 57 Longitudinal summary of proportion of patients with QoR pain score of 2 ('mostly free from severe or constant moderate pain' during the recovery period) and with pain on arrival scores in the range 0–1 ('none' or 'mild') by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)

Metrics	Study epoch		
	Baseline	Feedback v1	Feedback v3
QORPain2			
Count	803	3174	1852
Proportion	73.74%	63.51%	68.95%
PainU2			
Count	–	6838	6875
Proportion	–	77.35%	77.82%

v, version.

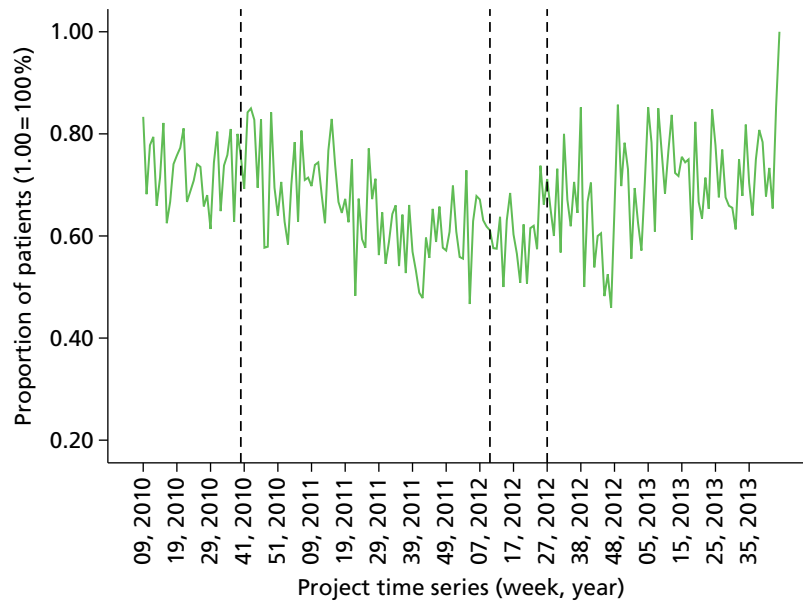


FIGURE 14 Time series for proportion of patients with QoR pain score of 2 ('mostly free from severe or constant moderate pain') during the recovery period. Each data point represents an average of all QoR data observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.



FIGURE 15 Time series for proportion of patients with pain on arrival in recovery score in the range 0-1 ('none' or 'mild'). Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

The association between longitudinal trends in pain data and the specific feedback interventions was estimated using ITS models as described in the statistical methods section above. For the QoR Scale pain data, it was possible to evaluate both main hypotheses as stated below in *Table 58* (the formulation of causal theory underpinning these hypotheses is described at length earlier in this section and elsewhere in the report). For the pain on arrival in recovery measure, because data collection was initiated at the end of the study baseline period (and concurrently with the onset of basic feedback), only the second hypothesis concerning the effects of implementing enhanced feedback over basic feedback could be tested.

In order to evaluate the hypotheses, ITS models were fitted with two interrupts for the QoR pain measure and with one interrupt for the pain on arrival dependent measure. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary’s main theatres.

For the QoR pain data, the fitted time series model explained 28.1% of the variance ($R = 0.530$; $F = 14.352$; $p < 0.001$). The model parameter estimates are reproduced in *Table 59*. In the model output, both temporal parameters for the introduction of enhanced feedback reached statistical significance. Controlling for the effects of all other parameters, implementation of enhanced feedback was associated with an increase in the proportion of patients reporting freedom from severe pain during recovery of 7.2% ($p < 0.01$). Introduction of enhanced feedback was additionally associated with a positive increase in the rate of change in the proportion of patients reporting freedom from severe pain over time, trending towards improvement ($\beta = 0.004$; $p < 0.001$).

TABLE 58 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in increase in proportion of patients reporting freedom from pain during the recovery period
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in proportion of patients reporting freedom from pain during recovery (QoR Scale) 2. increase in the proportion of patients reporting no or mild pain on arrival in recovery

H, hypothesis.

TABLE 59 Model parameter estimates for proportion of patients reporting freedom from pain during recovery (QoR Scale), including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.736	0.031		23.875	0.000
Baseline trend	-0.001	0.002	-0.351	-0.343	0.732
Feedback (basic v1)	0.012	0.036	0.043	0.321	0.749
Trend after feedback v1	-0.001	0.002	-0.659	-0.674	0.501
Feedback (enhanced v3)	0.072	0.027	0.362	2.710	0.007
Trend after feedback v3	0.004	0.001	0.800	6.080	0.000

v, version.

For the proportion of patients reporting no or mild pain on arrival in recovery measure, the fitted time series model explained 49.7% of the variance ($R = 0.705$; $F = 51.766$; $p < 0.001$). The model parameter estimates are reproduced in *Table 60* (note that in this model, the effects of implementation of enhanced feedback only is examined). From the model output, it can be determined that the onset of enhanced feedback (version 3) was associated with significant positive effects on both the level and the slope of the dependent measure. Holding all other parameters constant, the onset of enhanced feedback over basic feedback was associated with an increase of 12% in the proportion of patients reporting no or mild pain on arrival in recovery ($p < 0.001$). The onset of enhanced feedback was additionally associated with a positive increase in the rate of change in the proportion of patients reporting no or mild pain over time, trending towards improvement ($\beta = 0.003$; $p < 0.001$).

Sensitivity analyses

Effect on lower 50th percentile anaesthetist subgroup

In order to investigate the effect of the quality indicator feedback programme on cases performed by the lower 50th percentile of ranked anaesthetists, ITS models were fitted to a time series data set comprising weekly aggregated scores for the lowest 50th percentile ranked anaesthetists. Two metrics were used for this analysis: (1) the proportion of patients reporting freedom from pain during recovery (QoR Scale), and (2) the proportion of patients reporting no or mild pain on arrival in recovery. The period used to calculate rankings for the QoR-based measure was the full study baseline period. Owing to the absence of comparable baseline data, the period used to calculate rankings for the pain on arrival measure was the first feedback intervention stage (version 1 basic feedback). Of the 44 anaesthetists in the pilot group, 30 had sufficient data across the study phases, including the assessment period, to provide rank estimates of the QoR-based measure, and 40 had sufficient data for the pain-on-arrival measure. The resulting time series samples were, therefore, based on the surgical cases of 15 and 20 anaesthetists in the lower 50th percentile, respectively. The formal hypothesis for this analysis may be stated as follows in *Table 61*.

TABLE 60 Model parameter estimates for proportion of patients reporting no or mild pain on arrival in recovery, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback version 3 and change in trend following onset of feedback version 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.925	0.013		72.560	0.000
Baseline trend	-0.003	0.000	-1.680	-12.423	0.000
Feedback (enhanced v3)	0.120	0.019	0.713	6.248	0.000
Trend after feedback v3	0.003	0.000	0.856	7.462	0.000

v, version.

TABLE 61 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup

Main effect under investigation	Hypothesis
H3: effect of performance feedback on lower-ranked anaesthetists	<p>Implementation^a and escalation of performance feedback will improve quality of care delivered by the subgroup of anaesthetists ranked in the lower half of the distribution of postoperative pain measures and any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort. Improved quality of care is detectable by an increase in:</p> <ol style="list-style-type: none"> 1. the proportion of patients reporting freedom from pain during recovery (QoR Scale) 2. the proportion of patients reporting no or mild pain on arrival in recovery

H, hypothesis.

a Note that the implementation hypothesis is not testable for the second measure.

For the QoR Scale pain measure, the fitted time series model explained 8.7% of the variance in the dependent measure ($R = 0.295$; $F = 3.548$; $p < 0.005$). Of the time series parameters, the change in slope following implementation of enhanced feedback was a significant predictor of the proportion of patients reporting freedom from severe postoperative pain ($\beta = 0.003$; $p < 0.005$). In the primary analysis there was a significant positive effect of enhanced feedback on freedom from postoperative pain. In the lower 50th percentile group, the enhanced feedback iteration did not reach significance in terms of stepwise change in the dependent measure, but the improvement in slope following the second iteration was significant.

For the proportion of patients reporting no or mild pain on arrival in recovery, the fitted time series model explained 44.5% of the variance ($R = 0.667$; $F = 42.182$; $p < 0.001$). In the lower 50th percentile group, the model parameter effects echoed those of the findings of the primary analysis with the implementation of enhanced feedback being associated with a significant change in both level ($\beta = 0.138$; $p < 0.001$) and slope ($\beta = 0.004$; $p < 0.001$) of the dependent measure. The beta coefficients in both cases were larger in the lower 50th percentile group.

Statistical control of case age, gender and American Association of Anaesthetists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*, above. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

In analysis of the proportion of patients reporting freedom from severe pain (QoR Scale item), the full model including covariates explained 30.9% of the available variance ($R = 0.556$; $F = 9.114$; $p < 0.001$). In the covariate-only model, and holding the effects of the other covariates constant, ASA rating was negatively associated with proportion of patients free from severe pain ($\beta = -0.294$; $p < 0.001$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, ASA score retained a statistically significant unique effect ($\beta = -0.177$; $p < 0.01$). In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of enhanced feedback being associated with a significant improvement both in the level of the proportion of patients reporting severe pain ($\beta = 0.064$; $p < 0.05$) and in the trend ($\beta = 0.003$; $p < 0.001$).

For the proportion of patients reporting no or mild pain on arrival in recovery, the full model including covariates explained 51.4% of the available variance ($R = 0.717$; $F = 27.127$; $p < 0.001$). In the covariate-only model, and holding the effects of the other covariates constant, ASA score was negatively associated with pain on arrival ($\beta = -0.280$; $p < 0.001$) and age was positively associated ($\beta = 0.007$; $p < 0.01$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, none of the covariates retained a statistically significant unique effect. In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of enhanced feedback being associated with a significant stepwise increase in the proportion of patients reporting no or mild pain ($\beta = 0.113$; $p < 0.001$) and a small improvement in the rate of change over time ($\beta = 0.003$; $p < 0.001$).

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project lifecycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for the proportion of patients reporting freedom from postoperative pain, the model accounted for 6.8% of the available variance in the dependent measure ($R = 0.260$; $F = 4.510$; $p < 0.01$). Similar to the result from the primary analysis, the onset of feedback yielded no significant time series parameters.

In terms of the three-stage intervention model, when fitted for the proportion of patients reporting freedom from postoperative pain, the model accounted for 29.4% of the available variance ($R = 0.542$; $F = 10.832$; $p < 0.001$). With the roll-in period for enhanced feedback accounted for in the model, none of the time series parameters reached significance. The positive stepwise effect of enhanced feedback observed in the primary analysis was not replicated in the three-stage model.

When the model including roll-in period was fitted to the proportion of patients reporting no or mild pain on arrival in recovery, 55.5% of the variance was accounted for ($R = 0.745$; $F = 38.676$; $p < 0.001$). In the time series parameters, both the onset of rapid iteration and the onset of the enhanced feedback period were associated with a significant positive stepwise increase in the level of the proportion of patients reporting no or mild pain, replicating the observed effect of enhanced feedback on level in the primary analytic model.

Variations in case inclusion according to site and admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, two further ITS models were fitted, the first based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site, and the second based on time series data calculated from including all perioperative cases across all study sites within the trust. The former model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained due to missing administrative data. The second model was based on the broader multisite data set that was initiated during the baseline feedback epoch. Although baseline data exist at the primary study site, no comparable multisite data exist for the secondary sites and, therefore, in the interests of a fair test using a homogenous data set, the baseline period is excluded from the multisite time series models. The methodological rationale and procedural details of these secondary analyses are outlined in full in *Methods*, above.

When the time series model based on elective GA cases was fitted for the proportion of patients reporting freedom from severe pain (QoR Scale), the model explained 13.8% of the available variance ($R = 0.371$; $F = 4.222$; $p < 0.005$). In terms of the time series parameters and similar to the results from the primary analysis, the onset of enhanced feedback was associated with a positive effect on the level of proportion of patients free from pain ($\beta = 0.098$; $p < 0.05$). The positive change in slope observed in the primary model in response to onset of enhanced feedback was not reproduced here. When the time series model based on elective GA cases was fitted for the proportion of patients reporting no or mild pain on arrival in recovery, the model explained 39.3% of the available variance ($R = 0.627$; $F = 26.120$; $p < 0.001$). In terms of the time series parameters and similar to the primary analysis results, the onset of enhanced feedback had a significant positive association with freedom from pain, resulting in both a positive change in level ($\beta = 0.126$; $p < 0.001$) and slope ($\beta = 0.004$; $p < 0.001$).

The multisite case inclusion model accounted for 33.7% of the available variance in the proportion of patients reporting no or mild pain on arrival in recovery ($R = 0.581$; $F = 21.053$; $p < 0.001$). In terms of the time series parameters, similar to the results from the primary analysis, the onset of enhanced feedback was associated with both a positive effect on level ($\beta = 0.052$; $p < 0.001$) and slope ($\beta = 0.001$; $p < 0.001$) of the dependent measure.

Postoperative nausea

Summary statistics

Postoperative nausea was quantified using two metrics for the evaluation of the quality improvement initiative, as described in the methodological description of research measures earlier in this section. The two metrics used in this analysis are:

- QORPONV2: freedom from PONV during the recovery period (a subitem from the QoR Scale analysed in *Patient-reported quality of recovery*). Calculated as the proportion of patients reporting being mostly free from nausea, dry retching or vomiting (the lowest nausea category of a three-category ordinal response scale).
- Nausea0: 3-point nausea scale score completed by the PACU nurse based on experience of the patient’s full stay in recovery. Calculated as the proportion of patients scoring 0 (total absence of nausea).

Over the course of the project and for the pilot anaesthetist group, 8783 valid surgical cases with QoR PONV data were recorded at St Mary’s Hospital main recovery suite (representing 38.7% of recorded cases within the research data set). For the second nausea measure, 17,659 valid surgical cases were recorded at St Mary’s Hospital main recovery suite (representing 77.9% of recorded cases within the research data set).

Overall, across the duration of the project, the proportion of patients reporting freedom from PONV during the recovery period on the QoR PONV scale was 91.3% (8018 cases). In terms of the nausea scale score, the proportion of valid cases reporting total absence of nausea was 94.1% (16,620 cases).

Table 62 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, the proportion of patients reporting freedom from PONV during the recovery period (QoR Scale) decreased between the baseline and basic feedback condition and subsequently increased between the basic and enhanced feedback condition. The second nausea scale measure was introduced concurrently with the basic feedback intervention, meaning that no baseline data exist for this measure. The overall proportion of patients with total absence of nausea during recovery according to this measure decreased by 0.09% between the basic and enhanced feedback conditions.

TABLE 62 Longitudinal summary of proportion of patients with QoR PONV score of 2 (‘mostly free from nausea, dry retching or vomiting’ during the recovery period) and with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period) by study epoch (data based on pilot anaesthetist group and all cases at St Mary’s main theatres)

Metrics	Study epoch		
	Baseline	Feedback v1	Feedback v3
QORPONV2			
Count	996	4534	2488
Proportion	91.63%	90.68%	92.28%
Nausea0			
Count	–	8336	8223
Proportion	–	94.30%	93.91%

v, version.

Interrupted time series analysis

The time series charts for both postoperative nausea measures, by project week, are included in *Figures 16* and *17*, respectively. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

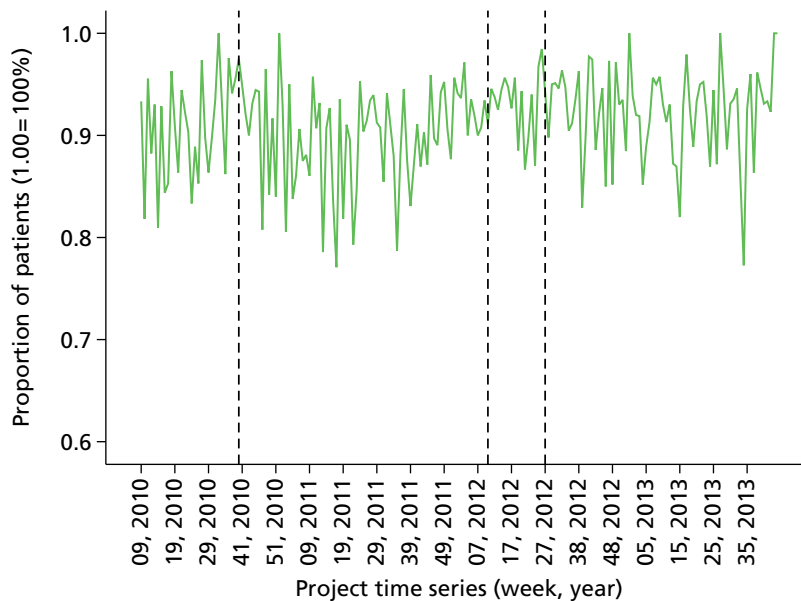


FIGURE 16 Time series for proportion of patients with QoR PONV score of 2 ('mostly free from nausea, dry retching or vomiting' during the recovery period). Each data point represents an average of all QoR data observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

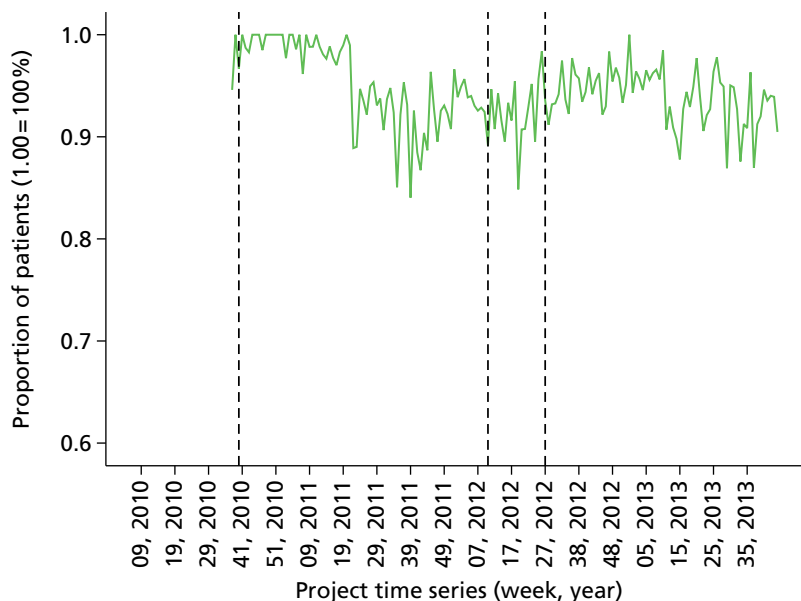


FIGURE 17 Time series for proportion of patients with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period). Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

As described previously, the primary ITS model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

The association between longitudinal trends in nausea data and the specific feedback interventions was estimated using ITS models as described in the statistical methods section above. For the QoR Scale PONV data, it was possible to evaluate both main hypotheses as stated below in *Table 63* (the formulation of causal theory underpinning these hypotheses is described at length earlier in this section and elsewhere in the report). For the nurse-reported nausea during recovery measure, because data collection was initiated at the end of the study baseline period (and concurrently with the onset of basic feedback), only the second hypothesis concerning the effects of implementing enhanced feedback over basic feedback could be tested.

In order to evaluate the hypotheses, ITS models were fitted with two interrupts for the QoR PONV measure and with one interrupt for the nurse-reported nausea-dependent measure. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary’s Hospital main theatres.

For the QoR nausea data, the fitted time series model explained 7.4% of the variance ($R = 0.272$; $F = 2.943$; $p < 0.05$). The model parameter estimates are reproduced in *Table 64*. In the model output, only the parameter for the onset of basic feedback reached statistical significance. Controlling for the effects of all other parameters, implementation of basic feedback was associated with a decrease in the proportion of patients reporting freedom from nausea during recovery of 4.6% ($p < 0.05$). Introduction of enhanced feedback had no determinable effect.

For the proportion of patients with nurse-recorded nausea scores of 0, the fitted time series model explained 38.1% of the variance ($R = 0.617$; $F = 32.196$; $p < 0.001$). The model parameter estimates are reproduced in *Table 65* (note that in this model, the effects of implementation of enhanced feedback only is examined). From the model output, it can be determined that the onset of enhanced feedback (version 3) was associated with significant positive effects on both the level and the slope of the dependent measure. Holding all other parameters constant, the onset of enhanced feedback over basic feedback was associated with an increase of 5.8% in the proportion of patients free from nausea during recovery ($p < 0.001$). The onset of enhanced feedback was additionally associated with a positive increase in the rate of change in the proportion of patients reporting no or mild pain over time, trending towards improvement ($\beta = 0.001$; $p < 0.005$).

TABLE 63 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in increase in proportion of patients reporting freedom from postoperative nausea during the recovery period
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in: <ol style="list-style-type: none"> 1. increase in proportion of patients reporting freedom from postoperative nausea during the recovery period. (QoR Scale) 2. increase in the proportion of patients for whom a nausea score of 0 has been recorded during the recovery period by PACU nurses

H, hypothesis.

TABLE 64 Model parameter estimates for proportion of patients with QoR PONV score of 2 ('mostly free from nausea, dry retching or vomiting'), including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.878	0.017		50.340	0.000
Baseline trend	0.002	0.001	2.124	1.826	0.070
Feedback (basic v1)	-0.046	0.020	-0.343	-2.238	0.026
Trend after feedback v1	-0.001	0.001	-1.613	-1.452	0.148
Feedback (enhanced v3)	0.008	0.015	0.082	0.539	0.591
Trend after feedback v3	0.000	0.000	-0.203	-1.363	0.175

v, version.

TABLE 65 Model parameter estimates for proportion of patients with PACU nurse-completed nausea scores of 0 (representing total absence of nausea during the recovery period), including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback version 3 and change in trend following onset of feedback version 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.996	0.006		164.666	0.000
Baseline trend	-0.001	0.000	-1.395	-9.292	0.000
Feedback (enhanced v3)	0.058	0.009	0.802	6.328	0.000
Trend after feedback v3	0.001	0.000	0.379	2.975	0.003

v, version.

Sensitivity analyses

Effect on lower 50th percentile anaesthetist subgroup

In order to investigate the effect of the quality indicator feedback programme on cases performed by the lower 50th percentile of ranked anaesthetists, ITS models were fitted to a time series data set comprising weekly aggregated scores for the lowest 50th percentile ranked anaesthetists. Two metrics were used for this analysis: (1) the proportion of patients reporting freedom from postoperative nausea during the recovery period (QoR Scale), and (2) the proportion of patients for whom a nausea score of 0 had been recorded during the recovery period by PACU nurses. The period used to calculate rankings for the QoR-based measure was the full study baseline period. Owing to the absence of comparable baseline data, the period used to calculate rankings for the nurse-recorded nausea measure was the first feedback intervention stage (version 1 basic feedback). Of the 44 anaesthetists in the pilot group, 30 had sufficient data across the study phases, including the assessment period, to provide rank estimates of the QoR-based measure and 40 had sufficient data for the nurse-recorded nausea measure. The resulting time series samples were therefore based on the surgical cases of 15 and 20 anaesthetists in the lower 50th percentile, respectively. The formal hypothesis for this analysis may be stated as follows in *Table 66*.

TABLE 66 Hypothesis formulated for effect on lower 50th percentile anaesthetist subgroup

Main effect under investigation	Hypothesis
H3: effect of performance feedback on lower-ranked anaesthetists	<p>Implementation^a and escalation of performance feedback will improve quality of care delivered by the subgroup of anaesthetists ranked in the lower half of the distribution of postoperative nausea measures, and any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort. Improved quality of care is detectable by an increase in:</p> <ol style="list-style-type: none"> 1. proportion of patients reporting freedom from postoperative nausea during the recovery period (QoR Scale) 2. proportion of patients for whom a nausea score of 0 had been recorded during the recovery period by PACU nurses

H, hypothesis.

a Note that the implementation hypothesis is not testable for the second measure.

For the QoR PONV measure, the fitted time series model explained 2% of the variance in the dependent measure ($R = 0.246$; $F = 2.407$; $p < 0.05$). Of the time series parameters, the change in slope following implementation of basic feedback (intervention 1) was a significant predictor of the proportion of patients reporting freedom from nausea and vomiting ($\beta = -0.003$; $p < 0.05$), a negative effect. The negative effect on the level of the QoR PONV measure detected by the primary analysis was not evident in analysis of the lower 50th percentile group data.

For the proportion of patients with a nurse-reported nausea score of 0, the fitted time series model explained 25.8% of the variance ($R = 0.508$; $F = 18.238$; $p < 0.001$). In the lower 50th percentile group, the model parameter effects echoed those of the findings of the primary analysis, with the implementation of enhanced feedback being associated with both a significant change in level ($\beta = 0.075$; $p < 0.001$) and slope ($\beta = 0.001$; $p < 0.001$) of the dependent measure. The beta coefficient for the level change in the dependent measure was larger in the lower 50th percentile group.

Statistical control of case age, gender and American Association of Anaesthetists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*, above. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

For the proportion of patients reporting freedom from PONV (QoR Scale item), the full model including covariates explained 7.7% of the available variance ($R = 0.278$; $F = 1.706$; not significant). In the covariate-only model, none of the covariates reached significance. In the full model with covariates and time series parameters, holding the effects of all other predictors constant, the onset of basic feedback was associated with a significant negative association with freedom from postoperative nausea ($\beta = -0.056$; $p < 0.05$), replicating the result from the primary model.

For the proportion of patients scoring zero on the nurse-reported postoperative nausea measure, the full model including covariates explained 38.4% of the available variance ($R = 0.620$; $F = 16.016$; $p < 0.001$). In the covariate-only model and holding the effects of the other covariates constant, ASA score was negatively associated with freedom from postoperative nausea ($\beta = -0.077$; $p < 0.01$) and age was positively associated ($\beta = 0.003$; $p < 0.01$). In the full model with covariates and time series parameters, holding the effects of all other predictors constant, none of the covariates retained a statistically significant unique effect. In terms of the time series parameters, a similar pattern of intervention effects was observed as for the primary analysis, with the onset of enhanced feedback being associated with a significant stepwise increase in proportion of patients rated without nausea ($\beta = 0.057$; $p < 0.001$) and a small improvement in the rate of change in the dependent measure over time ($\beta = 0.001$; $p < 0.05$).

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project lifecycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for the proportion of patients reporting freedom from postoperative nausea, the model accounted for 6.4% of the available variance in the dependent measure ($R = 0.254$; $F = 4.269$; $p < 0.01$). Similar to the result from the primary analysis, the onset of feedback was associated with a negative shift in the proportion of patients free from nausea ($\beta = -0.040$; $p < 0.05$).

In terms of the three-stage intervention model, when fitted for the proportion of patients reporting freedom from postoperative nausea, the model accounted for 8.5% of the available variance ($R = 0.291$; $F = 2.400$; $p < 0.05$). With the roll-in period for enhanced feedback accounted for in the model, none of the time series parameters reached significance. The negative stepwise effect of basic feedback observed in the primary analysis was not replicated in the three-stage model.

When the model including roll-in period was fitted to the proportion of patients with nurse-reported absence of nausea, 40.9% of the variance was accounted for ($R = 0.640$; $F = 21.465$; $p < 0.001$). In the time series parameters, the onset of the enhanced feedback period was associated with a significant positive stepwise increase in the level of the proportion of patients with nurse-reported freedom from nausea, replicating the observed effect of enhanced feedback on level in the primary analytic model.

Variations in case inclusion according to site and admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, two further ITS models were fitted, the first based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site, and the second based on time series data calculated from including all perioperative cases across all study sites within the trust. The former model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained owing to missing administrative data. The second model was based on the broader multisite data set that was initiated during the baseline feedback epoch. Although baseline data exist at the primary study site, no comparable multisite data exist for the secondary sites and, therefore, in the interests of a fair test using a homogenous data set, the baseline period is excluded from the multisite time series models. The methodological rationale and procedural details of these secondary analyses are outlined in full in *Methods*, above.

When the time series model based on elective GA cases was fitted for the proportion of patients reporting freedom from PONV (QoR Scale), the model explained 6% of the available variance ($R = 0.245$; $F = 1.687$; not significant). In terms of the time series parameters and in contrast to the results from the primary analysis, the onset of enhanced feedback was associated with a negative change in slope of the proportion of patients free from nausea ($\beta = -0.002$; $p < 0.05$). The negative change in level associated with the onset of basic feedback in the primary model was not reproduced in the case-limited model.

When the time series model based on elective GA cases was fitted for proportion of patients with nurse reported absence of postoperative nausea, the model explained 30.7% of the available variance ($R = 0.554$; $F = 17.881$; $p < 0.001$). In terms of the time series parameters and similar to the primary analysis results, the onset of enhanced feedback had a significant positive association with freedom from nausea, resulting in a positive change in both level ($\beta = 0.052$; $p < 0.005$) and slope ($\beta = 0.002$; $p < 0.05$).

The multisite case inclusion model accounted for 6.3% of the available variance in the proportion of patients with nurse-reported freedom from nausea ($R = 0.251$; $F = 2.787$; $p < 0.05$). In terms of the time series parameters and in contrast to the results from the primary analysis, the model yielded no significant associations. The significant positive effects on slope and level observed in response to the onset of enhanced feedback in the primary model were not replicated in the multisite data set.

Surgical site infection rate

Summary statistics

Over the course of the project and for the pilot anaesthetist group, 21,730 valid surgical cases for which SSI data could be generated based on *International Classification of Diseases, Tenth Edition (ICD-10)* diagnostic codes in the hospital administrative record were accumulated, for patients who underwent surgery at St Mary’s main theatres during the course of the quality improvement initiative (representing 95.9% of recorded cases within the research data set). The proportion of patients with a SSI was calculated based on the presence of one or more SSI-specific diagnostic codes in the patient’s administrative record as described earlier in *Methods*.

Overall, across the duration of the project, the proportion of valid cases registering positive for a SSI diagnosis was 1.2% (268 cases).

Table 67 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, the proportion of patients with a SSI diagnosis increased by 0.05% between the baseline and first feedback condition, and then decreased by 0.45% between the basic and enhanced feedback conditions.

Interrupted time series analysis

The time series chart for the proportion of patients with a SSI-specific diagnostic code, by project week, is depicted in Figure 18. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

As described previously, the primary ITS model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

TABLE 67 Longitudinal summary of proportion of patients with a SSI-specific diagnostic code by study epoch (data based on pilot anaesthetist group and all cases at St Mary’s main theatres)

Metrics	Study epoch		
	Baseline	Feedback v1	Feedback v3
SSIDiagAll			
Count	15	149	104
Proportion	1.40%	1.45%	1.00%

v, version.

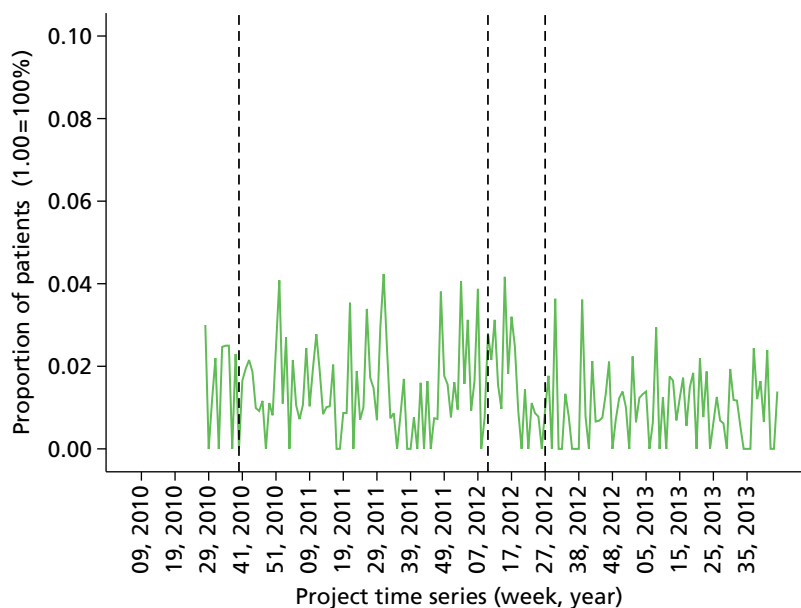


FIGURE 18 Time series for the proportion of patients with a SSI-specific diagnostic code. Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

The association between longitudinal trends in SSI rates and the specific feedback interventions was estimated using ITS models (the formulation of causal theory underpinning these hypotheses is described in more detail in *Methods, Research aims*). For the SSI data, it was possible to evaluate both main evaluative hypotheses, with some limitations. Owing to the lack of administrative data at the start of the project timeline in the first 20 weeks of the baseline period, no mortality data could be generated, meaning that the baseline period is significantly truncated. The research hypotheses are stated in *Table 68*.

In order to evaluate both hypotheses, ITS models were fitted with two interrupts, representing the onset of basic and enhanced feedback, for the proportion of patients with a SSI-specific diagnostic code as dependent variable. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary's main theatres.

For SSI rate, the fitted time series model explained a small proportion of the variance (5.6%) and the segmented regression model did not fit the observed data well, as indicated by the non-significant *F*-statistic ($R = 0.238$; $F = 1.973$; not significant). The model parameter estimates are reproduced in *Table 69*. None of the fitted parameters reached statistical significance.

TABLE 68 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in a reduction in the proportion of patients with a SSI-specific diagnostic code
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in a reduction in the proportion of patients with a SSI-specific diagnostic code

H, hypothesis.

TABLE 69 Model parameter estimates for the proportion of patients with a SSI-specific diagnostic code, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	0.010	0.023		0.430	0.668
Baseline trend	0.000	0.001	1.195	0.283	0.777
Feedback (basic v1)	-0.004	0.006	-0.091	-0.641	0.522
Trend after feedback v1	0.000	0.001	-1.057	-0.253	0.801
Feedback (enhanced v3)	-0.006	0.003	-0.281	-1.799	0.074
Trend after feedback v3	-2.552×10^{-5}	0.000	-0.054	-0.347	0.729

v, version.

Sensitivity analyses

Statistical control of case age, gender and American Association of Anaesthetists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*, above. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

For the proportion of patients with one or more SSI-specific diagnostic codes, the full model including covariates explained 6.3% of the available variance ($R = 0.251$; $F = 1.362$; not significant). In the covariate-only model, none of the covariates reached statistical significance. In the full model with covariates and time series parameters, again none of the model parameters reached statistical significance, replicating the results of the primary analytic model.

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project lifecycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for SSI rate, the model accounted for 3.5% of the available variance in the dependent measure ($R = 0.186$; $F = 1.990$; not significant). Similar to the primary model, the onset of feedback was associated with no significant time series parameters.

When the three-stage model was fitted to SSI rate, it accounted for 8.3% of the available variance ($R = 0.289$; $F = 2.119$; $p < 0.05$). With the roll-in period for enhanced feedback accounted for in the model, and in contrast to the results from the primary analysis, the three-stage model yielded a small but significant negative change in the trend post-roll-in period and a small but significant positive shift in the trend post enhanced feedback onset.

Variations in case inclusion according to site and admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, two further ITS models were fitted, the first based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site, and the second based on time series data calculated from including all perioperative cases across all study sites within the trust. The former model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained owing to missing administrative data. The second model was based on the broader multisite data set that was initiated during the baseline feedback epoch. Although baseline data exist at the primary study site, no comparable multisite data exist for the secondary sites and, therefore, in the interests of a fair test using a homogenous data set, the baseline period is excluded from the multisite time series models. The methodological rationale and procedural details of these secondary analyses are outlined in full in *Methods*, above.

When the time series model based on elective GA cases was fitted for SSI rate, the model explained 4.4% of the available variance ($R = 0.210$; $F = 1.214$; not significant). In terms of the time series parameters and in contrast to the results from the primary analysis, the onset of enhanced feedback was associated with a negative change in slope of SSI rate ($\beta = -0.001$; $p < 0.05$).

The multisite case inclusion model accounted for 7.9% of the available variance in SSI rate ($R = 0.280$; $F = 3.530$; $p < 0.05$). In terms of the time series parameters, similar to the results from the primary analysis, none of the parameters reached statistical significance.

Thirty-day mortality following surgery

Summary statistics

Over the course of the project and for the pilot anaesthetist group, 14,962 valid surgical cases for which 30-day mortality data could be generated based on the hospital administrative record were accumulated for patients who underwent surgery at St Mary's Hospital main theatres during the course of the quality improvement initiative (representing 66% of recorded cases within the research data set). The proportion of patients with 30-day mortality was calculated based on administrative records of days until recorded mortality as described earlier in *Methods*.

Overall, across the duration of the project, the proportion of valid cases registering positive for 30-day mortality was 0.8% (118 cases).

Table 70 provides a summary of the longitudinal data by study epoch and intervention phase [baseline, simple feedback (version 1) and enhanced feedback (version 3)]. Within the three main phases of the study, the proportion of patients with 30-day mortality increased by 0.32% between the baseline and first feedback conditions, and increased again by 0.29% between the basic and enhanced feedback conditions.

TABLE 70 Longitudinal summary of proportion of patients with 30-day mortality following surgery by study epoch (data based on pilot anaesthetist group and all cases at St Mary's main theatres)

Metrics	Study epoch		
	Baseline	Feedback v1	Feedback v3
Mort30Day			
Count	3	48	67
Proportion	0.35%	0.67%	0.96%

v, version.

Interrupted time series analysis

The time series chart for the proportion of patients with 30-day mortality following surgery, by project week, is depicted in *Figure 19*. The four possible intervention phases are delineated within the time series by three interrupts, representing:

1. onset of basic feedback (version 1)
2. onset of rapid development of feedback format (version 2 – a minor phase representing an unstable, iterative version of the feedback that was designed to pilot features of the enhanced feedback model)
3. onset of enhanced feedback model (version 3).

As described previously, the primary ITS model fitted to evaluate the feedback intervention focuses on the implementation of basic and enhanced feedback (i.e. versions 1 and 3), effectively dividing the project timeline into three main study epochs delineated by the first and third interrupts in the time series. The more complex four-phase model was investigated in subsequent sensitivity analyses.

The association between longitudinal trends in mortality rates and the specific feedback interventions was estimated using ITS models (the formulation of causal theory underpinning these hypotheses is described in more detail in *Methods, Research aims*). For the mortality data, it was possible to evaluate both main evaluative hypotheses, with some limitations. Owing to the lack of administrative data at the start of the project timeline in the first 20 weeks of the baseline period, no mortality data could be generated, meaning that the baseline period is significantly truncated. The research hypotheses are stated in *Table 71*.

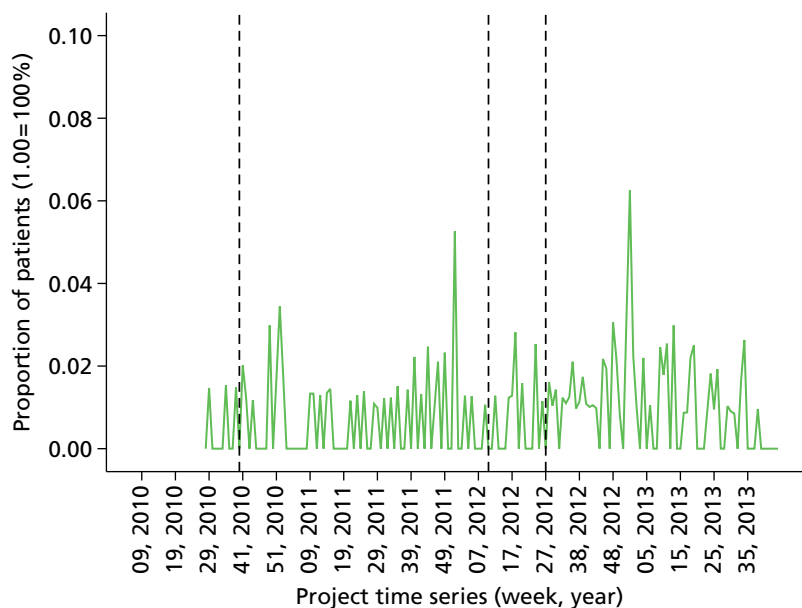


FIGURE 19 Time series for the proportion of patients with 30-day mortality following surgery. Each data point represents the proportion of patients observed in each successive week of the project. The vertical dashed lines represent interrupts in the timeline associated with iterations of the data feedback intervention.

TABLE 71 Main evaluative hypotheses

Main effect under investigation	Hypothesis
H1: effect of introduction of basic feedback	Implementation of basic feedback from anaesthetic quality indicators at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in a reduction in the proportion of patients experiencing 30-day mortality following surgery
H2: effect of enhancing basic feedback	Implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve quality of anaesthetic care, resulting in a reduction in the proportion of patients experiencing 30-day mortality following surgery

H, hypothesis.

In order to evaluate both hypotheses, ITS models were fitted with two interrupts, representing the onset of basic and enhanced feedback, for the proportion of patients with 30-day mortality as dependent variable. Time series data for the analysis were based on the weekly aggregated data from all surgical cases performed by the pilot anaesthetist group at St Mary's Hospital main theatres.

For mortality rate, the fitted time series model explained 6.6% of the variance ($R = 0.257$; $F = 2.328$; $p < 0.05$). The model parameter estimates are reproduced in *Table 72*. Holding all other temporal intervention effects constant, the onset of enhanced feedback was associated with both a stepwise increase of 0.8% in mortality ($p < 0.05$) and a very small negative change in slope, trending towards improvement ($\text{beta} < -0.001$; $p < 0.05$).

TABLE 72 Model parameter estimates for the proportion of patients with 30-day mortality following surgery, including beta weight coefficients and associated significance of *t*-statistics. Time series model parameters include baseline trend, onset of feedback intervention versions 1 and 3, and change in trend following onset of interventions versions 1 and 3

Model parameters	Unstandardised coefficients		Standardised coefficients		
	Beta	Standard error	Beta	<i>t</i>	Significance
Constant (intercept)	-0.007	0.022		-0.309	0.758
Baseline trend	0.000	0.001	2.160	0.514	0.608
Feedback (basic v1)	0.000	0.006	-0.008	-0.055	0.956
Trend after feedback v1	0.000	0.001	-2.104	-0.506	0.614
Feedback (enhanced v3)	0.008	0.003	0.387	2.493	0.014
Trend after feedback v3	0.000	0.000	-0.312	-2.015	0.046

v, version.

Sensitivity analyses

Statistical control of case age, gender and American Association of Anaesthetists score

In order to investigate the robustness of the findings from the primary ITSA of the impact of the feedback intervention on perioperative process and outcome parameters, sensitivity analysis was performed to adjust for patient age, gender and ASA score, as detailed in *Methods*. Mean weekly patient age, weekly percentage of male patients and mean weekly ASA score metrics were calculated and entered into the regression model as covariates, prior to entry of intervention time series parameters.

For the proportion of patients registering positive for 30-day postoperative mortality, the full model including covariates explained 9.5% of the available variance ($R = 0.308$; $F = 2.119$; $p < 0.05$). In the covariate-only model, ASA score was significantly positively associated with mortality (beta = 0.021; $p < 0.05$) and age negatively associated (beta = -0.001; $p < 0.021$). In the full model with covariates and time series parameters, ASA score retained statistical significance (beta = 0.019; $p < 0.05$). The onset of enhanced feedback was associated with a significant stepwise increase in 30-day mortality (beta = 0.009; $p < 0.05$), replicating the result from the primary analysis. The significant improvement in slope associated with enhanced feedback in the primary analytic model for mortality was not replicated in the covariate model.

Intervention timeline variants

As detailed in the methods, an important aspect of the sensitivity analysis undertaken was to explore the fit of two variant project timeline models suggested by analysis of the development of the initiative over the course of the project lifecycle. The first variant incorporated a simpler, single-stage intervention model to examine the effect of simply introducing feedback to the anaesthetist cohort. The second variant was a more complex three-stage model that isolated and controlled for the period immediately preceding the onset of enhanced feedback, in which the feedback underwent a period of rapid iterative development.

When the single-stage intervention model was fitted for mortality rate, the model accounted for 1.8% of the available variance in the dependent measure ($R = 0.134$; $F = 1.025$; not significant). Similar to the primary model, the onset of feedback was associated with no significant time series parameters.

When the three-stage model was fitted to mortality rate, it accounted for 6.7% of the available variance ($R = 0.259$; $F = 1.676$; not significant). With the roll-in period for enhanced feedback accounted for in the model, none of the time series parameters reached statistical significance.

Variations in case inclusion according to site and admission/anaesthetic type

In order to investigate whether or not the observed results from the primary analysis were robust to variations in case inclusion criteria, two further ITS models were fitted, the first based on time series data calculated from including only 'elective, general anaesthetic' cases at the primary study site, and the second based on time series data calculated from including all perioperative cases across all study sites within the trust. The former model necessitated analysis of a truncated intervention timeline in which both the baseline and enhanced feedback epochs were constrained owing to missing administrative data. The second model was based on the broader multisite data set that was initiated during the baseline feedback epoch. Although baseline data exist at the primary study site, no comparable multisite data exist for the secondary sites and, therefore, in the interests of a fair test using a homogenous data set, the baseline period is excluded from the multisite time series models. The methodological rationale and procedural details of these secondary analyses are outlined in full in *Methods*, above.

When the time series model based on elective GA cases was fitted for mortality rate, the model explained 0.9% of the available variance ($R = 0.096$; $F = 0.248$; not significant). In terms of the time series parameters, and in contrast to the results from the primary analysis, no significant effects were observed.

The multisite case inclusion model accounted for 2.5% of the available variance in mortality rate ($R = 0.157$; $F = 1.042$; not significant). In terms of the time series parameters, in contrast to the results from the primary analysis, no parameters reached statistical significance. The changes in slope and level identified by the primary analysis in response to the onset of enhanced feedback were not replicated in the multisite data set.

Discussion

In this section of the report, the findings from the primary and secondary analysis of the perioperative data collected in the quasi-experimental substudy will be summarised and their implications for the evaluative hypotheses considered.

Summary of primary findings against hypotheses

The primary analysis addressed two evaluative hypotheses (1 and 2) concerning the impact of basic and enhanced feedback, respectively, on a number of anaesthetic and perioperative quality indicators recorded for all surgical patients passing through the main recovery suite at the primary intervention site, St Mary's Hospital. A number of ITS models were fitted to the data, using segmented regression, to test the main hypotheses. In the primary ITS models, the feedback intervention was represented as a serial, two-stage intervention. This meant that hypotheses 1 and 2 were operationalised, for research purposes, as the observed effects of implementing basic feedback within a feedback-naive cohort and then escalating the intensity of that feedback (the 'enhanced' feedback condition) while controlling for prior longitudinal trends in the data at each stage. The ability to model serial effects of an intervention over time was essential in order to account for the iterative and temporally dependent nature of complex, real-world quality improvement initiatives, and it remains an implicit question for this research to understand the extent to which a quasi-experimental approach can successfully evaluate an intervention of this nature.

The results from modelling the impact of the feedback initiative over time on the various perioperative measures were mixed, with both positive and negative results across the various dimensions of quality assessed (*Table 73* gives an overview of primary results by perioperative area).

In terms of the study hypotheses, the following findings were observed from the primary analysis. Hypothesis 1 was concerned with the effect of implementing basic feedback where no prior feedback existed:

- Hypothesis 1: implementation of routine departmental anaesthetic quality monitoring and basic feedback at the level of the individual consultant anaesthetist will improve professional practice and improve anaesthetic care as evidenced by positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality.

The observed response to the implementation of basic feedback in the quality indicators assessed was, on the whole, limited, and the hypothesised benefits of implementing basic feedback were generally not observed in the data. While the average weekly temperature of patients arriving in recovery increased in response to the onset of basic feedback, in contrast, the weekly proportion of patients arriving in recovery with temperature $< 36^{\circ}\text{C}$ increased between the baseline and basic feedback epochs, too. Patient-reported QoR Scale scores showed no response to the implementation of basic feedback, with evidence to suggest that the proportion of patients reporting high scores may actually have decreased following the implementation of basic feedback. Patient-reported freedom from severe postoperative pain did not respond to the onset of basic feedback and the proportion of patients reporting freedom from postoperative nausea appeared to decline between the baseline and basic feedback epochs. No significant effects of the implementation of basic feedback were observed for either SSI rate or 30-day postoperative mortality rate. Given the largely negative results of implementing basic feedback, the implications for data reliability, particularly during the baseline period, will be discussed in more detail later in this section.

TABLE 73 Summary table of main quasi-experimental findings for primary analytic aims, by anaesthetic/perioperative quality area. The observed effects of implementing basic feedback within a feedback-naive cohort and then escalating the intensity of that feedback (the ‘enhanced’ feedback condition) are described, having controlled for longitudinal trends in the data

Anaesthetic quality indicator	Dependent measure used in analysis (and unit)	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
Patient temperature	Mean patient temperature on arrival in recovery (in °C)	Improvement in level observed After controlling for pre-intervention trends, patients were on average 0.082 °C warmer following implementation of basic feedback	Complex effect observed After controlling for pre-intervention trends, patients were on average 0.064 °C warmer following implementation of enhanced feedback, but the rate of change in temperature over time got worse (change in trend = -0.002 °C)
	Proportion of patients arriving in recovery < 36 °C (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 3.3% fewer patients arrived in recovery with temperature above the recommended 36 °C following implementation of basic feedback	Detrimental shift in rate of change observed After controlling for pre-intervention trends, the rate of change in the proportion of patients arriving in recovery with temperature above the recommended 36 °C got worse following implementation of enhanced feedback (change in trend < 0.1%)
QoR	Mean QoR Scale score (0–16; low- to high-quality recovery)	No significant effect observed After controlling for pre-intervention trends, no change in mean patient QoR Scale score was observed following implementation of basic feedback	Improvement in rate of change observed After controlling for pre-intervention trends, the rate of change in mean patient QoR Scale score improved following implementation of enhanced feedback (change in trend = 0.009 scale score)
	Proportion of patients with QoR score above 14 (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 9.4% fewer patients reported high QoR Scale scores (above 14) following implementation of basic feedback	Improvement in rate of change observed After controlling for pre-intervention trends, the rate of change in the proportion of patients reporting high QoR Scale scores (above 14) improved following implementation of enhanced feedback (change in trend = 0.001%)
Postoperative pain	Proportion of patients reporting freedom from severe pain during recovery (QoR item) (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in the proportion of patients reporting freedom from severe pain or constant moderate pain during recovery was observed following implementation of basic feedback	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 7.2% more patients reported freedom from severe or constant moderate pain during recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients reporting freedom from pain improved following implementation of enhanced feedback (change in trend = 0.004%)

Anaesthetic quality indicator	Dependent measure used in analysis (and unit)	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
	Proportion of patients reporting no or mild pain on arrival in recovery (excluding baseline period) (% proportion)	Not evaluated owing to absence of baseline data	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 12% more patients reported no or mild pain on arrival in recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients reporting no or mild pain improved following implementation of enhanced feedback (change in trend = 0.003%)
Postoperative nausea	Proportion of patients reporting freedom from PONV during recovery (QoR item) (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 4.6% fewer patients reported freedom from PONV following implementation of basic feedback	No significant effect observed After controlling for pre-intervention trends, no change in the proportion of patients reporting freedom from PONV was observed following implementation of enhanced feedback
	Proportion of patients with nurse-recorded absence of nausea during stay in recovery (excluding baseline period) (% proportion)	Not evaluated owing to absence of baseline data	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 5.8% more patients had nurse-recorded absence of nausea during their stay in recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients with nurse-recorded absence of nausea improved following implementation of enhanced feedback (change in trend = 0.001%)
SSI rate	Proportion of patients with a SSI-specific diagnostic code (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in SSI rate was observed following implementation of basic feedback	No significant effect observed After controlling for pre-intervention trends, no change in SSI rate was observed following implementation of enhanced feedback
Mortality	Proportion of patients with 30-day mortality following surgery (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in 30-day postoperative mortality rate was observed following implementation of basic feedback	Complex effect observed After controlling for pre-intervention trends, on average, 30-day postoperative mortality rate increased in level (0.8%) following implementation of enhanced feedback, but the rate of change in mortality rate over time improved slightly (change in trend < -0.001%)

H₁, hypothesis.
Beneficial effects of feedback are shaded dark green, detrimental effects are shaded blue, unambiguous or complex effects are shaded light green and those not evaluated are shaded grey.

Hypothesis 2 was concerned with the effect of implementing enhanced feedback in a group that had been receiving routine basic feedback previously:

- Hypothesis 2: implementation of enhanced feedback at the level of the individual consultant anaesthetist will improve professional practice and improve anaesthetic care as evidenced by positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality.

In contrast to the apparent effects of basic feedback, escalating the intensity of feedback through implementation of an enhanced feedback protocol had a more positive effect across a greater range of measures. After implementation of enhanced feedback, patients were, on average, warmer on arrival in recovery, although the rate of change in mean patient temperature got slightly worse. Again, the mean patient temperature and proportion of patients with temperature < 36 °C metrics did not show consistent results for the effects of enhanced feedback, with the latter demonstrating a detrimental response in the rate of change over time. Both the mean patient-reported QoR Scale score and proportion of patients reporting high-quality recovery showed improvement in rate of change, but not level, between basic and enhanced feedback conditions. The two measures of postoperative pain demonstrated consistent positive responses to the implementation of the enhanced feedback protocol, with significant improvement in both level and rate of change in the proportion of patients reporting freedom from severe pain and no or mild pain on arrival in recovery. The proportion of patients with nurse-reported absence of nausea similarly increased in response to the implementation of enhanced feedback, coupled with an improvement in the rate of change in this measure, although no significant effect was detected for patient-reported freedom from postoperative nausea. No significant effect of the implementation of enhanced feedback on SSI rate was detected. Thirty-day postoperative mortality appeared to show a complex response to the implementation of enhanced feedback, with a small increase in level coupled with a modest improvement in the rate of change over time.

Taken as a whole and in contrast to the observed effects of basic feedback, the implementation of enhanced feedback, as an escalation of the intensity of a basic feedback programme, was found to have a significant positive impact on a broad range of perioperative quality indicators, including mean patient temperature and QoR Scale score. The strongest apparent positive impact of enhanced feedback, however, was demonstrated for postoperative pain measures, both self-reported for duration in recovery and on arrival in recovery, and for nurse-recorded freedom from postoperative nausea.

Implications of secondary analysis for effect on perioperative quality

In order to investigate the robustness of the primary study findings, their sensitivity to variations in case inclusion criteria and to address a number of secondary research questions, a number of secondary analyses were performed for each perioperative quality measure. By way of a summary, the results from the secondary and sensitivity analyses are compiled in *Table 74*.

An important secondary research question for the study was whether or not the feedback initiative would have a significant effect on quality indicators for cases performed by anaesthetists who ranked in the lower half of the distribution of their peer group in the baseline phase of the study (hypothesis 3).

- Hypothesis 3: implementation and escalation of performance feedback at the level of the individual consultant anaesthetist will improve professional practice in anaesthetists ranked lowest at baseline. Any beneficial effects of feedback will be strongest for this subgroup, compared with the whole cohort.

TABLE 74 Summary table of findings from ITSA of primary evaluative measures and sensitivity analyses. Significant effects are recorded as either positive improvement or negative detrimental, relative to the effect of the feedback intervention. Temporal effects are additionally described as either stepwise shifts in level between experimental conditions, or change in slope denoting a shift in the rate of change in a variable over time between experimental conditions. For the secondary sensitivity analysis, the findings relative to those of the primary analysis for each quality indicator are described

Anaesthetic quality indicator	Primary analysis		Secondary and sensitivity analyses		Effect of inclusion of multisite data (excluding baseline and early FB period; excluding QoR items)	
	Dependent measure used in analysis	Effect of implementation of basic FB (H1)	Effect of implementation of enhanced FB (H2)	Effect on lower 50th percentile anaesthetist subgroup (H3)		Effect of project timeline variants: simple (one-stage) and complex (three-stage) timelines
Patient temperature	Mean patient temperature on arrival in recovery	Positive: stepwise improvement in level	Ambiguous: stepwise improvement in level paired with detrimental change in slope	Not evaluated	Positive effects of FB not replicated (model N/S)	Ambiguous effect of enhanced FB not reproduced. Negative detrimental effect on slope observed
	Proportion of patients arriving in recovery < 36 °C	Negative: stepwise detrimental change	Negative: detrimental change in slope	N/S	Simple effect of FB: positive – stepwise improvement in level. Complex effect of FB: positive effect of basic FB, positive effect of roll-in, negative effect of enhanced FB	Negative effect of enhanced FB replicated
					Simple effect of FB: negative – stepwise detrimental change. Complex effect of FB: negative effect of basic FB only	Negative effect of enhanced FB replicated
					Negative effect of basic FB replicated	Negative effect of enhanced FB replicated
					Negative effect of enhanced FB replicated	Negative effect of enhanced FB not replicated

continued

TABLE 74 Summary table of findings from ITSA of primary evaluative measures and sensitivity analyses. Significant effects are recorded as either positive improvement or negative detrimental, relative to the effect of the feedback intervention. Temporal effects are additionally described as either stepwise shifts in level between experimental conditions, or change in slope denoting a shift in the rate of change in a variable over time between experimental conditions. For the secondary sensitivity analysis, the findings relative to those of the primary analysis for each quality indicator are described (*continued*)

		Primary analysis			Secondary and sensitivity analyses			Effect of inclusion of multisite data (excluding baseline and early FB period, excluding QoR items)
Anaesthetic quality indicator	Dependent measure used in analysis	Effect of implementation of basic FB (H1)	Effect of implementation of enhanced FB (H2)	Effect on lower 50th percentile anaesthetist subgroup (H3)	Effect of controlling for age, gender and ASA score	Effect of project timeline variants: simple (one-stage) and complex (three-stage) timelines	Effect of limiting case inclusion criteria (elective, GA procedures)	
QoR	Mean QoR Scale score	N/S	Positive: improvement in slope	Not evaluated	N/S effect of baseline FB not replicated (i.e. detrimental effect present following control of covariates)	Simple effect of FB: negative – stepwise detrimental change. Complex effect of FB: negative change in slope for basic feedback inclusion)	N/S effect of baseline feedback not replicated (i.e. detrimental effect present following narrower case inclusion)	Not evaluated
	Proportion of patients with QoR score above 14	Negative: stepwise detrimental change	Positive: improvement in slope	Positive: improvement in slope (enhanced FB)	Positive effect of enhanced FB replicated	Simple effect of FB: negative – stepwise detrimental change. Complex effect of FB: negative change in slope for basic FB	Positive effect of enhanced FB not replicated	Not evaluated
					ASA score was a significant predictor			

Primary analysis		Secondary and sensitivity analyses			Effect of inclusion of multisite data (excluding baseline and early FB period; excluding QoR items)			
Anaesthetic quality indicator	Dependent measure used in analysis	Effect of implementation of basic FB (H1)	Effect of implementation of enhanced FB (H2)	Effect on lower 50th percentile anaesthetist subgroup (H3)		Effect of controlling for age, gender and ASA score	Effect of project timeline variants: simple (one-stage) and complex (three-stage) timelines	Effect of limiting case inclusion criteria (elective, GA procedures)
Postoperative pain	Proportion of patients reporting freedom from severe pain during recovery (QoR item)	N/S	Positive: stepwise improvement in level and improvement in slope	Positive: improvement in slope (enhanced FB)	N/S effect of basic FB replicated Positive effects of enhanced FB replicated ASA score was a significant predictor	Simple effect of FB: N/S. Complex effect of FB: N/S	N/S effect of baseline FB replicated. Positive: stepwise improvement in level from enhanced FB replicated (but not change in slope)	Not evaluated
	Proportion of patients reporting no or mild pain on arrival in recovery (excluding baseline period)	Not evaluated	Positive: stepwise improvement in level and improvement in slope	Positive: stepwise improvement in level and improvement in slope (enhanced FB)	Positive effect of enhanced FB replicated	Simple effect of FB: not evaluated. Complex effect of FB: positive effect of both roll-in and enhanced FB	Positive effect of enhanced FB replicated	Positive effect of enhanced FB replicated

continued

TABLE 74 Summary table of findings from ITSA of primary evaluative measures and sensitivity analyses. Significant effects are recorded as either positive improvement or negative detrimental, relative to the effect of the feedback intervention. Temporal effects are additionally described as either stepwise shifts in level between experimental conditions, or change in slope denoting a shift in the rate of change in a variable over time between experimental conditions. For the secondary sensitivity analysis, the findings relative to those of the primary analysis for each quality indicator are described (*continued*)

Primary analysis		Secondary and sensitivity analyses						
Anaesthetic quality indicator	Dependent measure used in analysis	Effect of implementation of basic FB (H1)	Effect of implementation of enhanced FB (H2)	Effect on lower 50th percentile anaesthetist subgroup (H3)	Effect of controlling for age, gender and ASA score	Effect of project timeline variants: simple (one-stage) and complex (three-stage) timelines	Effect of limiting case inclusion criteria (elective, GA procedures)	Effect of inclusion of multisite data (excluding baseline and early FB period; excluding QoR items)
Postoperative nausea	Proportion of patients reporting freedom from PONV during recovery (QoR item)	Negative: stepwise detrimental change	N/S	Negative: detrimental change in slope (basic FB)	Negative effect of basic FB replicated N/S effect of enhanced FB replicated	Simple effect of FB: negative – stepwise detrimental change. Complex effect of FB: N/S	Negative effect of basic FB not replicated but detrimental effect on slope of enhanced FB now observed	Not evaluated
	Proportion of patients with nurse-recorded absence of nausea during stay in recovery (excluding baseline period)	Not evaluated	Positive: stepwise improvement in level and improvement in slope	Positive: stepwise improvement in level and improvement in slope (enhanced FB)	Positive effect of enhanced FB replicated	Simple effect of FB: not evaluated. Complex effect of FB: positive effect of enhanced FB	Positive effect of enhanced FB not replicated. Now N/S	
SSI rate	Proportion of patients with a SSI-specific diagnostic code	N/S	N/S	Not evaluated	N/S effect of baseline FB replicated N/S effect of enhanced FB replicated	Simple effect of FB: N/S. Complex effect of FB: ambiguous – positive effect of roll-in period and negative effect of enhanced FB	N/S effects of FB not replicated. Significant improvement in slope observed in response to enhanced FB, but overall model fit not significant	N/S effect of enhanced FB replicated

Primary analysis		Secondary and sensitivity analyses						
Anaesthetic quality indicator	Dependent measure used in analysis	Effect of implementation of basic FB (H1)	Effect of implementation of enhanced FB (H2)	Effect on lower 50th percentile anaesthetist subgroup (H3)	Effect of controlling for age, gender and ASA score	Effect of project timeline variants: simple (one-stage) and complex (three-stage) timelines	Effect of limiting case inclusion criteria (elective, GA procedures)	Effect of inclusion of multisite data (excluding baseline and early FB period; excluding QoR items)
Mortality	Proportion of patients with 30-day mortality following surgery	N/S	Ambiguous: stepwise detrimental change paired with improvement in slope	Not evaluated	N/S effect of baseline feedback replicated	Simple effect of FB: N/S. Complex effect of FB: N/S	N/S effect of baseline FB replicated	Ambiguous effect of enhanced FB not replicated (now N/S)
					Ambiguous effect of enhanced FB not replicated (negative stepwise effect only observed)	Ambiguous effect of enhanced FB not replicated (now N/S)		

FB, feedback; H, hypothesis; N/S, not significant.

When the impact of basic and enhanced feedback was examined for the proportion of patients with temperature below the recommended 36 °C for the lower 50th percentile group, the detrimental trends in the data observed for the broader cohort group were not detected for the lower 50th percentile group. The positive improvement in trend in proportion of patients with high QoR scores after enhanced feedback for the full cohort was replicated for the lower 50th percentile group, with a slightly larger effect size, and no detrimental effect of baseline feedback was detected (as had been present for the full cohort). Similarly, for both the proportion of patients reporting no or mild pain on arrival in recovery and the proportion of patients with nurse-reported absence of nausea, the significant improvements in level and slope observed for the full cohort were replicated in the lower 50th percentile cohort, with larger effect sizes. Based on these findings, the third study hypothesis (3) is supported for the effects of enhanced feedback on anaesthetic quality indicator scores for the lower-ranked anaesthetist group, in the areas of postoperative pain management, control of postoperative nausea and overall patient-reported QoR.

Given the broad inclusion of surgical case types in the sample at the primary study site and variations in the patient group profiles of participating anaesthetists, it was important to investigate the effects on the results from the primary analysis, of controlling for patient age, gender and ASA score as proxies for case mix and condition severity. For the patient temperature measures, the covariate model confirmed the apparent trends towards increasing proportions of patients arriving in recovery cold, following both iterations of feedback and similarly replicated the observed effect of enhanced feedback on mean patient temperature data. For the QoR score measures, the covariate model replicated the effect of enhanced feedback on mean QoR score, but not proportion of high QoR scores. The effect of enhanced feedback on proportion of patients with high QoR scores was attenuated by longitudinal variation in ASA score, which retained significance in the model. The effects of feedback following control of ASA score, age and gender in the remaining models, including the effects of enhanced feedback on postoperative pain and nausea measures, were largely unchanged. Interestingly in all the covariate models, only ASA score retained significance as a predictor of both high QoR scores and the QoR-based postoperative pain measure, when entered with the intervention parameters, demonstrating the unique contribution of disease severity to variance in postoperative pain and QoR, irrespective of longitudinal feedback effects. Overall, the conclusion drawn from covariate analysis was that the observed positive effects of enhanced feedback on patient temperature data, mean QoR score, and measures of both postoperative pain and postoperative nausea, were robust after controlling for longitudinal variation in disease severity, patient age and gender.

For variations in case, inclusion criteria were modelled, when a stricter case inclusion criteria was used, limited to elective GA cases, despite the loss in statistical power the implementation of enhanced feedback was still associated with significant improvement in level and slope in both proportion of patients with nurse-reported freedom from nausea and freedom from pain on arrival in recovery. This finding lends further confidence to the conclusions from the primary analysis in these areas.

In considering the findings from the secondary analyses, comparison of the model fit statistics for each ITS model can provide additional information concerning the validity of each set of specified parameters relative to observed trends in the data. Direct statistical comparison of the proportion of variance explained by each model was not possible as the secondary analyses were not nested models (i.e. they were based on different sample populations influenced by differing case inclusion criteria and listwise case exclusion, and also different time series parameters); therefore, statistical analysis of change in the *F*-statistic between nested models could not be reported. Taken at a more descriptive, superficial level, however, comparison of the model fit statistics across sensitivity analyses and across dependent variables can provide an indication of how well the project timelines, covariate models and case inclusion criteria accounted for longitudinal variation in different perioperative indicators. This information is summarised from the results section in *Table 75*.

TABLE 75 Summary table of model fit statistics for ITS models fitted under the various primary and secondary analytic aim categories. Within each cell, the proportion of variance in the dependent variable explained by the overall model is indicated (synonymous with the square of the multiple *R*-statistic, expressed as a percentage), along with the *F*-statistic from ANOVA (a comparable indicator of the degree to which the model fits the data) and the associated significance of *F*

Anaesthetic quality indicator	Dependent measure used in analysis	Primary analysis		Secondary analyses				
		Two-stage intervention model (basic and enhanced feedback)	Lower 50th percentile anaesthetist subgroup	Covariate model controlling for age, gender and ASA score	Simple (one-stage) project timeline variant	Complex (three-stage) project timeline variant	Elective GA case inclusion criteria	Multisite data set
Patient temperature	Mean patient temperature on arrival in recovery	35.2%; <i>F</i> = 19.967; <i>p</i> < 0.001	N/A	27.4%; <i>F</i> = 7.674; <i>p</i> < 0.001	27.9%; <i>F</i> = 23.988; <i>p</i> < 0.001	37.4%; <i>F</i> = 15.513; <i>p</i> < 0.001	30.6%; <i>F</i> = 11.625; <i>p</i> < 0.001	16.4%; <i>F</i> = 8.114; <i>p</i> < 0.001
	Proportion of patients arriving in recovery < 36 °C	19.3%; <i>F</i> = 8.812; <i>p</i> < 0.001	16.8%; <i>F</i> = 7.5; <i>p</i> < 0.001	23.5%; <i>F</i> = 6.243; <i>p</i> < 0.001	16.9%; <i>F</i> = 12.589; <i>p</i> < 0.001	20.5%; <i>F</i> = 6.714; <i>p</i> < 0.001	23.9%; <i>F</i> = 8.313; <i>p</i> < 0.001	20.3%; <i>F</i> = 10.520; <i>p</i> < 0.001
QoR	Mean QoR	11.1%; <i>F</i> = 4.535; <i>p</i> < 0.005	N/A	16.6%; <i>F</i> = 4.048; <i>p</i> < 0.001	3.6%; <i>F</i> = 2.271; <i>N/S</i>	13.7%; <i>F</i> = 4.089; <i>p</i> < 0.001	8.9%; <i>F</i> = 2.572; <i>p</i> < 0.05	N/A
	Scale score	7.3%; <i>F</i> = 2.864; <i>p</i> < 0.05	8%; <i>F</i> = 7.5; <i>p</i> < 0.01	12.4%; <i>F</i> = 2.878; <i>p</i> < 0.01	4.9%; <i>F</i> = 3.164; <i>p</i> < 0.05	10.9%; <i>F</i> = 3.138; <i>p</i> < 0.005	6.9%; <i>F</i> = 1.949; <i>N/S</i>	N/A
Postoperative pain	Proportion of patients reporting freedom from severe pain during recovery (QoR)	28.1%; <i>F</i> = 14.352; <i>p</i> < 0.001	8.7%; <i>F</i> = 3.548; <i>p</i> < 0.005	30.9%; <i>F</i> = 9.114; <i>p</i> < 0.001	6.8%; <i>F</i> = 4.510; <i>p</i> < 0.01	29.4%; <i>F</i> = 10.832; <i>p</i> < 0.001	13.8%; <i>F</i> = 4.222; <i>p</i> < 0.005	N/A
	Proportion of patients reporting no or mild pain on arrival in recovery	49.7%; <i>F</i> = 51.766; <i>p</i> < 0.001	44.5%; <i>F</i> = 42.182; <i>p</i> < 0.001	51.4%; <i>F</i> = 27.127; <i>p</i> < 0.001	N/A	55.5%; <i>F</i> = 38.676; <i>p</i> < 0.001	39.3%; <i>F</i> = 26.120; <i>p</i> < 0.001	33.7%; <i>F</i> = 21.053; <i>p</i> < 0.001

continued

TABLE 75 Summary table of model fit statistics for ITS models fitted under the various primary and secondary analytic aim categories. Within each cell, the proportion of variance in the dependent variable explained by the overall model is indicated (synonymous with the square of the multiple R-statistic, expressed as a percentage), along with the F-statistic from ANOVA (a comparable indicator of the degree to which the model fits the data) and the associated significance of F (*continued*)

Anaesthetic quality indicator	Dependent measure used in analysis	Primary analysis		Secondary analyses				
		Two-stage intervention model (basic and enhanced feedback)	Lower 50th percentile anaesthetist subgroup	Covariate model controlling for age, gender and ASA score	Simple (one-stage) project timeline variant	Complex (three-stage) project timeline variant	Elective GA case inclusion criteria	Multisite data set
Postoperative nausea	Proportion of patients reporting freedom from PONV during recovery (QoR)	7.4%; F = 2.943; p < 0.05	2%; F = 2.407; p < 0.05	7.7%; F = 1.706; N/S	6.4%; F = 4.269; p < 0.01	8.5%; F = 2.4; p < 0.05	6%; F = 1.687; N/S	N/A
		38.1%; F = 32.196; p < 0.001	25.8%; F = 18.238; p < 0.001	38.4%; F = 16.016; p < 0.001	N/A	40.9%; F = 21.465; p < 0.001	30.7%; F = 17.881; p < 0.001	6.3%; F = 2.787; p < 0.05
SSI rate	Proportion of patients with nurse-recorded absence of nausea during stay in recovery	5.6%; F = 0.238; N/S	N/A	6.3%; F = 1.362; N/S	3.5%; F = 1.990; N/S	8.3%; F = 2.119; p < 0.05	4.4%; F = 1.214; N/S	7.9%; F = 3.530; p < 0.05
		6.6%; F = 2.328; p < 0.05	N/A	9.5%; F = 2.119; p < 0.05	1.8%; F = 1.025; N/S	6.7%; F = 1.676; N/S	0.9%; F = 0.248; N/S	2.5%; F = 1.042; N/S
Mortality	Proportion of patients with 30-day mortality following surgery							

N/A, not applicable; N/S, not significant. Models that did not achieve significant predictive fit and cells which represent combinations that were not possible to analyse are shaded.

From comparison of the *F*-statistics it is clear that the two-stage intervention model (basic and enhanced feedback), used in the primary analysis, fitted trends in some perioperative indicators better than others. Measures such as patient temperature, the PACU pain measure and the nurse-reported nausea measure showed good fit for the two-stage model, in comparison with the QoR Scale score and its PONV item, for which a relatively small amount of variance was accounted for by the two-stage model. The primary analytic models accounted for a very small proportion of the available variance in the SSI and 30-day mortality measures and, indeed, proportion of variance explained for these measures was very small across all model variants. This finding is to be expected, as many peri- and postoperative factors influence downstream outcomes such as infection and mortality, beyond the control of the attending anaesthetist. In comparison, a much larger proportion of variance was accounted for by the two-stage feedback intervention model for both postoperative pain measures and, to a lesser degree, postoperative nausea and patient temperature. Proportionately, the anaesthetist has considerably more control over postoperative pain and nausea than other potential surgical outcomes.

In considering the overall model fit for the covariate models, as is expected when additional parameters are added to a model, the proportion of variance accounted for increases. Noteworthy, however, is that the model fit (as indicated by the *F*-statistic) generally did not improve with the addition of the covariates to the two-stage intervention model, indicating limited change in the models predictive capability when ASA score, age and gender were combined with the time series parameters associated with the onset and trends linked to basic and enhanced feedback.

In terms of the question of whether or not the two-stage intervention model was the best fit for the observed variance in the perioperative indicators, the results indicate that this was largely the case for the majority of measures, and certainly the two-stage models fitted for the QoR Scale score, both pain metrics, nurse-reported nausea, SSI and mortality rates demonstrated better fit than either the simpler or more complex intervention model variants tested. In contrast, a simple single-phase intervention model (baseline – feedback) explained marginally more variance in the patient temperature measures than any other timeline variant, and the same was the case for the QoR postoperative nausea measure.

Data reliability

The inconsistent effects of basic and enhanced feedback observed in this study call into question the reliability of the baseline data and data collected in the early phase of this project. The contradictory nature of a number of the results for specific measures suggests that statistical artefacts or anomalies in the baseline data set are a valid potential confounding factor for several parts of the analysis and may provide an alternative explanation for the apparent negative effects of implementation of basic feedback, followed by positive or ambiguous effects of enhanced feedback (the latter having occurred after data collection practices had become better standardised). In the history of the project, several iterations were made to the data collection protocol to improve the reliability of recording, the theatre administration system was implemented and became available for use by the project team and during the initial feedback period a researcher was appointed to work with the PACU team to develop the data collection and feedback system. The subsequently developed data collection process had a number of validation steps to improve the reliability and consistency of the data (*Figure 20*). It is likely that these developments had an impact on longitudinal trends in the time series data.

The potential for anomalous baseline effects is evident from reviewing the time series charts for proportion of patients with temperature < 36 °C and for proportion of patients with no or mild pain, the latter in particular displaying a strong U-shaped curve, as opposed to the hypothesised linear, stepwise trend. It may, therefore, be the case that a quality monitoring and feedback intervention acts first to stabilise and improve the reliability of data collection before more subtle effects on underlying clinical processes driving the measures may be detected. The effect of quality improvement programmes on the reliability of measurement over time is often reported in case studies and any evaluation of such an intervention must

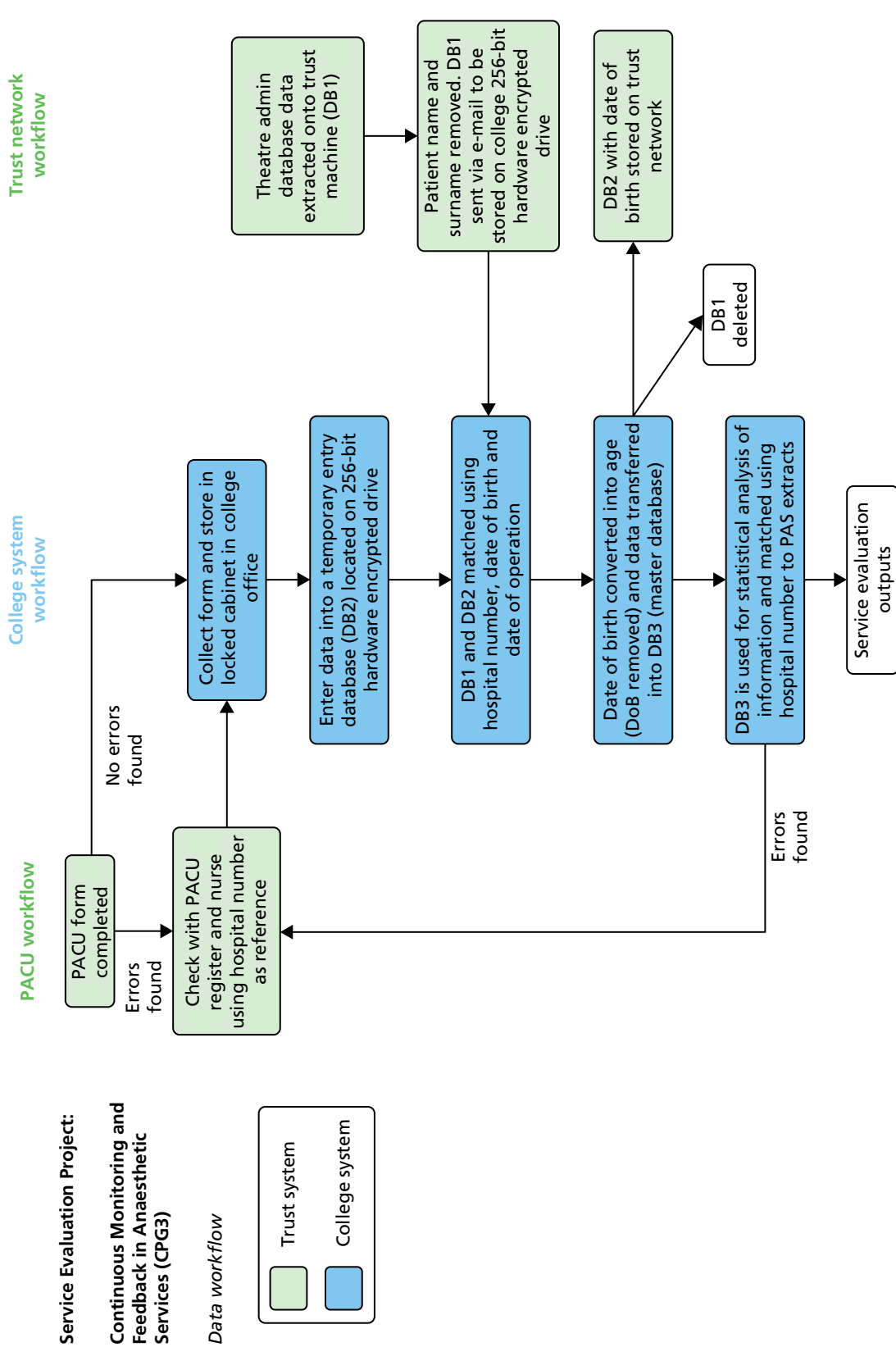


FIGURE 20 Schematic of information workflow developed following appointment of researcher to oversee data collection in PACU and development of feedback system (March 2011). Data at all times are stored on the trust network or on a 256-bit hardware encrypted drive at the college. Databases containing full patient identifiable information do not leave the trust network. Patient name and surname are not transferred to the college system at any point. Hospital number, date of birth and sex are transferred to allow analysis of subgroups and matching/data validation. E-mails with data are deleted after sending from 'sent items' and 'deleted items': DB, database; DoB, date of birth; PAS, patient administration system.

be able to take this effect into account. Any effects of inaccuracy in baseline measurement were potentially exacerbated by the relatively short baseline collection period, compared with the intervention conditions and the presence of missing data, both administrative and recovery indicators. Again, the trend towards increasing valid case frequency over time in the primary study site indicates improvement in the reliability of the data collection protocol over time.

Regarding potential issues with the baseline measures, there are two possible potential options for further investigation. The first involves post-hoc assignment of a roll-in period in which presumably data collection was initially unreliable and improved in consistency over time. This period could then be excluded from the analysis, allowing evaluation of an apparently cleaner trend line that embodies more signal from the feedback intervention. The second option would be to identify the offending data collection period as previously indicated and control for the slope of the trend during that period. Both options, however, involve relaxing some experimental control relative to a 'fair test' of our intervention, with knock-on effects for the robustness of the evaluation and the conclusions we are able to draw about the effects of implementing feedback. In this sense both options would involve removing or confounding the project baseline period in some way and would relegate the evaluation of a test of whether or not implementing enhanced feedback had additional effects over receiving basic feedback. These further analyses remain viable options for further exploration of the data set, however.

Study limitations

The design for this study was a single-group longitudinal study using ITSA. While the strength of this design is linked to its ability to model and control for temporal trends in a longitudinal data set (as discussed in the rationale for selection of the time series design in *Statistical methods*), the lack of control group poses challenges in terms of allocating causality for any observed apparent effects of interruption in the time series to the underlying intervention timeline. The main threat to validity occurs owing to the possibility that observed trends in the time series might be explained by specific extraneous confounders, such as concurrent changes to the local perioperative service, its staffing or organisation. In the context of this study, owing to limitations in the final data set for analysis as a result of missing administrative data and the organic nature of quality improvement programmes resulting in spread of the intervention over time to additional sites, the resulting variable case profile within the time series may have confounded any analysis of macro-level data that sought to determine an overall effect of this quality improvement initiative across the perioperative units in which it was implemented. Similarly, analysis of the effects of the intervention at the primary study site, St Mary's Hospital, was limited by the availability of contextual administrative data during the study baseline period, which meant that any analysis of cases limited to a homogenous subgroup would essentially be based on a truncated time series, with knock-on detrimental effects on the stability of parameter estimates for phases within the time series models. In order to address these limitations, multiple subanalyses were performed based on specific case inclusion criteria that maximised use of the available data. While maximising statistical power in each subanalysis, this strategy makes the evaluation more complex in terms of analysis and interpretation, especially as achieving a homogenous sample of perioperative cases throughout the duration of the project timeline and which had all been similarly exposed to the intervention was impossible without sacrificing some element of experimental control, such as the ability to homogenise anaesthetic type and admission type.

Regression to the mean is always a possible alternative explanation for observed temporal effects in longitudinal studies, especially where the baseline group was selected on the basis of low scores on a measure.¹⁸⁶ The adverse effects of regression to the mean are proportional to the degree of measurement error (random variation) present in a data set. In order to reduce the opportunity for regression to the mean as an alternative explanation for effects in the analysis of the lower 50th percentile of anaesthetists' quality indicator scores, we have maximised the number of case observations on which the rank ordering of anaesthetists was based. The rank ordering should, therefore, be a robust estimate.

The research suggests that it is imperative that data feedback initiatives designed to improve quality of care are based on effective measurement and data administration processes that are embedded within routine perioperative practice and capable of producing robust and reliable data over time. Information technology (IT) support and integration within electronic patient administration systems, where mature, well designed and well maintained, has the potential to greatly support this aim. This is particularly important given the potential sensitivity of certain perioperative quality indicators to extraneous variation in the reliability and quality of data collection protocols, and the need to control the signal-to-noise ratio, as was demonstrated in this study. Comparatively, variations in certain indicators due to underlying fluctuations in quality of care may represent very small effect sizes that are difficult to detect without sufficient investment in obtaining robust, reliable and longitudinal data.

In terms of the implications for evaluation of quality improvement interventions, this work highlights the need for mixed-methods designs that are sensitive to complex effects of interventions across social and technical dimensions of care provision. Evaluative designs for quality improvement programmes should additionally be sensitive to longitudinal variation, as is evident from the need for examination of time series trends in the response to multiple or serial interventions and for the examination of the reliability of soft data collected over time. Furthermore, the experience within this project of attempting to fit linear models to describe complex serial intervention effects demonstrates the limits of even the most flexible quasi-experimental designs, necessitating the combination of complementary designs that are sensitive to context and process.

Chapter 9 Impact on perioperative productivity

This chapter reports the rationale, methodology and findings from the productivity work stream of the evaluation. *Box 7* contains all key research aims from the original project protocol, with those of greatest relevance to the productivity analysis presented in bold.

The productivity analysis addresses these aims by exploring the impact of the feedback initiative on perioperative efficiency. These outcomes are explored both from the perspective of end-users and through a more objective analysis of effects on the metric, WWT. Through the exploration of the experiences of end-users there is an opportunity to investigate other contextual factors which have the potential to influence an individual's ability to make improvements to the WWT experienced by their patients. In doing this we can gain a broader understanding of the system in which the feedback initiative is embedded and the interaction that this has with overall effects.

Introduction

Hospital productivity and efficiency are prominent governmental objectives. Financial pressures, health-care policies and quality monitoring increase the demand to optimise hospital productivity. Financially, the NHS aimed to save £20B by 2014.¹⁸⁷ Clinical pathways and patient flow predominantly influence hospital productivity. Central to this is the perioperative pathway; surgical procedures currently account for approximately 9% of a hospital's total costs.¹⁸⁸

BOX 7 Primary and secondary aims from original protocol

Research aims

To evaluate the impact of a departmental continuous quality monitoring and multilevel feedback initiative on the quality of anaesthetic care and efficiency of perioperative workflow within a London teaching hospital over a 2-year period. Data will be analysed at multiple time points over the course of the project in order to generate both formative and summative information to support development and evaluation of the initiative.

To employ a quasi-experimental time series design to provide robust evidence concerning the impact of a serial data feedback intervention on anaesthetic quality indicators while controlling for baseline variance.

To contribute to the evidence base for valid and reliable quality indicators for anaesthetic care including effective patient experience measures.

To document the main features of the data feedback intervention as it develops through the CLAHRC programme for replication at other sites, including definition of metrics, data processes, feedback format and action mechanisms.

To assess the perceived acceptability and utility of this information system for individual end-users, the clinical department and other organisational stakeholders, using a formative, mixed-methods design.

The PACU has been identified as a critical step in surgical patient flow, impacting theatre work-streams. Research shows that health professionals of differing disciplines vary in their perceptions of the causes of delays. Delays disrupt productivity and working hours as well as patient satisfaction and experience. Evidence shows a dependency between waiting time and theatre throughput.¹⁸⁹ The PACU is not designed to care for postoperative patients post recovery from anaesthesia for prolonged periods. Additionally, patients awaiting discharge are occupying limited bed spaces assigned to surgical patients on the upstream theatre list. In the worst cases, this can prevent admission to PACU from theatre resulting in theatre delays and cancellations on the day of surgery. Patients overstay in the PACU where delays in transfer occur, resulting in increased pressure on PACU staff. Overstaying patients consume considerable amounts of nursing time, threatening staff-to-patient ratios for recovering surgical patients. While there will always be some amount of WWT due to the distance between the PACU and the ward, and handover times, excessive wait times of 30 minutes or more have a considerable detrimental effect on perioperative workflow. This issue and how to improve patient transfer times has received little direct attention in the research literature to date. Educational initiatives have been shown to improve delays through behaviour modification.¹⁹⁰

The provision of feedback can influence clinical practice and performance through motivation and reinforcement of 'good practice'.^{13,49,54,57,135,148,191-193} These effects are enforced if feedback is part of a strategy, targeted at decision-makers. Furthermore, studies show that feedback has pronounced effects if prior baseline compliance is low.¹⁹⁴ Statistical process control, adopted from industry, has been identified as a means of monitoring and providing feedback through the use of run charts. It allows the detection, analysis and evaluation of performance variations.¹⁹⁵

Effective feedback has been defined as 'feedback in which information about previous performance is used to promote positive and desirable development'. Emerging research has yielded characteristics which enhance this. For example, feedback should be timely, credible, specific and focused on tasks rather than individuals.^{105,135,156,191,196} However, these factors can be subjective. Furthermore, feedback efficacy has sociotechnical influences, and context needs to be considered. Archer states that a 'feedback culture' needs embedment, with reciprocal learning from stakeholders. Facilitation, sequential support and adherence to the planned change have been reported to maximise feedback effects.¹⁹⁶

To establish a successful, sustainable monitoring and feedback system, end-user input is required. NHS ward leads, theatre managers and bed managers have vital roles in co-ordinating hospital care. Engagement is a prerequisite to sustainability. This chapter specifically reports the impact and lessons learnt from the initiative from the managerial and patient-flow perspective. The aims were to understand the impact of performance feedback at ward level on patient flow, namely the timely transfer of surgical patients from the PACU to surgical wards.

Methods

Study design

A mixed-methods evaluation of the effects of the feedback intervention on productivity.

Qualitative data

The time point 1 interviews that had been conducted with the perioperative service manager, the lead nurse for the PACU and six surgical nursing leads from the primary site were reanalysed to identify and interpret themes related to productivity.

Quantitative data

Ward wait time was defined as the interval between the receiving ward being contacted after the patient was deemed ready for discharge from the PACU unit and handover of the patient to the ward nurse. PACU staff recorded key transfer time points on a paper form for each patient, along with a number of additional QoR indicators. Data were then entered into an IT-based theatre administration system from which WWT interval was calculated.

Analysis

Interview data were analysed based on the principles of grounded theory. The constant-comparative technique was used to refine categories of information until data saturation occurred. The category structure was reviewed and refined by a multidisciplinary team comprising two social sciences researchers and a research assistant with a clinical background.

We used *t*-tests to identify significant differences in WWT pre and post feedback both for the aggregated hospital level data and for individual surgical wards. Analysis was conducted using SPSS version 20.

Results

Factors influencing patient transfer efficiency

Conflicting departmental priorities

Some leads acknowledged that discharge delays and optimising ward transfer times were not a priority for their department, impacting the extent of their engagement and motivation to improve patient transfer. Conflicting priorities lay between hospital productivity, achieving targets and clinical priorities.

I need to engage my staff in delivering good care, increasing knowledge . . . targets . . . new equipment . . . And that's just to name a few.

I don't see, if there's no pressure coming from theatres, why it should be so important.

Discrepancies in professional judgement

Influencing these priorities were perceptions of patients' best interests. Qualitative analysis revealed discrepancies here, concerning benefits of high staff-to-patient ratio in recovery conflicting with unsatisfactory environmental factors such as noise, continuity of care and comfort.

It depends on where you're going. Sometimes for the patients it might be better that they stay in recovery because they get one-to-one nursing for a while.

Staying in recovery on a trolley for x number of hours is not very good patient experience . . . you want to come up to the ward where it's much more comfortable, where you can have continuity of care and all of that.

System interdependencies and awareness of upstream consequences on patient flow

Although the value of optimising patient transfer was acknowledged, interviewees described how components of the patient pathway are interdependent; multiple factors impact patient flow. Links between bed availability, patient discharge and pharmacy were described.

If your discharges are delayed, the EDCs [electronic discharge checklists] are not done on time . . . The beds are not cleared as quickly as you'd like . . .

In addition, interviewees identified that a lack of understanding of these interdependencies and their upstream effects contributed to delays. For example, the implications of bypassing procedures, misallocation of patients to colleague's 'empty beds' and the subsequent impact on theatre lists. The importance of understanding these upstream effects was that they were reported to impact staff satisfaction, busyness and patient discharge.

What they don't understand is the impact, OK? They just see a space and their patient can stay in that space . . . Because they don't care about the other person's list.

Last night for instance . . . The recovery staff had to work longer and harder.

Oh, our nurses have gone for a tea break. That's great but my nurses aren't going to get for one because your nurses are having one, you know.

So no, I think it's very useful and it shows . . . because before looking at how long it was taking, an hour and a half to get somebody down to a particular ward, I mean that's appalling. So that just shows a reflection on delaying discharge and stuff.

Interviewees highlighted the need to target upstream effects via education, in particular education of system processes, encouraging junior doctors to prioritise electronic discharges and ensuring that health professionals understand the impact of their actions on the hospital system.

When you're bringing in the junior doctors . . . maybe just telling them the importance of the EDCs . . . map the patient's pathway and show them the impact on timely discharges, the impact it has on the whole trust . . .

Lack of resources and management of resources

In line with this, specific barriers were the perceived pressure on resources, bed availability and time. Interlinking and heightening this were the multiple nurse duties. Further pressure was perceived through emergency cases.

I think I know the reasons for delay. It's either because people are on their breaks and there's not enough staff on the ward to leave it safe, or it's that we don't have beds.

We have a problem with the emergency work . . . I think it's increased significantly since we got the trauma unit and I think that's had an impact on beds as well.

Variation in patient transfer was perceived through availability of departmental resources and ability to plan for/allocate patients to beds. When asked if their respective ward had issues with WWT, one respondent replied:

Not for this floor, no, because we have the beds specifically for the patients.

The big issue with the bed managers for me is when we have outlying patients . . . They were really good on Monday and they moved those two outliers out of the ward as soon as they could, and by the end of the day we did get all our patients in.

Poor communication

Communication was regarded as fundamental to the productivity of system work streams. Poor communication was described as inhibiting productivity and invoking interdisciplinary friction. These communication issues, mainly bed allocation and communication with recovery, were reflected at multiple points.

Bed managers don't always give us this information in the morning of who has got beds and who's not got beds.

They ring me every morning now and I might say that sometimes it just about does my head in, because I get enough phone calls.

But when I ask which patient sometimes the person that's rung isn't even aware of what list is being done.

Poor co-ordination between teams

Coinciding with communication issues, poor interdisciplinary team working was perceived to negatively impact patient care, staff satisfaction and service efficiency. This was exacerbated by the separate nature of health-care roles – working in parallel rather than in synergy, power struggles and task prioritisation. Furthermore, staff identified that abuse of the system by colleagues created friction between health-care professionals, impacting perceptions of 'norms' and their behaviour in the future. Interdisciplinary teamwork was also affected by professional judgement, including differences in the perceived time a patient was ready to be transferred owing to his or her individual needs, pain being one example.

This is something missing here . . . the teamwork between medics and nurses here.

There might be the inclination like, well we went across before and they weren't ready so we'll leave it for another 5 or 10 minutes.

Because when we go across to get a patient and they're still in pain, it's always oh, well that's just come on.

A general theme that emerged was the need for a culture change in the work patterns of health professionals. The main themes here were the need to increase unity, teamwork and understanding of the consequences of one's actions on patient flow.

So, this needs to change, the culture of the doctors here working completely separately from the nursing staff and coming and giving orders and nothing else.

Sometimes the culture is that it is not my patient so you have to wait for that nurse who is allocated to that patient to pick the patient up. So it's changing the culture as well.

Inadequate discharge process

A recurring barrier was inadequacy of the current discharge process, through lack of compliance with protocols or lack of a discharge system. Bottlenecks in this process were timely completion of EDCs by junior doctors, and the delays this invokes for pharmacy. To facilitate improvements, interviewees recommended potential changes, namely nurse-led discharge, specific discharge times and increased organisation.

It's been an ongoing problem, and it's just like hitting your head against a brick wall. Some doctors will not start EDC until the day pharmacy have said they must.

It's the discharge process on the ward . . . I don't have the facts of how it goes on, but I want to assume that it's not happening as it should be . . .

You need to be around first thing in the morning to do your ward rounds and you need to discharge people. I think there could be an awful lot more nurse led discharge.

What I ideally would like is a discharge time in the morning . . . to tell my women . . . yes you'll be going home tomorrow and the discharge time is 9 o'clock.

Effective management

An active manager was perceived to positively influence patient flow. In particular, engagement, working with clinical teams was perceived to positively impact interdisciplinary team working.

I'm one of the few ward managers that actually goes up there, and I don't see why not. I'm a nurse. And I just think you need to be seen and be on the shop floor and to get it going.

Moreover, interviewees depicted how managerial support and representation from the trust was required to optimise results and unite the organisation.

There are some problems that we need support from higher people . . . like dealing with the medical team which is completely separate, we need someone to actually be our voice and be our representative.

Need for a system-wide approach

The interviews highlighted that hospital productivity is not purely about effective surgery; rather a system-wide approach is needed to optimise efficiency, patient satisfaction and patient flow rates.

Theatres is only a tiny bit of the whole process. It's a huge systems process.

We need an entire system shift.

Pre-assessment is a good place to have the right information about what they're coming in for and their length of stay. But I suppose that's a wider issue for the trust.

No, because there's a whole-systems approach that needs to change before we get anywhere. And people have been bleating on about the same thing for at least 7 years that I've been here.

To achieve this, we need to optimise the discharge system through organisation, maximising the roles of existing staff as well as the potential to create new staff roles. To maximise impact of the initiative, many interviewees identified that further involvement of higher managerial levels and senior engagement would be required.

I would like to see that data used to produce change . . . I don't think we're knocking on the right doors. And it might be that you need to go to the director of nursing and say to her, have you seen this data, is there anything you can do to help . . .

I think you need to take a different approach higher up.

Effectiveness of feedback

At aggregated level, combined analysis of WWT across all eight wards showed no significant improvement from before (mean = 51.24; SD = 69.18) to after (mean = 51.35; SD = 72.96) initiation of the monitoring and feedback initiative, $t(12742) = -0.09$; $p = 0.930$. Participants reported a range of experiences relating to feedback use and factors influencing patient transfer.

Need for the project/justification for the project

Before the initiative, a consensus about lack of feedback emerged. Information was obtained through personal or departmental audits. However, staff were aware of current problems with patient transfer.

The problem before was that the staff on the ward wouldn't come down to get patients, for hours on end . . . They had other priorities like going on a break and giving medication, and then coming down.

Well when I first came here I was stunned at how long it was taking people to get people out of theatre, I couldn't believe it, I'd never met that before, so long and so bad.

Sometimes the theatre couldn't bring a patient out from theatres because there was no space, and I would either ask one of my nurses to take the patient to the ward, if there's a bed, they would take the patient to the ward to create a space for a patient in theatres. It was as bad as that.

Didn't have feedback or anything at previous trusts.

Before this project started I felt like I didn't have any support to manage this problem, and when [project lead] started with slowly looking into the issue and then developing into this project it has kind of provided me with a support system of trying to find out where the bottlenecks are with regards to the discharge of patients to the wards.

So no, I think it's very useful and it shows . . . because before looking at how long it was taking, an hour and a half to get somebody down to a particular ward, I mean that's appalling.

I have talked to [service manager] about possibly meeting her to look at maybe streamlining things.

Impact of the project

The introduction of continuous monitoring and feedback of patient transfer data collected in the PACU was perceived to have had a positive impact on patient transfer and upstream perioperative service efficiency. Interviewees described how the initiative allowed them to identify areas requiring improvement. The perceived mechanism of this was via stimulating and increasing professional awareness.

I think if say we always used to get the kids within 10 minutes, and then it suddenly changed to like 20 minutes that's quite a big change so we could definitely look on the ward to see whether that was staffing issues or whether the nurses just didn't want to go down straightaway or didn't feel they needed to go down straightaway. So yeah, I think if there was a big change we could look into that.

They're very informative, they're helpful to see how long your waiting times are, and I could see that my area wasn't performing very well.

I think it made me aware of how long our patients were really in theatre, in recovery before they came back.

It's good to know where you are and where you can improve, and it's also good to see where wards are finding it difficult and they are improving.

Suppose if there was an area that was always scoring low then they'd need to look into why that was happening and see if they could address the reasons.

I think it's been very, very useful and shows where wards need to improve everywhere, so yeah.

Well yeah, they give me the actual times which I wouldn't be 100% aware or if it wasn't for the reports.

They're very informative, they're helpful to see how long your waiting times are, and I could see that my area wasn't performing very well, but it is one thing on a long list of things I'd like to improve.

So over the time it has raised awareness of the ward staff that they need to come down and get patients as soon as they are called.

Engagement

Stakeholders engaged with the project and peers via active data dissemination and verbal communication of results. Acknowledgement of successes and efforts was perceived to be of particular value. Some leads were propelled by financial gains: meeting trust targets and avoiding breaches and fines.

The, you know, biggest satisfaction is just to show it [the feedback] to staff and say thank you.

They [the research team] will send me a message saying, 'Well done'.

And they help us as well to avoid breaches . . . which is a very important issue because obviously hospital's fined £10,000 per breach.

Evaluation of the initiative

Although overall support for the initiative was high, interviewees identified potential areas for improvement. Many interviewees perceived that 'time nursing staff are away from ward', 'porter waiting times' and real-time information, 'what's happening in the ward' impact figures and should be added to metrics recorded. There were also further concerns with data accuracy.

What I find isn't measured is the length of time that a nurse is away from the ward. Because it can vary from anything from 20 to 45 minutes.

It's been very useful, although as I say I do sometimes dispute whether it's accurate by what time we leave the ward.

Well I think it's important but we've also, I think we've got to look at what's happening in the actual ward at the time.

We found that *t*-test analysis revealed varying results from individual wards. Some showed significant improvements in patient transfer, for example [$t(1313) = 3.85; p < 0.01$], [$t(897) = 2.84; p < 0.01$] and [$t(499) = 2.51; p < 0.05$]. In contrast, other wards significantly worsened, for example [$t(3008) = -6.14; p < 0.01$] and [$t(2512) = -4.61; p < 0.01$]. Finally, some wards showed no significant difference after implementation of the initiative, for example [$t(2976) = 0.46; p > 0.05$], [$t(462.05) = 0.325; p > 0.05$] and [$t(545.54) = 0.30; p > 0.05$].

Discussion

This component of the evaluation supports the value of a continuous patient transfer monitoring and feedback initiative to improve perioperative efficiency, service delivery and patient/staff satisfaction. Qualitative analysis of the views of end-users within the initiative demonstrates the many local and systemic factors necessary for timely patient transfer and improved perioperative patient flow. It additionally suggests mechanisms by which the data feedback intervention implemented might impact on local processes, staff attitudes and, ultimately, WWT.

As an intermediary step in the perioperative pathway, PACU is susceptible to delays and productivity losses, necessitating an efficient workstream.¹⁹⁷ Interviewees recognised the impact of the initiative on theatre productivity, through the potential to reduce theatre backlogging and positively impact finances. Support for the data feedback was widespread within the qualitative analysis. The interventional methods used – continuous monitoring and quantitative measures – have been advocated to improve patient flow.¹⁹⁸ Furthermore, WWT as a metric has been identified as one of the main theatre-planning performance measures.¹⁸⁹

Receipt of the feedback was viewed to increase awareness, with secondary benefits on hospital productivity. More specifically, interviewees utilised the feedback to identify departmental weaknesses and areas in which improvements could be made. The feedback reports were actively disseminated within departments, with verbal communication of results. Interviewees reported how acknowledgement of achievements and efforts stimulated improvements and engagement. In particular, poor management of bed resources and the financial implications of this on the trust were perceived to be areas that could be improved on.

In line with the barriers identified on receipt of feedback, literature highlights how service efficiency and operating throughput is multifactorial, influenced by, for example, bed availability, emergency admissions and ward distance from theatre.¹⁸⁹ Resources, especially bed availability, were a strong theme throughout the interviews, identified as one of the main barriers to achieving an optimal WWT. This is consistent with existing literature; Samarth *et al.* found that delays in bed availability accounted for 99.6% of delayed patients leaving PACU. Variations in length of stay have been reported as directly related to delays.¹⁹⁷ In concordance with the perceived pressures that emergency cases place on bed resources, Cardoen *et al.* described how such 'uncertainty' negatively impacts productivity, and should be accounted for in theatre planning.¹⁸⁹

Ineffective communication, team skills and patient discharge were also identified as barriers. In particular, discrepancies in professional judgement and poor interdisciplinary teamwork were regarded as significant barriers to positive change. This is supported by Mazzei *et al.*, who report that different disciplines harbour perceptual differences to the causes of delays.¹⁹⁰ The evidence base shows that bed allocations and communication, in particular liaison through downstream pathways and other professionals, are critical roles within the perioperative patient flow process.¹⁹⁵ The mainstay of effective communication in PACU has been shown to involve channelling from leaders, for example the charge nurse, to other hospital areas.¹⁹⁷ Moreover, as a focus of governmental policy, ineffective discharge strategies have long been identified as an area for improvement. A strong dependency was found between patient discharge, bed resources and WWT, highlighting the need to optimise these key pathways. In particular, the concepts of nurse-led discharge, specific discharge times and education were recommended. A further theme was the notion that professionals do not currently understand 'upstream effects', that is the consequences of their actions, resulting in a compelling need to educate individuals to consider the impact and system as a whole.

Qualitative analysis did, however, reveal facilitators to achieving an optimal WWT. Overarching this lay the themes of effective communication and co-ordination, for example by bed managers in addition to the need for active contribution by clinical managers. This was perceived to positively influence teamwork, efficiency, satisfaction and productivity. Although most ward leads described the initiative as useful and recognised the importance of active clinical engagement, significant improvements in WWT were reflected in only three of the eight wards studied. This is in contrast to Overdyke *et al.*,¹⁹⁰ who found that an awareness education initiative improved operating room efficiency overall, in terms of reducing turnover time by 16 minutes. In line with the qualitative themes, this study, among others, described how delays were mainly due to staff unavailability, patient transportation and congestion.

To explain these results, many factors need to be taken into account. Contextual influences of wards will naturally impact WWT; these have been shown to influence interprofessional interactions, motivation and skill to achieve outcomes as well as organisational readiness for change.^{190,199} Some wards showed no

overall change in performance, yet displayed consistent high achievement in the metrics monitored. Research shows that improvement post feedback implementation is more substantial if prior performance is poor.¹⁹⁴ Other wards had more active managers, financial incentives, shorter distances to PACU and higher staff-to-patient ratios. The impact of hospital layout and ward location have been identified as potential areas to target to improve patient flow;²⁰⁰ admittedly, these factors cannot be modified easily.

Effective feedback is that feedback which promotes positive development. Although support for the initiative was high and awareness was raised, respondents highlighted areas of the initiative which required amendments. Qualitative analysis revealed the need for accurate feedback, concerns about data accuracy and the desire for additional metrics, such as real-time information. Studies have shown that effective feedback needs to be credible, specific and account for context.^{105,135,156,191,196}

Although stakeholders valued the initiative, they expressed the requirement for further organisational changes to aid progression. The requirement for a culture change and a system-wide approach were the main themes which emerged regarding sustainability of the project. In concordance with the qualitative analysis, Karvonen *et al.*²⁰⁰ stated that improvements in patient flow can be made via organisational change, for example through optimisation of job function and management. Human factors, influencing people and their behaviour, play a role. Intertwined within this are environmental factors, teamwork and process design. The application of human factors to health care has been shown to positively influence safety culture, teamwork, communication and systems design.²⁰¹ Interviewees described how they envisaged change to be made subsequent to the data feedback; however, more support from senior trust managerial levels was required. These issues are also reflected by Mann *et al.*,¹⁵⁰ who showed that sustained, valid feedback, improvements in safety culture, involvement of senior leaders and a shared goal were the main factors required for feedback project sustainability. Senior management commitment has been shown to be central to development in this area.²⁰¹ Strong parallels have been seen between managerial values and organisational climate.²⁰²

Although there is strong consensus of support for the initiative, the interviews highlighted the need for an infrastructure to support change, and shared goals targeted through a system-wide approach. A strong theme emerged that what was required to achieve this was culture change: proactivity, forward thinking and understanding of hospital dynamics. This exploration has allowed greater understanding of the links between workplace culture and the impact of the initiative, through the identification of staff perceptions, attitudes and needs. The overall benefits of this initiative were perceived to influence patient satisfaction, patient flow and patient experience as well as staff dynamics and the trust as a whole. Despite the need to plan patient flow, and appreciate the impact of clinical pathways on outcomes, further research is required to understand these key elements and provide a reliable framework to optimise productivity.²⁰³ A prominent research agenda for quality improvement is the identification of methods that sustain a successful project.¹⁵⁰ We have isolated the need for active leadership, effective communication and the support of higher managerial levels, which, in combination with an overall climate for patient service, safety and improvement, form the multidimensional construct required.¹⁹⁹ In order to influence patient outcomes and reap economic benefits, we need to plan patient flow.

In conclusion, this mixed-methods evaluation supports the value of performance-indicator feedback to improve perioperative productivity. We have described local barriers and facilitators; such end-user perspectives will be invaluable in the future design and implementation of initiatives targeting productivity, efficiency and patient flow. Overall, a multifaceted approach addressing system-wide issues is required. Strong managerial leadership to engender improvement, support staff and act as a voice for change is a prerequisite.

The limitations of this work include the small qualitative sample size. Owing to contextual-influences, *t*-test analyses may not be generalisable. However, the qualitative themes are strong and consistent, providing basis for planning interventions.

Chapter 10 Synthesis and conclusions

Taken as a whole, the findings from this evaluation provide rich information concerning the effects of a comprehensive, long-term anaesthetic quality monitoring and feedback initiative on multiple dimensions of service performance. Furthermore, they provide insight into the process of development that took place within this initiative and of interactions between context, intervention and user, and document the experiences and perceptions of the anaesthetists who participated as end-users and codesigners of the feedback.

Based on review of the relevant literature in quality improvement, it was hypothesised that the mechanisms of effect of this intervention would operate at two levels: (1) stimulation of local culture change, promoting more open and constructive use of feedback on quality of care delivered at unit level, and (2) stimulation of personal professional behaviour change through normative comparisons with colleagues and enhanced insight into variations and outcomes associated with personal practice. A mixed-methods design with quasi-experimental and realist qualitative investigative components provided the ideal approach for exploring these hypotheses.

In the quasi-experimental evaluation of the impact of the initiative on perioperative quality indicators using ITS models, three specific hypotheses were explored concerning the effects of feedback. Hypothesis 1 was concerned with the effect of implementing basic feedback where no prior feedback existed. It stated that implementation of routine departmental anaesthetic quality monitoring and basic feedback at the level of the individual consultant anaesthetist would improve professional practice and anaesthetic care. Improvement would be detectable by positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality indicators.

The observed response to the implementation of basic feedback in the quality indicators assessed was, on the whole, limited, and the hypothesised benefits of implementing basic feedback were generally not observed in the data. While the average weekly temperature of patients arriving in recovery increased in response to the onset of basic feedback, in contrast, the weekly proportion of patients arriving in recovery with a temperature under 36 °C increased between the baseline and basic feedback condition, too. Patient-reported QoR Scale scores showed no response to the implementation of basic feedback, with evidence to suggest that the proportion of patients reporting high scores may actually have decreased following the implementation of basic feedback. Patient-reported freedom from severe postoperative pain did not respond to the onset of basic feedback, and the proportion of patients reporting freedom from postoperative nausea appeared to decline between the baseline and basic feedback epochs. No significant effects of the implementation of basic feedback were observed for either SSI rate or 30-day postoperative mortality rate. For the vast majority of measures, hypothesis 1 is, therefore, rejected, with the exception of the apparent effect of introduction of basic feedback on the average patient temperature on arrival in recovery. Examination of longitudinal variation in the baseline data set, however, offered a potential alternative explanation for the largely negative results of implementing basic feedback, in terms of issues with the reliability of data collection and data administration in the earlier phases of the study.

Hypothesis 2 was concerned with the effect of implementing enhanced feedback in a group that had been receiving routine basic feedback previously. It stated that provision of enhanced feedback including detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data and institution-wide dissemination, would facilitate personal professional learning from case experience and stimulate further improvement in quality of anaesthetic care.

Escalating the intensity of feedback through implementation of an enhanced feedback protocol had a positive effect across a greater range of measures than implementation of basic feedback. After implementation of enhanced feedback, patients were on average warmer on arrival in recovery, though the rate of change in mean patient temperature got slightly worse. Again, the mean patient temperature and proportion of

patients with a temperature below 36 °C metrics did not show consistent results for the effects of enhanced feedback, with the latter demonstrating a detrimental response in the rate of change over time. Both the mean patient-reported QoR Scale score and proportion of patients reporting high-quality recovery showed improvement in rate of change, but not level, between the basic and enhanced feedback conditions.

The two measures of postoperative pain demonstrated consistent positive responses to the implementation of the enhanced feedback protocol, with significant improvement in both level and rate of change in the proportion of patients reporting freedom from severe pain and no or mild pain on arrival in recovery. The proportion of patients with nurse-reported absence of nausea similarly increased in response to the implementation of enhanced feedback, coupled with an improvement in the rate of change in this measure, although no significant effect was detected for patient-reported freedom from postoperative nausea. No significant effect of the implementation of enhanced feedback on SSI rate was detected. Thirty-day postoperative mortality appeared to show a complex response to the implementation of enhanced feedback, with a small increase in level coupled with a modest improvement in the rate of change over time.

Hypothesis 3 stated that implementation and escalation of performance feedback at the level of the individual consultant anaesthetist would improve professional practice in anaesthetists ranked lowest at baseline and that any beneficial effects of feedback will be strongest for this subgroup compared with the full cohort. Based on the study findings, the third study hypothesis is supported for the effects of enhanced feedback on anaesthetic quality indicator scores for lower-ranked anaesthetists, in the areas of postoperative pain management, control of postoperative nausea and overall patient-reported QoR.

A simple overview of the time series analysis findings for the primary study hypotheses is provided in *Table 76*.

TABLE 76 Summary of findings from ITSA for primary research hypotheses

Anaesthetic quality indicator	Dependent measure used in analysis	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
Patient temperature	Mean patient temperature on arrival in recovery	Positive	Positive and negative
	Proportion of patients arriving in recovery < 36 °C	Negative	Negative
QoR	Mean QoR Scale score	Not significant	Positive
	Proportion of patients with QoR score above 14	Negative	Positive
Postoperative pain	Proportion of patients reporting freedom from severe pain during recovery	Not significant	Positive
	Proportion of patients reporting no or mild pain on arrival in recovery	Not evaluated	Positive
Postoperative nausea	Proportion of patients reporting freedom from PONV during recovery	Negative	Not significant
	Proportion of patients with nurse-recorded absence of nausea during stay in recovery	Not evaluated	Positive
SSI rate	Proportion of patients with a SSI-specific diagnostic code	Not significant	Not significant
Mortality	Proportion of patients with 30-day mortality following surgery	Not significant	Positive and negative

H, hypothesis.

Positive effects are shaded dark green, negative effects are shaded blue, positive and negative effects are shaded light green and those effects not evaluated are shaded grey.

Taken as a whole, the implementation of enhanced feedback, as an escalation of the intensity of a basic feedback programme, was found to have a significant positive impact on a broad range of perioperative quality indicators, including mean patient temperature and QoR Scale score. The strongest apparent positive impact of enhanced feedback, however, was demonstrated for postoperative pain measures, both self-reported for duration in recovery and on arrival in recovery, and for nurse-recorded freedom from postoperative nausea. The observed positive effects of enhanced feedback on patient temperature data, mean QoR score, and measures of both postoperative pain and postoperative nausea were robust after covariate analysis in which longitudinal variation in disease severity (ASA), patient age and gender were controlled. When further statistical models were fitted based on stricter case inclusion criteria limited to elective GA cases, despite the loss in statistical power the implementation of enhanced feedback was still associated with significant improvement in level and slope in both proportion of patients with nurse-reported freedom from nausea and proportion of patients with freedom from pain on arrival in recovery, lending further confidence in these findings.

To complement the statistical findings, the qualitative workstream components were designed to investigate longitudinal processes and contextual mechanisms operating within the programme, from the broad perspectives of individual, social and local cultural factors. Issues associated with local culture and transparency were evident at both qualitative time points, which suggests that the intervention was not always existing in an entirely open and constructive environment. There was certainly an ongoing need to avoid associations with performance management if the intervention was to be trusted and engaged with by clinician end-users. However, it was noticeable that anaesthetists became more open with their data over time and were less demanding of complete anonymity. This was demonstrated through their involvement and positive appraisal of specialty-level group work and interaction with peer experts. In fact, there was a high demand for social interaction around the feedback in general and the need for this to be formalised was regularly expressed. It was considered to be hugely important to have a peer leading the project and this may go some way in explaining why the culture surrounding feedback became more open over time. Clinicians found it easier to place their trust in a fellow anaesthetist who was creating a collective culture for improvement than in somebody who was external to the department.

Interviewees put forward a multitude of use cases that demonstrate utilisation of the performance data to make improvements and test the effect of changes to behaviour on outcomes. This was particularly evident at time point 2, which suggests that the initiative had a cumulative effect and may explain the apparent effectiveness of the second feedback phase in influencing perioperative indicators. This is to be expected from an iterative quality improvement intervention with phased implementation. As time passes, end-users are more able to identify the benefits in having and using the feedback data, which leads to an increase in engagement. It was evident in our analysis that benefits go wider than just having the opportunity to improve practice. In particular, levels of engagement were linked to recognising how the data could be used to satisfy the requirements of revalidation, be useful in personal professional appraisals and, therefore, save time and effort on the part of the anaesthetists.

Individuals had certainly engaged with the feedback throughout the project lifespan and clearly valued its presence in their department. However, they also recognised a number of flaws and felt that they needed more personalised feedback if they were able to effectively improve their professional performance in the long term. The need for personal feedback is a key finding of the project. End-users need to be confident that the data they are using are clinically meaningful; if they are not, they are far more likely to engage in processes of message rejection. Receiving and acting on data about one's own individual performance is clearly linked to the concepts of professionalism and excellence. Anaesthetists associated their involvement in the initiative with their professional identity and the need to strive for excellence. The quantification of their performance enabled them to objectively assess their conceptual ideas of excellence, and this sometimes contradicted their professional identity and led to behaviour change.

The results of the survey studies demonstrate that it is possible to develop a valid and reliable survey to capture end-user perceptions on local quality monitoring and workplace climate. In terms of which characteristics of feedback predicted end-user perceptions of usefulness, the findings from this study suggest that anaesthetists perceive a range of factors as important in determining the usefulness of feedback. Specifically, the local departmental context and its support of quality improvement is an important determinant of how instrumental feedback from monitoring quality indicators is likely to be. Furthermore, feedback that is tailored to be relevant to the personal professional practice of the individual clinician is an important predictor of usefulness. In terms of the feedback content and design characteristics that anaesthetists value most, the perceived credibility of the data and the local relevance of the quality indicators are paramount.

In terms of perioperative productivity, the evaluation supports the value of a continuous patient transfer monitoring and feedback initiative to improve perioperative efficiency, service delivery and patient/staff satisfaction. Qualitative analysis of the views of end-users within the initiative demonstrates the many local and systemic factors necessary for timely patient transfer and improved perioperative patient flow. It additionally suggests mechanisms by which the data feedback intervention implemented might impact on local processes, staff attitudes and, ultimately, patient transfer efficiency. Although there is strong consensus of support for the initiative, the interviews highlighted the need for an infrastructure to support change, and shared goals targeted through a system-wide approach. A strong theme that emerged as required to achieve this was that of culture change: proactivity, forward thinking and understanding of hospital dynamics. This exploration has allowed greater understanding of the links between workplace culture and the impact of the initiative, through the identification of staff perceptions, attitudes and needs. The overall benefits of this initiative were perceived to influence patient satisfaction, patient flow and patient experience as well as staff dynamics and the trust as a whole.

It is clear that feedback is a multifaceted interaction between format, focus and recipient, and this should be taken into consideration in the development of interventions alongside recognition of complexity and context. The clinicians involved in this initiative identified numerous benefits of this type of initiative and identified particular factors which support the effective transition from data, to information, to action and improvement. The system was regarded as providing end-users with the evidence that they required to demonstrate their professional ability and identify any areas in which they would benefit from further support and training. The findings from this realist qualitative evaluation of user perceptions and experiences are valuable for designing similar initiatives in the future that will effectively support clinicians in developing their practice. Qualitative views highlight key areas of challenge for existing and future systems, which must be overcome if such systems are to realise their full potential.

Chapter 11 Study recommendations

Although the evaluation was unable to demonstrate a positive impact of implementing a simple, passive monthly feedback report on baseline perioperative indicators, convincing longitudinal benefits were demonstrated where investment in refinement and enhancement of the feedback protocol was made in collaboration with end-users and within a broader framework of active engagement and constructive dialogue centred around the data. Findings from the qualitative research and evaluative survey demonstrate the positive response of clinicians to this type of initiative and their willingness to interact with a sustained and comprehensive information system in order to understand variations in care delivered personally and in collaboration with colleagues.

While the challenges of reliable measurement of patient satisfaction with anaesthesia and the experience of postoperative pain and recovery are well established in the research literature, our research has highlighted the considerable scope and opportunity that exists for future research and development to address the questions of how we can best use the data resulting from such initiatives and how we transform those data into intelligence for specific professional groups.

Recommendations for future research

Our mixed-methods evaluation of an intensive and sustained quality monitoring and feedback programme for anaesthetists describes an intervention model that draws on the audit and feedback interventions that are increasingly described within the implementation science literature and elsewhere.^{149,152,176,204,205} Recent work in implementation science has sought to strengthen the research base for mounting effective audit and feedback interventions through establishing an agenda for future research that highlights key areas such as characteristics of the setting and recipient, cointerventions, mode of delivery and nature of the content of feedback.²⁰⁵ The model described in this report draws on improvement science methods and philosophy (process control and continuous evaluation) and may, therefore, be more accurately conceptualised as a continuous data feedback intervention. Based on our mixed-methods analysis and evaluation, we understand the effects of this model in terms of a positive impact on sociocognitive processes at the level of both the individual and professional groups, as a means of both stimulating a collective response to quality of care issues and modifying personal professional behaviour. The continuous data feedback model, as we implemented it, was additionally noteworthy, we believe, owing to two factors: (1) the strong focus on clinicians as end-users for the information produced by the process and corresponding strong sense of clinical ownership and engagement that the model invoked, and (2) the fact that the collaborative and personal responses to the information provided by the initiative were stimulated by the intervention model itself and without the explicit threat of formal professional sanctions or organisational regulation for perceived suboptimal clinical performance.

While the implementation of this model in our study was specific to anaesthesia, we believe that the broader intervention concept and model is generic and highly portable across specialties and areas of clinical practice. The increasing use of electronic patient record and administration systems within health-care institutions across a broad spectrum of clinical areas provides an important opportunity for development of information processes to ensure that we make optimal 'secondary use' of available administrative data and provide health-care professionals with the intelligence that they need to monitor variation, identify and understand opportunities for improvement and evaluate changes introduced into practice. Furthermore, the continuous data feedback intervention developed in this study is portable in another important way: in its mode of implementation. In our work we used a combination of hospital information systems, administrative databases and manual data collection and entry, followed by manual report generation using commonly available database tools. With increasing implementation and sophistication of electronic record and hospital information systems for clinical data, the potential for integration of a continuous feedback model within the design of such systems is large with considerable scope for future innovation and evaluation. Indeed, at the

time of writing, the authors are aware of a number of parallel data feedback interventions in perioperative service areas that are at varying stages of integration within local IT programmes, including one initiative that seeks to replicate and test the model implemented with this study within a different institutional context directly, albeit with stronger IT support. In terms of future productive directions for research and development linked to extension of the current study, the following are suggested:

1. **Investigation of the degree to which the current model of a continuous feedback intervention will port successfully to other clinical specialties and areas:** we have made a case for the generic nature of this intervention as a portable concept that can be instantiated for a wide range of professional groups, quality indicators and metrics. Within acute care, development and evaluation of this intervention model with surgeons as the target professional group is one viable extension of our current work, building on several reported initiatives in this area to incorporate the social or collaborative features of our intervention model. Intensive care is a similarly inviting prospect owing to the data-rich nature of the intensive care unit environment, and beyond acute care, general practice may provide a further potential environment for application owing to the high case volume and pressure for rapid turnaround, which we know from our research is a precursor of paucity in useful feedback for practitioners. A further viable research question concerns how this intervention model may be effectively implemented for clinical units or departments, multidisciplinary teams and more dispersed collaborations of health-care practitioners. There is considerable scope, at a fundamental level, to replicate this initiative in further similar and dissimilar areas of practice in order to strengthen the evidence base for this concept and broaden scientific understanding of effective design and context for data feedback interventions. At a higher level of analysis, once a corpus of evaluative case studies has been achieved, there exists the possibility of more incisive comparative analysis and synthesis to establish a firmer evidence base within this area.
2. **In-depth costs–benefits analysis of a continuous quality monitoring model compared with conventional discontinuous audit:** the question of whether or not continuous monitoring is more effective as a basis for regulation of quality of care than periodic assessment is central to the case for continuous data feedback interventions. At a national or indeed any superorganisational level, both routine surveillance and periodic inspection are implemented by various agencies for quality assurance purposes. At the level of the local health-care organisation and the clinical unit, clinical audit is established as the norm, with varying schedules dependent on local culture, resources and governance. Continuous monitoring and feedback may well be too resource-intensive for widespread adoption, despite the policy emphasis on quality accounts in recent years. The logical next step is to increase our understanding of the potential costs and benefits of this approach through robust economic evaluation. As has been previously discussed, the opportunities for contemporary IT solutions to the challenges of data management, intelligent real-time analysis and user-centred reporting of clinical data should reduce the burden of implementing and sustaining a continuous monitoring and feedback model, but it remains to be seen whether or not the potential benefits to patients of providing real-time feedback to health-care providers outweigh the costs of implementing such systems.
3. **Investigation of how variations in context relating to the maturity of local electronic record and information systems affect implementation and outcome of data feedback initiatives:** our experience in interacting with clinicians and researchers interested in data feedback interventions in perioperative service areas suggests that the technological basis of implementation is a key potential enabler and warrants further study. Local service contexts may vary in terms of the ready availability of data relating to key quality criteria, data administration and reporting capabilities. Within ‘data-rich’ environments, reliable feedback may be more routinely available. Furthermore, clinical groups used to operating within a data-rich environment may be more receptive to feedback interventions, for reasons highlighted in the current study linked to confidence in the data and the possession of requisite skills and departmental processes for the ready assimilation, interpretation and response to signals in routine data. The study of the interaction between the use of data and the local organisational context is likely to be an important future implementation science dimension for this area of development and hence an important focus for health services research.

4. **Evidence synthesis to support enhanced use of data in quality improvement initiatives:** in order to support the aforementioned research areas and develop an evidence base for data feedback interventions, there is a need for further evidence synthesis, building on some of the established reviews summarised in the initial section of this report. Important insight could be gained from conducting both broad scoping reviews to summarise the diverse multidisciplinary fields of research relevant to the design and implementation of data feedback interventions, along with more systematic synthesis and meta-analysis of published studies and existing systematic reviews that focus on specific topic areas related to this discipline. Two such areas that are particularly relevant to development of the work reported in this study are (1) effective feedback and use of data from quality indicators in health care, and (2) the secondary use of data from routinely maintained and electronic systems to support quality improvement.

In addition to the future focus of research in this area, our study holds important research implications for the methodologies that are employed to evaluate future quality improvement interventions in a health-care setting. While none of the specific methods employed within this study are novel (though use of a sensitivity analytic approach with time series models to investigate variations in case inclusion arising from inconsistencies in our data set had some degree of novelty), the combination of methods in a mixed qualitative/quasi-experimental design represents a contemporary approach to evaluation of complex interventions. The combination of qualitative and quantitative methods, employed over a longitudinal time frame, may serve as a useful exemplar for research and evaluation in evolving fields such as implementation and improvement science, in which understanding how an intervention interacts with its context of implementation is an important component of the evaluation. Owing to the close partnership between clinical leads and academic evaluators within this project, the research team had a level of access to the programme participants that allowed the depth of data collection and analysis required by the research and which is often difficult to obtain in academically led studies. In addition, this study has demonstrated the utility and flexibility of ITSA as a tool for statistical evaluation of longitudinal trends in data resulting from the operation of multiple iterative intervention events, such as the type commonly implemented within quality improvement projects using Plan-Do-Study-Act (PDSA) and similar methodologies to embed changes. While we had to impose some quasi-experimental discipline on the time course for development and implementation of the intervention, the ITS analytic model, when fitted as segmented regression, has been shown to be flexible and capable in the evaluation of this type of serial intervention. Future research in improvement and implementation science might productively consider the ITS design and its data requirements when attempting to understand the effects of longitudinal programmes.

Finally, although patient and public involvement was not explicitly an aim of this study, contact with patients at collaborative forums has suggested a number of areas in which future development might potentially utilise lay and expert patient perspectives to help shape future perioperative patient satisfaction and feedback initiatives. During the course of the NIHR CLAHRC project which funded implementation of this initiative, the research team had the opportunity to film a short video concerning the patient experience of recovery from surgery, interspersed with researcher and clinician commentary, which linked our patient representative's experience to clinical and academic perspectives on a positive patient experience and how the proposed intervention mechanism might support improvement in patient experience. This integrated narrative formed an important communication device for dissemination of the aims of the project both to the public and clinical audiences. In our view, there exists considerable scope to engage patients in processes to develop more effective feedback concerning the patient experience of recovery from surgery to inform the development of patient satisfaction measures and to facilitate communication between patient and anaesthetist in the pre- and postoperative periods. The patient's experience of recovery and perioperative care is an important dimension of performance of the care delivery process and should be promoted in the context of existing clinical and productivity metrics for feedback.

Implications for practice

The study findings and conclusions hold many implications for developments in practice associated with anaesthetic quality monitoring and feedback in particular, and for the successful implementation of clinical data-driven continuous quality improvement more generally. Based on the process evaluation and analysis of implementation factors reported in this study, there exists considerable transferable learning that might support future efforts to translate research knowledge in this area into practice by introducing similar interventions into an applied clinical setting.

We have outlined what we see as the important conceptual features of the intervention model evaluated in this research in the preceding section, as part of the stimulus for future research replication and development in this area. Research questions aside, our experience in developing and implementing this model suggests several important points for development of effective data feedback interventions and for clinical groups that would seek to implement similar processes as a model for effective data-driven continuous quality improvement.

1. **The potential to develop clinically relevant intelligence from comprehensive, routinely collected and locally available data sources should be developed:** broadly speaking, in the development of this intervention, we were surprised by the volume of routinely collected perioperative administrative data residing in various hospital systems and databases that were not effectively integrated and which could potentially have been harnessed sooner, to provide real-time intelligence on variations in care within a clinically facing information system. Just as the inputs and processes that contribute to variations in the performance of care delivery systems are complex and multivariate in nature, the monitoring and control systems that are put in place to regulate these outputs must demonstrate similar requisite variety. Harnessing a broad spectrum of pre-existing and bespoke process and outcome data sources is, therefore, likely to be imperative to provide facility for the type of disaggregation, control of case mix and provision of contextual data that renders routinely collected quality data useful for specific clinical groups and professional subgroups. Where continuous data collection and feedback may be regarded as resource intensive, considerable efficiency savings can be made where new data processes can be linked and embedded within existing or developing electronic data administration and reporting systems.
2. **The localised relevance of data (from the perspective of the end-user) is paramount in effective feedback design and clinical engagement:** in terms of engaging clinicians in quality monitoring and feedback, our experience was that features of the data set and feedback reports themselves could serve as important engagement mechanisms in capturing the attention of busy health-care professionals and ensuring that individuals engaged with information contained within their reports. Introducing individual case-level breakdown of data to identify individual patients that were in pain, cold or nauseous was useful in personalising and making the information relevant to the day-to-day practice of the anaesthetist. Engagement was further improved by efforts to disaggregate data to the extent that comparisons could be made with other anaesthetists performing similar procedures. Aggregated data did not identify where improvement efforts could be directed and were open to individual anaesthetist bias and subjective interpretation. Developing a subspecialty focus within the data set additionally supported a higher intensity of interaction between individuals in order to share best practice developed by exemplar performers and to develop collective solutions to common quality challenges, supporting the social aspects of our enhanced intervention model which our analysis suggests was essential in achieving improvement in practice. Notably, our regression analysis of the factors predicting perceived usefulness of data feedback similarly demonstrated the primacy of local relevance and credibility above other factors. A further important engagement factor was ultimately aligning the feedback with local departmental processes and broader specialty policy agendas. Demonstrating the usefulness of the feedback report as evidence of fitness to practice for local clinician appraisal processes and for clinician revalidation further stimulated engagement and uptake.

- 3. The benefits of data feedback interventions may only be realised longitudinally and if the process and measures are sufficiently embedded and sustained:** at the level of the individual anaesthetist's data, effect sizes for an initiative of this nature may be small when one aggregates cases across some time period, such as a weekly or monthly average data point. Given the subjective, self-report nature of some patient experience measures used in this project, the resulting data set is inherently noisy, giving rise to common variation extraneous to the performance of the anaesthetist and broader perioperative process. Taking these two factors into account, it is important if the benefits of quality monitoring and feedback are to be realised for improvement purposes, that sufficient data accumulate in the system, at a sufficient frequency of sampling, for meaningful signals to be generated and interpreted. Our feedback reports featured a rolling window of 1 year's monthly data points for anaesthetists, and we would suggest that this time span is reasonable as a recommended minimum. A longer historical perspective may further enhance the ability of individuals and groups to separate signal from noise. Unfortunately, in a dynamic clinical and institutional environment, the factors which govern important parameters in service delivery rarely remain stable for long periods of time. Furthermore, we found that it takes time to develop and embed reliable data collection processes and for variation resulting from artefacts of the data collection process to settle to acceptable levels. For these reasons, it is important to strike the right balance in terms of the time horizon for data feedback and to use common sense in specification of any stable baseline period for comparisons.
- 4. Design of effective data feedback is predicated on sound statistical understanding of the data, intelligent application of analytics and multiple complementary data perspectives:** our work in this area, including multiple iterations of measurement, analysis and reporting processes, has emphasised the importance of these basic principles of effective informatics for the discovery and communication of meaningful patterns in clinical data. We have already commented on the importance of using contextual data to control for case-mix effects, but the importance of selection of an appropriate sampling period is additionally critical for understanding variations. Feedback from report recipients in our study suggested that monthly aggregation of case data was optimal, but this is additionally a matter of statistical consideration. Depending on the intensity of case volume accumulation over time at the level of the individual anaesthetist's case load, the correct sampling frequency is the one that provides a sufficiently robust and stable estimate of the underlying parameter, while maximising the frequency of data points for trending. Furthermore, aggregated summary statistics may artificially smooth a trend line, reducing the apparent effects of important outlying cases. While aggregated statistics such as average scores on a measure provide robust parameter estimates for trending, the information contributed by the presence of extreme outliers is very important information, where the interests of high-quality care and delivery of a reliable service (consistent high-quality care) are concerned. A patient in excruciating pain is not comforted by the knowledge that he or she is in a statistical minority and, similarly, an anaesthetist should not be reassured by an average score that well exceeds the accepted threshold if 1 in 20 consecutive cases falls well short of it. The solution we arrived at in the design of our intervention was to use multiple complementary perspectives on statistical data including varied forms of graphs and signposting reports. Effective data feedback should, therefore, report outliers as well as aggregated measures, should present a longitudinal perspective (variation over time) as well as cross-sectional (comparative peer performance at a single time point) and should include analysis at varied levels of granularity (so as to appraise the wood as well as the trees). Finally, our work emphasised the importance of understanding the psychometric and distributional properties of quality indicators and scale scores as a first step in designing data feedback. Understanding the statistical distribution of metrics and scale scores and their implications for detecting variation and selection of appropriate summary measures is paramount and may necessitate a period of tuning or calibration of the data collection protocol, if floor/ceiling effects or unusual/skewed distributions are to be avoided. Presenting clinicians with unduly noisy or unstable data is likely to hamper engagement, render the data as perceived to be not useful and prolong the time taken for clinicians to develop confidence in and facility with the feedback. We regard one of the critical successes of this project to be the demonstration of how routinely collected data can be translated into information that clinicians could use to drive improvement, through the careful design of a reporting mechanism.

5. **The social component of collaborative quality monitoring, feedback and action planning should not be overlooked in the design of clinical performance information systems:** the message that technical innovations in clinical information systems are embedded within a social or human and organisational context comes through strongly in our study findings. Indeed, we were not able to demonstrate any positive overall effect of implementation of passive data feedback provision and it was not until an enhanced model with 'softer' social components was implemented that our data feedback intervention demonstrated capability as an improvement mechanism. Our study suggests that effective data feedback interventions are true 'complex' interventions in that they must include processes for engagement of the local professional group, collaborative problem solving, sharing of best practice experience and development of a shared sense of responsibility for challenges in care delivery (in addition to that instilled within the individual). Consequently, feedback processes must be integrated within collaborative departmental review and monitoring processes, and time and support must be provided for interpretation, formulation and implementation of responses to the data, at both team and individual levels. Where the social dimension of this intervention is considered, a key issue is the level of anonymity that should be preserved within reports that employ peer comparisons as a stimulus for improvement. We suggest that, initially, an anonymised approach would be appropriate in the majority of circumstances and that this is linked to the importance of providing a safe and supportive environment for individuals to reflect on and develop confidence in the data. We additionally recommend that implementation of a data feedback initiative should be reinforced by messages from departmental leads that individuals will be appraised based on the degree to which they engage with personal performance data and respond to opportunities to enhance practice, rather than on the basis of any absolute performance scoring or formal peer comparison, once a threshold for agreed basic standards of care have been met. Our intervention model was designed to be supportive/formative, rather than punitive or an exercise in performance management. Interestingly, anaesthetists within our programme reported experiencing substantial perceived social pressure to conform to norms for acceptable performance within the department, from the stimulus of an anonymised peer-comparative report and without the threat of formal sanctions for statistically deviant performance. Furthermore, anaesthetists reported that framing this initiative as a managerially led financial or productivity drive would have rendered it less appealing than a clinically led exercise centred on patient experience. Of course, getting the social component of this complex intervention model right means ensuring that a local context exists conducive to open discussion of performance data. The local context is affected by cultural factors reinforced by specialty policy and agenda, by the messages employees receive from their employing institution, by local working climate (the shared values, practices and working norms of a unit or department) and by the interpersonal atmosphere and dynamics within teams. These factors vary as a function of context, meaning that a common rigidly defined intervention model may not fit all scenarios and features of the intervention, including the level of anonymity and protection provided for individuals, may need to be manipulated to fit a specific institutional context. Interestingly, over time the espoused need for anonymity in the feedback reports and broader discussions of performance within our study group declined, partly as a function of increasing trust and confidence in the data and the aims of the initiative, and partly in response to the desire to identify consistent high performers from whom to learn. The declining emphasis placed on anonymity in a data feedback initiative over time may, therefore, be an important indication of successful embedding, clinician engagement and confidence, and the development of a local context conducive to data-driven continuous improvement.

Summary of recommendations for development

The research findings give rise to the following implications and recommendations for development of data feedback interventions and scientific understanding of how they operate within health care:

1. The research suggests that quality monitoring and feedback interventions of the type implemented in this study represent an important quality improvement mechanism, especially where investment is made in their long-term development and sustainment.
2. The findings suggest that the design of feedback, its perceived intent, fitness for purpose and context of use are all important considerations for success.
3. It is essential to involve end-users in the development of the feedback system, not only at conception, but to foster an ongoing sense of ownership of the data and a willingness to interact with them.
4. It is important to pair passive data dissemination with support, active engagement and opportunities for intra- and interprofessional dialogue, concerning how to respond to evidence of variations and opportunities for improvement.
5. Continuous feedback can make the natural variation inherent in human-intensive processes, such as health care, visible to improvement efforts. In so doing, subjective and intangible phenomena, such as patient satisfaction, may be objectified for more constructive conversations, enhanced shared decision-making and better control.
6. In the development of monitoring and feedback systems, appropriate attention must be given to how data are used and converted into information for specific user groups, such as clinicians, rather than simply focusing on what to measure and how reliable those measures are.
7. The success of quality improvement interventions such as the data feedback initiative studied in this research should be evaluated using multiple dimensions, including social, organisational and professional outcomes, in addition to clinical end points.
8. While downstream postoperative outcomes may be insensitive to the effects of an anaesthetic quality feedback intervention, process-of-care measures, such as those associated with postoperative pain management, nausea and perioperative normothermia, are more receptive.
9. Within the health informatics field, considerable scope exists beyond this project to further test evolving theory and practice from improvement science and industrial process control related to how data can be used to support continuous improvement in process-based operations.
10. The trend towards a shift away from intermittent, snapshot audits of practice in favour of a continuous monitoring and continuous improvement model within health-care organisations should be the subject of further investigation in terms of its implications for patients and quality of care.

Acknowledgements

We are indebted to 115 consultant anaesthetists at Imperial College Healthcare. Above all, the authors gratefully acknowledge the commitment and generosity of 44 consultant anaesthetists based at St Mary's Hospital, whose willingness to participate in a long-term experiment to see if feedback could place another window on the system made this work possible. In this area, the academic leads wish to credit our co-author, Dr Glenn Arnold, for his persistent vision and drive.

The authors gratefully acknowledge the contributions of the many clinicians, recovery nurses, service managers, patient representatives and academics who have supported this project in one form or another. We specifically wish to mention the following individuals:

Dr Christopher Stonell.

Dr Mark Sacks.

Nyasha Chironga.

Catherine Riley.

Dr William Harrop-Griffiths.

Dr Helgi Johannsson.

Anne McKenna.

Dr Mark Palazzo.

Dr Alan Poots.

Cathy Howe.

Professor Derek Bell.

Professor Charles Vincent.

Dr Ramani Moonesinghe.

Professor Mike Grocott.

Jill Lloyd.

The views and conclusions outlined in this report remain the sole responsibility of the authors.

Contributions of authors

Dr Jonathan Benn (principal investigator, Imperial College London) had overall responsibility for the design and delivery of the project, supervised the researchers, and drafted and edited the final report.

Dr Glenn Arnold (co-investigator, Imperial College London) was a member of the steering group, acted as clinical lead for the feedback programme, conducted data collection and analysis, and reviewed and provided input to sections of the report.

Ms Danielle D’Lima (researcher, Imperial College London) conducted data collection and analysis, and drafted and edited the final report.

Mr Igor Wei (researcher, Imperial College London) conducted data collection and analysis, acted as operational assistant to the feedback programme and contributed to drafting of sections of the report.

Ms Joanna Moore (researcher, Imperial College London) conducted data collection and analysis, and contributed to drafting of sections of the report.

Ms Floor Aleva (visiting researcher, Imperial College London) conducted data collection and analysis and contributed to drafting of sections of the report.

Professor Andrew Smith (co-investigator, Lancaster University) was a member of the steering group, acted as academic advisor to the project, and reviewed and provided input to sections of the report.

Dr Alex Bottle (co-investigator, Imperial College London) was a member of the steering group, acted as statistical and academic advisor to the project, supervised one of the researchers, and reviewed and provided input to sections of the report.

Dr Stephen Brett (co-investigator, Imperial College London) was a member of the steering group, acted as senior academic advisor to the project, and reviewed and provided input to sections of the report.

Publications and dissemination

Papers

Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. *Br J Anaesth* 2012;**109**:80–91.

D’Lima D, Moore J, Bottle A, Brett S, Arnold G, Benn J. Developing effective feedback on quality of anaesthetic care: what are the important characteristics? *J Health Serv Res Policy* 2015;**20**:26–34.

D’Lima DM, Moore J, Bottle A, Brett SJ, Arnold GM, Benn J. Developing effective feedback on quality of anaesthetic care: what are its most valuable characteristics from a clinical perspective? *J Health Serv Res Policy* 2015;**20**(Suppl. 1):26–34.

Translational output

Benn J. Presentation of national survey of use of perioperative indicators to QUARC meeting – Royal College of Anaesthetists, February 2014, London, UK.

Benn J, Arnold G, D’Lima D, Moore J, Wei I, Riley C, *et al.* Quality feedback to clinicians: what indicators and how? Presentation at the Royal College of Anaesthetists Anniversary Meeting, March 2013, London, UK.

Benn J. Use of quality indicators: a national survey of perioperative units. *RCOA Bulletin*, 80, 2013, p. 21.

Arnold G. Quality indicators in anaesthesia. Converting data to information. Presentation at the Association of Anaesthetists of Great Britain and Ireland Winter Scientific Meeting 2013, London, UK.

Benn J, Moppett IK. What is normal or not? *Br J Anaesth* 2013;**110**:497–500.

Benn J, Arnold G. Quality indicators in anaesthesia. Presentation at the Royal College of Anaesthetists/AAGBI Joint Clinical Directors Meeting, 2012, London, UK.

Arnold G. Presentation at Anaesthetic Association of Great Britain and Ireland research seminar, 2012.

Benn J, Arnold G, Wei I, Riley C, Chironga N, Aleva F, *et al*. Using quality indicators in anaesthesia. Presentation at the Royal College of Anaesthetists Anniversary Meeting, 14–15 March 2012, London, UK.

Benn J. Getting maximum impact: Experience in the IMPAQT project. Invited presentation at the NIHR CLAHRC Winter Collaborative Delivery Event, January 2012, Royal College of Physicians, London, UK.

Arnold G. Quality indicators in anaesthesia. Converting data to information. Presentation at the Association of Anaesthetists of Great Britain and Ireland Winter Scientific Meeting 2013, London, UK.

Awards

D’Lima D, Moore J, Arnold G, Benn J. Impact of a local quality monitoring programme in anaesthesia: a mixed methods evaluation. Presentation given at Evidence Based Peri-Operative Medicine (EBPOM) Annual Conference, 2013, London, UK. Received award for best research presentation.

Benn J, Arnold G, D’Lima D, Moore J, Wei I, Poots A, *et al*. Determining the important characteristics of quality monitoring and feedback to improve anaesthetic care: a mixed methods evaluation. Health Services Research Network Annual Forum, 2014. Received the Improvement Science Award, Health Services Research Network (HSRN) Symposium 2014, for best Improvement Science project.

Conference contributions

Benn J, Arnold G, D’Lima D, Moore J, Wei I, Poots A, *et al*. Determining the important characteristics of quality monitoring and feedback to improve anaesthetic care: a mixed methods evaluation. HSRN annual forum, 2014, Nottingham, UK.

Benn J, D’Lima D, Moore J, Wei I, Arnold G. Enhanced feedback from perioperative quality indicators: studying the impact of a complex quality improvement intervention. Presentation at BMJ International Forum, Paris, 2014. *BMJ Qual Saf* 2014;**23**:351–2.

D’Lima D, Moore J, Arnold G, Benn J. Translating data into information for improvement: qualitative evaluation of a perioperative quality monitoring and feedback initiative. Accepted as a poster presentation at the BMJ International Forum on Quality and Safety in Healthcare, 2014, London, UK.

D’Lima D, Moore J, Arnold G, Sacks M, Stonell C, Benn J. Determining the important characteristics of quality monitoring and feedback to improve anaesthetic care. Presentation at ISQUA, October 2013, London, UK.

Moore J, D’Lima D, Arnold G, Sacks M, Stonell C, Benn J. Effective use of quality indicators in anaesthesia: structure and psychometric properties of an evaluative survey. Poster presented at ISQUA, October 2013, London, UK.

D'Lima D, Nelson A, Benn J. Using patient feedback to change clinical behaviour: the anaesthetists' perspective. Poster presented at Division of Health Psychology Annual Conference, 11–13 September 2013, Brighton, UK.

D'Lima D, Moore J, Arnold G, Benn J. Performance feedback to healthcare professionals: what supports behaviour change? Presentation given at European Health Psychology Society Annual Conference, 2013, Bordeaux, France.

D'Lima, D, Moore, J, Arnold, G, Benn, J. Impact of a local quality monitoring programme in anaesthesia: a mixed methods evaluation. Presentation given at Evidence Based Peri-Operative Medicine (EBPOM) Annual Conference, 2013, London, UK.

Moore J, D'Lima D, Arnold G, Sacks M, Stonell C, Benn J. Evaluation of anaesthetist-specific feedback on patients' quality of recovery. Euroanaesthesia, Barcelona, Spain, June 2012.

Benn J, D'Lima D, Moore J, Wei I, Arnold G. Improving the quality and efficiency of perioperative care. Abstract presented at the BMJ Forum for Quality and Safety, London, UK, 2013.

Arnold G, Chironga N, Benn J, Riley C, Sacks M, Stonell C. Monitoring quality of peri-operative care using statistical process control methods. Abstract presented at The Association for Perioperative Practice (AfPP) Annual Congress and Exhibition, Harrogate, UK, 14–16 October 2010.

Benn J, Wei I, Chironga N, Riley C, Aleva F, Arnold G. Development of a process for continuous monitoring and feedback of perioperative quality indicators, 17–20 April, 2012, Paris, France.

Riley C, Benn J, Arnold G, Stonell C. A continuous quality improvement approach to enhancing patients' experience of transitions of care in the post anaesthetic care unit. Abstract presented at the BMJ International Forum on Quality and Safety in Healthcare, 2011, Amsterdam, the Netherlands.

Wei I, Arnold, G, Riley C, Chironga N, Benn J. Development of a system for continuous quality monitoring, feedback and improvement in anaesthetic services. Poster presented at the Patient Safety Congress 2011, Birmingham, UK, 16–18 May 2011.

Benn J, Wei I, Chironga N, Riley C, Arnold G. Continuous monitoring and feedback of anaesthetic quality indicators. Abstract presented at The Association for Perioperative Practice 47th Congress and Exhibition, Bournemouth, UK, 19–21 October 2011.

Wei I, Arnold G, Chironga N, Benn J. Continuous monitoring and feedback for perioperative quality indicators. Poster presented at the Association for Perioperative Practice 47th Congress and Exhibition, Bournemouth, UK, 19–21 October 2011.

References

1. Department of Health. *High Quality Care for All: NHS Next Stage Review Final Report*. London: Department of Health; 2008.
2. Department of Health. *Equity and Excellence: Liberating the NHS*. London: The Stationery Office; 2010.
3. Francis R. *Independent Inquiry into Care Provided by Mid Staffordshire NHS Foundation Trust January 2005 – March 2009*. London: The Stationery Office; 2013.
4. Walley P, Rayment J, Cooke M. *Clinical Systems Improvement in NHS Hospital Trusts and Their PCTs: A Snapshot of Current Practice*. Institute for Innovation and Improvement and The University of Warwick; 2006.
5. Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, et al. Is health care getting safer? *BMJ* 2008;**337**:a2426. <http://dx.doi.org/10.1136/bmj.a2426>
6. Moonesinghe S, Tomlinson A. Quality improvement and revalidation: two goals, same strategy? *Br J Anaesth* 2011;**106**:447–50. <http://dx.doi.org/10.1093/bja/aer052>
7. Rubin P. Revalidation: a General Medical Council perspective. *Clin Med* 2010;**10**:112. <http://dx.doi.org/10.7861/clinmedicine.10-2-112>
8. Myerson K. Appraisal and revalidation need sensitive handling. *Anaesthesia* 2001;**56**:199–201. <http://dx.doi.org/10.1046/j.1365-2044.2001.01972.x>
9. Joint Committee on Good Practice. *Good Practice: A Guide for Departments of Anaesthesia, Critical Care and Pain Management*. London: Royal College of Anaesthetists and Association of Anaesthetists of Great Britain and Ireland; 2006.
10. General Medical Council. *Good Medical Practice*. London: GMC; 2006.
11. Kennedy I. *Learning from Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary 1984–1995*. The Bristol Royal Infirmary Inquiry, 2001.
12. Bent PD, Bolsin SN, Creati BJ, Patrick AJ, Colson ME. Professional monitoring and critical incident reporting using personal digital assistants. *Med J Aus* 2002;**177**:496–9.
13. Bolsin S, Solly R, Patrick A. The value of personal professional monitoring performance data and open disclosure policies in anaesthetic practice: a case report. *Qual Saf Health Care* 2003;**12**:295–7. <http://dx.doi.org/10.1136/qhc.12.4.295>
14. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;**44**(Suppl. 1):66–206. <http://dx.doi.org/10.2307/3348969>
15. Donabedian A. The evaluation of medical care programs. *Bull New York Acad Med* 1968;**44**:117–24.
16. Donabedian A. Criteria, norms and standards of quality: what do they mean? *Am J Public Health* 1981;**71**:409–12. <http://dx.doi.org/10.2105/AJPH.71.4.409>
17. Haller G, Stoelwinder J, Myles P, McNeil J. Quality and safety indicators in anesthesia: a systematic review. *Anesthesiology* 2009;**110**:1158–75. <http://dx.doi.org/10.1097/ALN.0b013e3181a1093b>
18. Gaba DM. Anaesthesiology as a model for patient safety in health care. *BMJ* 2000;**320**:785–8. <http://dx.doi.org/10.1136/bmj.320.7237.785>

19. Findlay G, Goodwin A, Protopapa K, Smith N. *Knowing the Risk: A Review of the Peri-Operative Care of Surgical Patients*. London: National Confidential Enquiry into Patient Outcome and Death; 2011.
20. Whitty P, Shaw I, Goodwin D. Patient satisfaction with general anaesthesia. *Anaesthesia* 1996;**51**:327–32. <http://dx.doi.org/10.1111/j.1365-2044.1996.tb07741.x>
21. Fung D, Cohen M. What do outpatients value most in their anesthesia care? *Can J Anesth* 2001;**48**:12–19. <http://dx.doi.org/10.1007/BF03019808>
22. Heidegger T, Husemann Y, Nuebling M, Morf D, Sieber T, Huth A, et al. Patient satisfaction with anaesthesia care: development of a psychometric questionnaire and benchmarking among six hospitals in Switzerland and Austria. *Br J Anaesth* 2002;**89**:863–72. <http://dx.doi.org/10.1093/bja/aef277>
23. Auquier P, Pernoud N, Bruder N, Simeoni MC, Auffray JP, Colavolpe C, et al. Development and validation of a perioperative satisfaction questionnaire. *Anesthesiology* 2005;**102**:1116. <http://dx.doi.org/10.1097/00000542-200506000-00010>
24. Capuzzo M, Landi F, Bassani A, Grassi L, Volta C, Alvisi R. Emotional and interpersonal factors are most important for patient satisfaction with anaesthesia. *Acta Anaesth Scand* 2005;**49**:735–42. <http://dx.doi.org/10.1111/j.1399-6576.2005.00738.x>
25. Chanthong P, Abrishami A, Wong J, Herrera F, Chung F. Systematic review of questionnaires measuring patient satisfaction in ambulatory anesthesia. *Anesthesiology* 2009;**110**:1061. <http://dx.doi.org/10.1097/ALN.0b013e31819db079>
26. Myles PS, Hunt JO, Nightingale CE, Fletcher H, Beh T, Tanil D, et al. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesth Analg* 1999;**88**:83–90.
27. Myles P, Weitkamp B, Jones K, Melick J, Hensen S. Validity and reliability of a postoperative quality of recovery score: the QoR-40. *Br J Anaesth* 2000;**84**:11–15. <http://dx.doi.org/10.1093/oxfordjournals.bja.a013366>
28. Myles P, Reeves M, Anderson H, Weeks A. Measurement of quality of recovery in 5672 patients after anaesthesia and surgery. *Anaesth Intensive Care* 2000;**28**:276–80.
29. Myles P, McLeod A, Hunt J, Fletcher H. Sex differences in speed of emergence and quality of recovery after anaesthesia: cohort study. *BMJ* 2001;**322**:710–11. <http://dx.doi.org/10.1136/bmj.322.7288.710>
30. National Institute for Health and Care Excellence. *Inadvertent Perioperative Hypothermia*. NICE Clinical Guideline 65. London: NICE; 2008.
31. Macario A. Which clinical anesthesia outcomes are important to avoid? The perspective of patients. *Anesth Analg* 1999;**89**:652.
32. Chung F, Mezei G. Factors contributing to a prolonged stay after ambulatory surgery. *Anesth Analg* 1999;**89**:1352–9.
33. Aubrun F, Paqueron X, Langeron O, Cortat P, Riou B. What pain scales do nurses use in the postanesthesia care unit? *Eur J Anaesthesiol* 2003;**20**:745–9. <http://dx.doi.org/10.1097/0003643-200309000-00012>
34. Dolin SJ, Cashman JN, Bland JM. Effectiveness of acute postoperative pain management: I. Evidence from published data. *Br J Anaesth* 2002;**89**:409–23. <http://dx.doi.org/10.1093/bja/89.3.409>
35. Gan TJ. Risk factors for postoperative nausea and vomiting. *Anesth Analg* 2006;**102**:1884–98. <http://dx.doi.org/10.1213/01.ANE.0000219597.16143.4D>

36. Kurz A, Sessler DI, Lenhardt R. Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization. *N Engl J Med* 1996;**334**:1209–16. <http://dx.doi.org/10.1056/NEJM199605093341901>
37. Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J, *et al.* Clinical indicators: development and applications. *Neth J Med* 2007;**65**:15–22.
38. Clarke A, Rao M. Developing quality indicators to assess quality of care. *Qual Saf Health Care* 2004;**13**:248–9. <http://dx.doi.org/10.1136/qshc.2004.011312>
39. Evans SM, Lowinger JS, Sprivulis PC, Copnell B, Cameron PA. Prioritizing quality indicator development across the healthcare system: identifying what to measure. *Intern Med J* 2009;**39**:648–54. <http://dx.doi.org/10.1111/j.1445-5994.2008.01733.x>
40. Johnston G, Crombie I, Alder E, Davies H, Millard A. Reviewing audit: barriers and facilitating factors for effective clinical audit. *BMJ* 2000;**9**:23. <http://dx.doi.org/10.1136/qhc.9.1.23>
41. Bowie P, Bradley NA, Rushmer R. Clinical audit and quality improvement – time for a rethink? *J Eval Clin Prac* 2012;**18**:42–8. <http://dx.doi.org/10.1111/j.1365-2753.2010.01523.x>
42. Audet A-MJ, Doty MM, Shamasdin J, Schoenbaum SC. Measure, learn, and improve: physicians' involvement in quality improvement. *Health Aff* 2005;**24**:843–53. <http://dx.doi.org/10.1377/hlthaff.24.3.843>
43. Davies H, Powell A, Rushmer R. *Healthcare Professionals' Views on Clinician Engagement in Quality Improvement: A Literature Review*. London: The Health Foundation; 2007.
44. Parand A, Burnett S, Benn J, Iskander S, Pinto A, Vincent C. Medical engagement in organisation-wide safety and quality-improvement programmes: experience in the UK Safer Patients Initiative. *Qual Saf Health Care* 2010;**19**:1–5. <http://dx.doi.org/10.1136/qshc.2009.036368>
45. Hysong S, Best R, Pugh J. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006;**1**:9. <http://dx.doi.org/10.1186/1748-5908-1-9>
46. Benn J, Koutantji M, Wallace L, Spurgeon P, Rejman M, Healey A, *et al.* Feedback from incident reporting: information and action to improve patient safety. *Qual Saf Health Care* 2009;**18**:11–21. <http://dx.doi.org/10.1136/qshc.2007.024166>
47. Shute VJ. Focus on formative feedback. *Rev Educ Res* 2008;**78**:153–89. <http://dx.doi.org/10.3102/0034654307313795>
48. Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;**44**:101–8. <http://dx.doi.org/10.1111/j.1365-2923.2009.03546.x>
49. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care* 2006;**15**:433–6. <http://dx.doi.org/10.1136/qshc.2006.018549>
50. Ramaprasad A. On the definition of feedback. *Behav Sci* 1983;**28**:4–13. <http://dx.doi.org/10.1002/bs.3830280103>
51. Davies HTO, Nutley SM. Developing learning organisations in the new NHS. *BMJ* 2000;**320**:998–1001. <http://dx.doi.org/10.1136/bmj.320.7240.998>
52. van der Veer S, de Keizer N, Ravelli A, Tenkink S, Jager K. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. *Int J Med Inform* 2010;**79**:305–23. <http://dx.doi.org/10.1016/j.ijmedinf.2010.01.011>
53. De Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. *Int J Qual Health Care* 2009;**21**:119–29. <http://dx.doi.org/10.1093/intqhc/mzn059>

54. Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2003;**3**:CD000259. <http://dx.doi.org/10.1002/14651858.cd000259>
55. Chaillet N, Dube E, Dugas M, Audibert F, Tourigny C, Fraser WD, et al. Evidence-based strategies for implementing guidelines in obstetrics: a systematic review. *Obstetr Gynecol* 2006;**108**:1234–45. <http://dx.doi.org/10.1097/01.AOG.0000236434.74160.8b>
56. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach* 2006;**28**:117–28. <http://dx.doi.org/10.1080/01421590600622665>
57. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *BMJ* 1991;**303**:398–402. <http://dx.doi.org/10.1136/bmj.303.6799.398>
58. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. *Qual Saf Health Care* 2004;**13**:26–31. <http://dx.doi.org/10.1136/qhc.13.1.26>
59. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ* 1995;**153**:1423–31.
60. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001;**39**(Suppl. 2):II2–45. <http://dx.doi.org/10.1097/00005650-200108002-00002>
61. Eason K. Local sociotechnical system development in the NHS National Programme for Information Technology. *J Inform Technol* 2007;**22**:257–64. <http://dx.doi.org/10.1057/palgrave.jit.2000101>
62. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. *J Biomed Inform* 2010;**43**:159–72. <http://dx.doi.org/10.1016/j.jbi.2009.07.002>
63. Grimshaw JM, Eccles MP, Walker AE, Thomas RE. Changing physicians' behavior: what works and thoughts on getting more things to work. *J Contin Educ Health Prof* 2002;**22**:237–43. <http://dx.doi.org/10.1002/chp.1340220408>
64. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64. <http://dx.doi.org/10.1136/qhc.12.6.458>
65. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 2007;**16**:203–7. <http://dx.doi.org/10.1136/qshc.2006.020776>
66. Carey RG. *Improving Healthcare with Control Charts: Basic and Advanced SPC Methods and Case Studies*. Milwaukee, WI: ASQ Quality Press; 2003.
67. Curran ET, Benneyan JC, Hood J. Controlling methicillin-resistant *Staphylococcus aureus*: a feedback approach using annotated statistical process control charts. *Infect Control Hosp Epidemiol* 2002;**23**:13–18. <http://dx.doi.org/10.1086/501961>
68. Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. San Francisco, CA: Jossey-Bass Publishers; 1996.
69. Mohammed M, Cheng K, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;**357**:463–7. [http://dx.doi.org/10.1016/S0140-6736\(00\)04019-8](http://dx.doi.org/10.1016/S0140-6736(00)04019-8)
70. Mohammed MA. Using statistical process control to improve the quality of health care. *Qual Saf Health Care* 2004;**13**:243–5. <http://dx.doi.org/10.1136/qshc.2004.011650>

71. Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008;**17**:137–45. <http://dx.doi.org/10.1136/qshc.2004.012047>
72. Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf* 2011;**20**:46–51. <http://dx.doi.org/10.1136/bmjqs.2009.037895>
73. Provost LP, Murray S. *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco, CA: Jossey-Bass; 2011.
74. Batalden P, Davidoff F, Marshall M, Bibby J, Pink C. So what? Now what? Exploring, understanding and using the epistemologies that inform the improvement of healthcare. *BMJ Qual Saf* 2011;**20**(Suppl. 1):i99–105. <http://dx.doi.org/10.1136/bmjqs.2011.051698>
75. Perla RJ, Parry GJ. The epistemology of quality improvement: it's all Greek. *BMJ Qual Saf* 2011;**20**(Suppl. 1):i24–7. <http://dx.doi.org/10.1136/bmjqs.2010.046557>
76. Benn J, Burnett S, Parand A, Pinto A, Vincent C. Factors predicting change in hospital safety climate and capability in a multi-site patient safety collaborative: a longitudinal survey study. *BMJ Qual Saf* 2012;**21**:559–68. <http://dx.doi.org/10.1136/bmjqs-2011-000286>
77. Wachter RM, Pronovost PJ. The 100,000 Lives Campaign: a scientific and policy review. *Jt Comm J Qual Patient Saf* 2006;**32**:621–7.
78. Locock L. Healthcare redesign: meaning, origins and application. *Qual Saf Health Care* 2003;**12**:53–7. <http://dx.doi.org/10.1136/qhc.12.1.53>
79. Shortell S, Bennett C, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q* 1998;**76**:593–624. <http://dx.doi.org/10.1111/1468-0009.00107>
80. Varughese A, Hagerman N, Kurth C. Quality in pediatric anesthesia. *Pediatr Anesth* 2010;**20**:684–96. <http://dx.doi.org/10.1111/j.1460-9592.2010.03329.x>
81. Runcie C. Assessing the performance of a consultant anaesthetist by control chart methodology. *Anaesthesia* 2009;**64**:293–6. <http://dx.doi.org/10.1111/j.1365-2044.2008.05762.x>
82. Fasting S, Gisvold S. Statistical process control methods allow the analysis and improvement of anesthesia care. *Can J Anesth* 2003;**50**:767–74. <http://dx.doi.org/10.1007/BF03019371>
83. Mayer EKBM, Bottle AP, Rao CB, Darzi AWHF, Athanasiou TPF. Funnel plots and their emerging application in surgery. *Ann Surg* 2009;**249**:376–83. <http://dx.doi.org/10.1097/SLA.0b013e31819a47b1>
84. Coory M, Duckett S, Sketcher-Baker K. Using control charts to monitor quality of hospital care with administrative data. *Int J Qual Health Care* 2008;**20**:31–9. <http://dx.doi.org/10.1093/intqhc/mzm060>
85. Plsek PE. Quality improvement methods in clinical medicine. *Pediatrics* 1999;**103**:e203–14.
86. Thor J, Lundberg J, Ask J, Olsson J, Carli C, Harenstam KP, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007;**16**:387–99. <http://dx.doi.org/10.1136/qshc.2006.022194>
87. Powell A, Rushmer R, Davies H. *A Systematic Narrative Review of Quality Improvement Models in Health Care*. Edinburgh: Quality Improvement Scotland; 2009.
88. Boaden R, Harvey G, Moxham C, Proudlove N. *Quality Improvement: Theory and Practice in Healthcare*. Coventry: NHS Institute for Innovation and Improvement; 2008.

89. Schouten LMT, Hulscher MEJL, Everdingen JJEv, Huijsman R, Grol RPTM. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ* 2008;**336**:1491–4. <http://dx.doi.org/10.1136/bmj.39570.749884.BE>
90. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, *et al.* An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;**355**:2725–32. <http://dx.doi.org/10.1056/NEJMoa061115>
91. Bosch M, Weijden Tvd, Wensing M, Grol R. Tailoring quality improvement interventions to identified barriers: a multiple case analysis. *J Eval Clin Pract* 2007;**13**:161–8. <http://dx.doi.org/10.1111/j.1365-2753.2006.00660.x>
92. Hulscher M, Laurant M, Grol R. Process evaluation on quality improvement interventions. *Qual Saf Health Care* 2003;**12**:40–6. <http://dx.doi.org/10.1136/qhc.12.1.40>
93. Kluger AN, DeNisi A. Feedback interventions: toward the understanding of a double-edged sword. *Curr Direct Psychol Sci* 1998;**7**:67–72. <http://dx.doi.org/10.1111/1467-8721.ep10772989>
94. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implement Sci* 2008;**3**:1–12. <http://dx.doi.org/10.1186/1748-5908-3-36>
95. Flick U. *An Introduction to Qualitative Research*. Berlin: Sage; 2014.
96. Dixon-Woods M, Leslie M, Bion J, Tarrant C. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q* 2012;**90**:548–91. <http://dx.doi.org/10.1111/j.1468-0009.2012.00674.x>
97. Bradley E, Holmboe E, Mattera J, Roumanis S, Radford M, Krumholz H. Data feedback efforts in quality improvement: lessons learned from US hospitals. *Qual Saf Health Care* 2004;**13**:26–31. <http://dx.doi.org/10.1136/qhc.13.1.26>
98. De Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. *Int J Qual Health Care* 2009;**21**:119–29. <http://dx.doi.org/10.1093/intqhc/mzn059>
99. Smith A, Mahajan R. National critical incident reporting: improving patient safety. *Br J Anaesth* 2009;**103**:623–5. <http://dx.doi.org/10.1093/bja/aep273>
100. Eccles MP, Hrisos S, Francis J, Kaner EF, Dickinson HO, Beyer F, *et al.* Do self-reported intentions predict clinicians' behaviour: a systematic review. *Implement Sci* 2006;**1**:28. <http://dx.doi.org/10.1186/1748-5908-1-28>
101. Benn J, Koutantji M, Wallace L, Spurgeon P, Rejman M, Healey A, *et al.* Feedback from incident reporting: information and action to improve patient safety. *Qual Saf Health Care* 2009;**18**:11–21. <http://dx.doi.org/10.1136/qshc.2007.024166>
102. Wallace L. Feedback from reporting patient safety incidents—are NHS trusts learning lessons? *J Health Serv Res Policy* 2010;**15**(Suppl. 1):75–8. <http://dx.doi.org/10.1258/jhsrp.2009.09s113>
103. Wallace LM, Spurgeon P, Benn J, Koutantji M, Vincent C. Improving patient safety incident reporting systems by focusing upon feedback—lessons from English and Welsh trusts. *Health Serv Manage Res* 2009;**22**:129–35. <http://dx.doi.org/10.1258/hsmr.2008.008019>
104. Ilgen DR, Fisher CD, Taylor MS. Consequences of individual feedback on behavior in organizations. *J Appl Psychol* 1979;**64**:349. <http://dx.doi.org/10.1037/0021-9010.64.4.349>
105. Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. *Br J Anaesth* 2012;**109**:80–91. <http://dx.doi.org/10.1093/bja/aes173>

106. Sapyta J, Riemer M, Bickman L. Feedback to clinicians: theory, research, and practice. *J Clin Psychol* 2005;**61**:145–53. <http://dx.doi.org/10.1002/jclp.20107>
107. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;**6**:CD000259. <http://dx.doi.org/10.1002/14651858.cd000259.pub3>
108. Ovretveit JC, Shekelle PG, Dy S, McDonald KM, Hempel S, Pronovost PJ, *et al*. How does context affect interventions to improve patient safety? *BMJ Qual Saf* 2011;**20**:604–10. <http://dx.doi.org/10.1136/bmjqs.2010.047035>
109. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Processes* 1991;**50**:179–211. [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
110. Festinger L. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press; 1957.
111. Skinner BF. 'Superstition' in the pigeon. *J Exp Psychol* 1948;**38**:168–72. <http://dx.doi.org/10.1037/h0055873>
112. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;**82**:581–629. <http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x>
113. Rogers EM. A prospective and retrospective look at the diffusion model. *J Health Commun* 2004;**9**(Suppl. 1):13–19. <http://dx.doi.org/10.1080/10810730490271449>
114. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. *Manag Sci* 1989;**35**:982–1003. <http://dx.doi.org/10.1287/mnsc.35.8.982>
115. Chuttur M. *Overview of the Technology Acceptance Model: Origins, Developments and Future Directions*. All Sprouts Content; 2009.
116. Holden RJ, Karsh B-T. The technology acceptance model: its past and its future in health care. *J Biomed Inform* 2010;**43**:159–72. <http://dx.doi.org/10.1016/j.jbi.2009.07.002>
117. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology* 2009;**43**:535–54. <http://dx.doi.org/10.1177/0038038509103208>
118. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, *et al*. Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res* 2007;**7**:148. <http://dx.doi.org/10.1186/1472-6963-7-148>
119. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, *et al*. Development of a theory of implementation and integration: normalization process theory. *Implement Sci* 2009;**4**:29. <http://dx.doi.org/10.1186/1748-5908-4-29>
120. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, *et al*. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med* 2010;**8**:63. <http://dx.doi.org/10.1186/1741-7015-8-63>
121. Gollop R, Whitby E, Buchanan D, Ketley D. Influencing sceptical staff to become supporters of service improvement: a qualitative study of doctors' and managers' views. *Qual Saf Health Care* 2004;**13**:108–14. <http://dx.doi.org/10.1136/qshc.2003.007450>
122. Gustafson DH, Sainfort F, Eichler M, Adams L, Bisognano M, Steudel H. Developing and testing a model to predict outcomes of organizational change. *Health Serv Res* 2003;**38**:751–76. <http://dx.doi.org/10.1111/1475-6773.00143>
123. Iles V, Sutherland K. *Organisational Change. A Review for Health Care Managers, Professionals and Researchers*. London: NCCSDO; 2001.

124. Pettigrew AM, Woodman RW, Cameron KS. Studying organizational change and development: challenges for future research. *Acad Manag J* 2001;**44**:697–713. <http://dx.doi.org/10.2307/3069411>
125. Scott T, Mannion R, Davies HT, Marshall MN. Implementing culture change in health care: theory and practice. *Int J Qual Health Care* 2003;**15**:111–18. <http://dx.doi.org/10.1093/intqhc/mzg021>
126. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care – an interactive sociotechnical analysis. *J Am Med Inform Assoc* 2007;**14**:542–9. <http://dx.doi.org/10.1197/jamia.M2384>
127. Ashford J, Eccles M, Bond S, Hall JA, Bond J. Improving health care through professional behaviour change: introducing a framework for identifying behaviour change strategies. *Br J Clin Govern* 1999;**4**:14–23. <http://dx.doi.org/10.1108/14664109910297146>
128. Steele CM. The psychology of self-affirmation: sustaining the integrity of the self. *Adv Exp Soc Psychol* 1988;**21**:261–302. [http://dx.doi.org/10.1016/S0065-2601\(08\)60229-4](http://dx.doi.org/10.1016/S0065-2601(08)60229-4)
129. Vygotsky LS. The Development of Scientific Concepts in Childhood. In Vygotsky L, Hanfmann E, Vakar G, editors. *Thought and Language: Studies in Communication*. Cambridge, MA: MIT Press; 1962. pp. 82–118.
130. Bandura A, Cervone D. Differential engagement of self-reactive influences in cognitive motivation. *Organ Behav Hum Decis Processes* 1986;**38**:92–113. [http://dx.doi.org/10.1016/0749-5978\(86\)90028-2](http://dx.doi.org/10.1016/0749-5978(86)90028-2)
131. Locke EA, Cartledge N, Koeppl J. Motivational effects of knowledge of results: a goal-setting phenomenon? *Psychol Bull* 1968;**70**:474–85. <http://dx.doi.org/10.1037/h0026737>
132. Smith A, Glavin R, Greaves J. Defining excellence in anaesthesia: the role of personal qualities and practice environment. *Br J Anaesth* 2011;**106**:38–43. <http://dx.doi.org/10.1093/bja/aeq308>
133. Buetow S. What motivates health professionals? Opportunities to gain greater insight from theory. *J Health Serv Res Policy* 2007;**12**:183–5. <http://dx.doi.org/10.1258/135581907781543111>
134. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991;**303**:398–402. <http://dx.doi.org/10.1136/bmj.303.6799.398>
135. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher* 2006;**28**:117–28. <http://dx.doi.org/10.1080/01421590600622665>
136. Hocking G, Weightman W, Smith C, Gibbs N, Sherrard K. Measuring the quality of anaesthesia from a patient's perspective: development, validation, and implementation of a short questionnaire. *Br J Anaesth* 2013;**111**:979–89. <http://dx.doi.org/10.1093/bja/aet284>
137. Moonesinghe SR, Tomlinson AA. Quality improvement and revalidation: two goals, same strategy? *Br J Anaesth* 2011;**106**:447–50. <http://dx.doi.org/10.1093/bja/aer052>
138. Francis R. *Report of the Mid-Staffordshire NHS Foundation Trust Public Inquiry*. London: The Stationery Office; 2010.
139. Barnett SF, Alagar RK, Grocott MPW, Giannaris S, Dick JR, Moonesinghe SR. Patient-satisfaction measures in anaesthesia: qualitative systematic review. *Anesthesiology* 2013;**119**:452–78. <http://dx.doi.org/10.1097/ALN.0b013e3182976014>
140. Castanelli D, Kitto S. Perceptions, attitudes, and beliefs of staff anaesthetists related to multi-source feedback used for their performance appraisal. *Br J Anaesth* 2011;**107**:372–7. <http://dx.doi.org/10.1093/bja/aer152>

141. Heidegger T, Saal D, Nübling M. Patient satisfaction with anaesthesia—Part 1: satisfaction as part of outcome—and what satisfies patients. *Anaesthesia* 2013;**68**:1165–72. <http://dx.doi.org/10.1111/anae.12347>
142. Powell A, Davies H, Thomson R. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;**12**:122–8. <http://dx.doi.org/10.1136/qhc.12.2.122>
143. Bird SM, Sir David C, Farewell VT, Harvey G, Tim H, Peter C. Performance indicators: good, bad, and ugly. *J R Stat Soc A Sta* 2005;**168**:1–27. <http://dx.doi.org/10.1111/j.1467-985X.2004.00333.x>
144. Jamtvedt G, Young J, Kristoffersen D, O'Brien M, Oxman A. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;**2**:CD000259. <http://dx.doi.org/10.1002/14651858.cd000259.pub2>
145. Wright J, Dugdale B, Hammond I, Jarman B, Neary M, Newton D, *et al*. Learning from death: a hospital mortality reduction programme. *JRSM* 2006;**99**:303–8. <http://dx.doi.org/10.1258/jrsm.99.6.303>
146. O'Reilly M, Talsma A, VanRiper S, Kheterpal S, Burney R. An anesthesia information system designed to provide physician-specific feedback improves timely administration of prophylactic antibiotics. *Anesthes Analg* 2006;**103**:908–12. <http://dx.doi.org/10.1213/01.ane.0000237272.77090.a2>
147. Kluger AN, Van Dijk D. Feedback, the various tasks of the doctor, and the feedforward alternative. *Med Educ* 2010;**44**:1166–74. <http://dx.doi.org/10.1111/j.1365-2923.2010.03849.x>
148. Alvero AM, Bucklin BR, Austin J. An objective review of the effectiveness and essential characteristics of performance feedback in organisational settings (1985–1998). *J Organ Behav Manag* 2001;**21**:3–29. http://dx.doi.org/10.1300/J075v21n01_02
149. Hysong S. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009;**47**:356–63. <http://dx.doi.org/10.1097/MLR.0b013e3181893f6b>
150. Mann K. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Acad Med* 2011;**86**:1120–7. <http://dx.doi.org/10.1097/ACM.0b013e318226abdd>
151. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf* 2012;**21**:13–20. <http://dx.doi.org/10.1136/bmjqs-2011-000010>
152. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Soc Sci Med* 2010;**70**:1618–25. <http://dx.doi.org/10.1016/j.socscimed.2010.01.039>
153. Cohen J, Cohen P, West S, Aiken L. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd edn. London: Lawrence Erlbaum Associates; 2003.
154. Kotter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators – a systematic review. *Implement Sci* 2012;**7**:21. <http://dx.doi.org/10.1186/1748-5908-7-21>
155. Lampe M. Appraisal and reassessment of the specialist in anaesthesia. *Best Pract Res: Clin Anaesthesiol* 2002;**16**:391–400. <http://dx.doi.org/10.1053/bean.2002.0219>
156. Evans RG, Edwards A, Evans S, Elwyn B, Elwyn G. Assessing the practising physician using patient surveys: a systematic review of instruments and feedback methods. *Fam Pract* 2007;**24**:117–27. <http://dx.doi.org/10.1093/fampra/cml072>

157. Veillard J, Champagne F, Klazinga N, Kazandjian V, Arah OA, Guisset A-L. A performance assessment framework for hospitals: the WHO regional office for Europe PATH project. *Int J Qual Health Care* 2005;**17**:487–96. <http://dx.doi.org/10.1093/intqhc/mzi072>
158. Briner M, Kessler O, Pfeiffer Y, Wehner T, Manser T. Assessing hospitals' clinical risk management: development of a monitoring instrument. *BMC Health Serv Res* 2010;**10**:337. <http://dx.doi.org/10.1186/1472-6963-10-337>
159. Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009;**47**:356–63. <http://dx.doi.org/10.1097/MLR.0b013e3181893f6b>
160. Johnston G, Crombie IK, Alder EM, Davies HTO, Millard A. Reviewing audit: barriers and facilitating factors for effective clinical audit. *Qual Health Care* 2000;**9**:23–36. <http://dx.doi.org/10.1136/qhc.9.1.23>
161. McCroskey JC. The use and abuse of factor analysis in communication research. *Hum Commun Res* 2006;**5**:375. <http://dx.doi.org/10.1111/j.1468-2958.1979.tb00651.x>
162. Huppelschoten AG, Aarts JW, van Empel IW, Cohlen BJ, Kremer JA, Nelen WL. Feedback to professionals on patient-centered fertility care is insufficient for improvement: a mixed-method study. *Fertility Sterility* 2013;**99**:1419–27. <http://dx.doi.org/10.1016/j.fertnstert.2012.12.024>
163. Birkmeyer NJ, Finks JF, Greenberg CK, McVeigh A, English WJ, Carlin A, et al. Safety culture and complications after bariatric surgery. *Ann Surg* 2013;**257**:260–5. <http://dx.doi.org/10.1097/SLA.0b013e31826c0085>
164. Davies H, Nutley S, Mannion R. Organisational culture and quality of health care. *Qual Health Care* 2000;**9**:111. <http://dx.doi.org/10.1136/qhc.9.2.111>
165. Ohrn A. Patient safety dialogue: evaluation of an intervention aimed at achieving an improved patient safety culture. *J Patient Saf* 2011;**7**:185–92. <http://dx.doi.org/10.1097/PTS.0b013e318230e702>
166. Burnett S, Benn J, Pinto A, Parand A, Iskander S, Vincent C. Organisational readiness: exploring the preconditions for success in organisation-wide patient safety improvement programmes. *Qual Saf Health Care* 2010;**19**:313–17. <http://dx.doi.org/10.1136/qshc.2008.030759>
167. Scholefield H. Embedding quality improvement and patient safety at Liverpool Women's NHS Foundation Trust. *Best Pract Res Clin Obstetr Gynaecol* 2007;**21**:593–607. <http://dx.doi.org/10.1016/j.bpobgyn.2007.02.005>
168. Speroff T, Nwosu S, Greevy R, Weinger MB, Talbot TR, Wall RJ, et al. Organisational culture: variation across hospitals and connection to patient safety climate. *Qual Safety Health Care* 2010;**19**:592–6. <http://dx.doi.org/10.1136/qshc.2009.039511>
169. Singer S. Identifying organizational cultures that promote patient safety. *Health Care Manag Rev* 2009;**34**:300–11. <http://dx.doi.org/10.1097/HMR.0b013e3181afc10c>
170. Firth Cozens J. Cultures for improving patient safety through learning: the role of teamwork. *Qual Health Care* 2001;**10**(Suppl. 2):ii26–31. <http://dx.doi.org/10.1136/qhc.0100026>
171. Frankel AS, Leonard MW, Denham CR. Fair and just culture, team behavior, and leadership engagement: the tools to achieve high reliability. *Health Serv Res* 2006;**41**:1690–709. <http://dx.doi.org/10.1111/j.1475-6773.2006.00572.x>
172. Garon M. Speaking up, being heard: registered nurses' perceptions of workplace communication. *J Nurs Manag* 2012;**20**:361–71. <http://dx.doi.org/10.1111/j.1365-2834.2011.01296.x>
173. Marshall T, Mohammed MA, Rouse A. A randomized controlled trial of league tables and control charts as aids to health service decision-making. *Int J Qual Health Care* 2004;**16**:309–15. <http://dx.doi.org/10.1093/intqhc/mzh054>

174. Benn J, Burnett S, Parand A, Pinto A, Iskander S, Vincent C. Perceptions of the impact of a large-scale collaborative improvement programme: experience in the UK Safer Patients Initiative. *J Eval Clin Pract* 2009;**15**:524–40. <http://dx.doi.org/10.1111/j.1365-2753.2009.01145.x>
175. Baker DF, Buckley MR. A historical perspective of the impact of feedback on behaviour. *J Manage Hist* 1996;**2**:21–33.
176. Foy R, Eccles M, Jamtvedt G, Young J, Grimshaw J, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Serv Res* 2005;**5**:50. <http://dx.doi.org/10.1186/1472-6963-5-50>
177. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;**321**:649–6.
178. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655.
179. Ovreteit J, Gustafson D. Evaluation of quality improvement programmes. *Qual Saf Health Care* 2002;**11**:270–5.
180. Schouten L, Grol R, Hulscher M. Factors influencing success in quality improvement collaboratives: development and psychometric testing of an instrument. *Implement Sci* 2010;**5**:84.
181. Benn J, Burnett S, Parand A, Pinto A, Iskander S, Vincent C. Studying large-scale programmes to improve patient safety in whole care systems: challenges for research. *Soc Sci Med* 2009;**69**:1767–76.
182. Speroff T. Study designs for PDSA Quality Improvement Research. *Qual Manage Health Care* 2004;**13**:17.
183. Eccles M, Grimshaw J, Campbell R, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 2003;**12**:47–52.
184. Shadish W, Cook T, Campbell D. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton, Mifflin and Company; 2002.
185. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002;**27**:299–309.
186. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005;**34**:215–20.
187. Stott DL. Whole team training. *J Perioper Pract* 2013;**23**:95.
188. Augusto V. Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Comput Ind Eng* 2010;**58**:231–8. <http://dx.doi.org/10.1016/j.cie.2009.04.019>
189. Cardoen B, Demeulemeester E, Beliën J. Operating room planning and scheduling: a literature review. *Eur J Operational Res* 2010;**201**:921–32. <http://dx.doi.org/10.1016/j.ejor.2009.04.011>
190. Overdyk FJ, Harvey SC, Fishman RL, Shippey F. Successful strategies for improving operating room efficiency at academic institutions. *Anesth Analg* 1998;**86**:896–906.
191. de Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. *Int J Qual Health Care* 2009;**21**:119–29. <http://dx.doi.org/10.1093/intqhc/mzn059>
192. Chaillet N, Dumont A. Evidence-based strategies for reducing cesarean section rates: a meta-analysis. *Birth* 2007;**34**:53–64. <http://dx.doi.org/10.1111/j.1523-536X.2006.00146.x>

193. Bienstock JL, Katz NT, Cox SM, Hueppchen N, Erickson S, Puscheck EE. To the point: medical education reviews—providing feedback. *Am J Obstetr Gynecol* 2007;**196**:508–13. <http://dx.doi.org/10.1016/j.ajog.2006.08.021>
194. Fakhri MG, Jones K, Rey JE, Berriel-Cass D, Kalinicheva T, Szpina S, *et al.* Sustained improvements in peripheral venous catheter care in non-intensive care units: a quasi-experimental controlled study of education and feedback. *Infect Control Hosp Epidemiol* 2012;**33**:449–55. <http://dx.doi.org/10.1086/665322>
195. Eslami S, Abu-Hanna A, Keizer N, Bosman R, Spronk P, Jonge E, *et al.* Implementing glucose control in intensive care: a multicenter trial using statistical process control. *Intensive Care Med* 2010;**36**:1556–65. <http://dx.doi.org/10.1007/s00134-010-1924-3>
196. Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2009;**44**:101–8. <http://dx.doi.org/10.1111/j.1365-2923.2009.03546.x>
197. Samarth CN GP. Process efficiency. Redesigning social networks to improve surgery patient flow. *J Healthc Inform Manag* 2009;**23**:20–6.
198. Hall R, Belson D, Murali P, Dessouky M. Modeling Patient Flows through the Healthcare System. In Hall R, editor. *Patient Flow: Reducing Delay in Healthcare Delivery*. Springer; 2006. http://dx.doi.org/10.1007/978-0-387-33636-7_1
199. Neal A, Griffin MA, Hart PM. The impact of organizational climate on safety climate and individual behavior. *Safety Science* 2000;**34**:99–109. [http://dx.doi.org/10.1016/S0925-7535\(00\)00008-4](http://dx.doi.org/10.1016/S0925-7535(00)00008-4)
200. Karvonen S. Patient flow analysis: planning a new surgery unit. *Br J Healthc Manag* 2012;**18**:96–102. <http://dx.doi.org/10.12968/bjhc.2012.18.2.96>
201. Carthey C. *Implementing Human Factors in Healthcare*. Clinical Human Factors Group; 2010.
202. Wallace J. The relationship between organisational culture, organisational climate and managerial values. *Int J Public Sector Manag* 1999;**12**:548–64. <http://dx.doi.org/10.1108/09513559910305339>
203. Rotter T, Kugler J, Koch R, Gothe H, Twork S, van Oostrum JM, *et al.* A systematic review and meta-analysis of the effects of clinical pathways on length of stay, hospital costs and patient outcomes. *BMC Health Serv Res* 2008;**8**:265. <http://dx.doi.org/10.1186/1472-6963-8-265>
204. Flottorp S, Jamtvedt G, Gibis B, McKee M. *Using Audit and Feedback to Health Professionals to Improve the Quality and Safety of Health Care. Policy Summary*. Copenhagen: WHO Regional Office for Europe, on behalf of the European Observatory on Health Systems and Policies; 2010.
205. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, *et al.* No more ‘business as usual’ with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014;**9**:14 <http://dx.doi.org/10.1186/1748-5908-9-14>

Appendix 1 Qualitative interview schedule: first time point

Evaluative stakeholder interview schedule

(A) Preamble to interviews

Provide standard information form to read. Prompt for clarity/questions. Obtain consent. Provide brief verbal overview of the project and aims. Start recording.

- Please could you introduce yourself and state your role/specialty for the recording?

(B) General views upon feedback

- In your view, what are the most important aspects of quality of care relevant to anaesthetics practice?
 - Prompt for perspectives of anaesthetists, patients, other HC [health-care] professionals, managers.
- Do you think anaesthetists generally get adequate feedback upon these aspects of quality of care?

(C) Evaluation of the current initiative

- What are your general thoughts about this initiative and the feedback reports that you receive? [Introduce feedback report template]
- What do you think of the quality indicators that are currently reported?
 - Have we missed anything important?
 - Respond to the following (prompt for explanation of ratings)

For each measure please rate (on a scale of 1–5)	PONV	Pain	Temperature
1: Importance to overall quality of anaesthetic care? (Validity)			
2: Confidence in the accuracy of the measure? (Reliability)			
3: Degree to which you can influence this measure? (Controllability)			

- (Review the matrix above and check that our interpretation is correct)
- What was your initial reaction to seeing your data?
 - Were the results as you expected?
 - What are the benefits? What do the reports tell you that you did not know before?
- How do you use the information contained within the reports?
 - Are there any examples of changes you have made to your practice?
 - Does the data help you identify and address underlying reasons for variations?
 - Do you think it's possible for you to influence the data through changing your practice?
 - If the data suggested there was an opportunity for improvement, would you change your practice?
 - How do you think anaesthetists should use the data?
 - What do you think might prevent anaesthetists from making improvements based upon the data?
 - How do you think the department should use the data?

- What do you think about the current report format?
 - Frequency, length, graphical/text content, technical complexity.
- Would you rather see your data compared with others, your data displayed over time, or both?
 - How do you feel about being compared with your colleagues? Prompt for any case-mix issues. Is competition important?

(D) Future development

- Are there any measures, features or functionality that you would like to see included in future versions of the reports?
 - Would you be interested in a longer report with a more detailed breakdown/analysis of your data?
- What further support could be provided for anaesthetists to use this data to improve care?
- Can you see a role for initiatives of this type in revalidation?

(E) Broader context

- Do you see any barriers to engagement with and utilisation of this initiative?
 - Any concerns around use of the data?
- Is there anything about the organisation or context in which you work that might make a system like this one more or less successful?
- Do you think there is an atmosphere of transparency here amongst the clinical group concerning quality issues?
 - Prompt for comfort with disclosing and discussing personal performance data with peers. Are such discussions constructive/punitive?
 - What other factors influence whether you are comfortable for your data to be collected and used in this way?
- What support from the broader organisation/department/specialty would you need to use this data effectively for continuous improvement?

Indicator ratings template for use during interview:

For each measure please rate:		PONV	Pain	Temperature
(low)	1—2—3—4—5 (high)			
1	Importance to overall quality of anaesthetic care?			
2	Confidence in the accuracy of the measure?			
3	Degree to which you can influence this measure?			

Appendix 2 Qualitative interview schedule: second time point

Introduction to interview.

Give out participant information sheet and complete consent form.

- Have you previously been interviewed as part of this project?
- If so . . .
- As you may know the feedback initiative has changed and developed significantly and is now a trust wide initiative. We have made many improvements to the reports based on previous interviews and discussion but we are still looking to develop them even further. In order to do this we would like to discuss a number of key areas with you today.
- Do you receive the feedback reports?
- How often?
- Do the reports contain all of your cases?
- What are your general impressions/reflections on this initiative and the fact that you are receiving these feedback reports?
- Have your views about the feedback reports changed or developed in any way over time?

Data to information:

- How do you make sense of the data? How do you give it meaning? How do you translate the data into something that is meaningful to you?
- What information do you think the individual clinician needs to get from these reports in order to learn and make improvements?
- What do you do with the information that we provide to you?
- Do you review it/carry it with you/discuss it with colleagues?
- You now receive data on the quality of recovery score. What do you think about receiving this data?
- How do you interpret this data?
- What is your understanding of what it shows/adds?

Information to action:

- How do you use the information that is provided to you?
- Have you changed your practice based on this feedback?
- Do you intend to change your practice based on the feedback reports that you receive?
- Are the feedback reports sufficient?
- If you wanted to change your practice would this initiative be enough to support you in doing so? If not, what is missing?
- How could you use this report for appraisal/revalidation purposes? Would it be useful? If not, why not? What is missing to help you do this?
- When revalidation is introduced would you be comfortable using this report as evidence for your fitness to practice? If not, why not?
- What do you think the consequences are of using this data to make improvements? Do you think they are good or bad?
- What do you think are the barriers and facilitators to using this data effectively to make improvements?
- What supports you in using this data?
- What makes it easier/more difficult to use this data to make improvements?
- Do you feel that you have the resources to effectively use the data that we provide you with?

- Do you feel any sense of responsibility to act upon the data that we provide you with?
- What would encourage you to act upon the data that we provide you with in the reports?
- How does the fact that the reports are anonymous help or hinder you in your use of them to make improvements?
- Have your perceptions about anonymity changed over time?

Conceptualisations of the report:

- What does the report show about you as a professional?
- What do you think it says about your department that you are involved in this type of initiative in the first place?
- Is this initiative compatible with current guidelines and best practice for perioperative units?
- How do the reports make you think or feel differently about your professional practice?

Context:

- How does your local environment affect the way in which you use the feedback reports?
- Do your peers support you in using the feedback reports? If so how?
- What do your peers think about you using this data to make improvements?
- What is the general feeling across the department about the use of this data to make improvements?

Further developments:

- How do you visualise this initiative going forward?
- How can we make further improvements to support you in your use of the reports?
- What could we change about the feedback that would encourage you to change your practice based upon it?

Appendix 3 Evaluative survey: feedback on quality of anaesthetic care (St Mary's Hospital)

Instructions to respondents

This questionnaire is concerned with your experience of the data that is fed back to you on the quality of care delivered by **anaesthetic services at St Mary's Hospital** and how that information is used. Where we refer to "feedback" we mean quantitative data from measures and indicators, rather than anecdotal reports or conversations. We ask that you give your name to enable follow-up measurement. Responses will be analysed and reported anonymously.

A) PERSONAL DETAILS

Name		Grade	Trainee		Year of qualification
Hospital & Dept	SMH Anaesthesia		Consultant		
Specialty					Country of qualification

B) FOCUS OF CURRENT QUALITY FEEDBACK

Dimensions of quality			
Do you receive regular quantitative feedback on each of the following dimensions of quality? Regular feedback = at least monthly. Please circle Yes or No.			
Yes	No	Clinical effectiveness of care (e.g. clinical outcomes, pain control, etc.)	B01
Yes	No	Compliance with best practice guidelines (e.g. hand hygiene, perioperative temperature control, etc.)	B02
Yes	No	Productivity and efficiency (e.g. waste, patient flow, throughput, delays)	B03
Yes	No	Financial performance (e.g. cost efficiency of care delivered)	B04
Yes	No	Patient safety (e.g. freedom from adverse events/critical incidents and hospital acquired infection)	B05
Yes	No	Patient experience (e.g. patient satisfaction, waiting times, clarity of info, quality of interaction, respect & dignity)	B06

Level of feedback			
Considering clinical effectiveness, patient safety and patient experience as a whole, please circle True or False for each of the following:			
True	False	I receive monthly or more regular feedback concerning the care delivered by the Trust or Hospital	B07
True	False	I receive monthly or more regular feedback concerning the care delivered by my department	B08
True	False	I receive monthly or more regular feedback concerning the care I delivered personally	B09

C) EFFECTIVENESS OF CURRENT QUALITY FEEDBACK

Please consider the adequacy of feedback you receive from quality indicators against each of the following dimensions and circle the appropriate number on the scale provided:

Your perceptions of current quality of care indicators in your area:		Completely inadequate							Excellent
C01	Comprehensiveness: The degree to which the data you receive is comprehensive and covers all important dimensions of care quality.	1	2	3	4	5	6	7	8
C02	Relevance: The degree to which care quality indicators are unambiguous and specific to our service area and the care we routinely deliver to patients.	1	2	3	4	5	6	7	8

C03	Reliability: The degree to which indicators are objective and reliable indicators of current standards of care, promoting confidence in the accuracy of the data over time.	1 2 3 4 5 6 7 8
C04	Improvability: The degree to which indicators measure aspects of care that you and your unit can have a direct impact upon through changing behaviour, the care process or local systems.	1 2 3 4 5 6 7 8

Your perceptions of the current feedback you receive:		Completely inadequate	Excellent
C05	Level of analysis: The degree to which the data you receive is broken down to a level that is directly relevant to you (e.g. for your team, your ward, your operating theatre, your patients).	1 2 3 4 5 6 7 8	
C06	Timeliness: The degree to which the frequency of feedback you receive helps you to monitor how care quality varies over time.	1 2 3 4 5 6 7 8	
C07	Means of communication: The degree to which the channel and method for dissemination (e.g. meetings, email, reports, posters) are useful and engaging.	1 2 3 4 5 6 7 8	
C08	Data presentation: The degree to which the format in which data is presented (e.g. tables, graphs, scorecards) is clear and easy to use, with the right amount of data presented.	1 2 3 4 5 6 7 8	
C09	Data credibility: The degree to which the data is viewed as credible and from a trustworthy, unbiased source.	1 2 3 4 5 6 7 8	
Your perceptions of how quality of care data is used for improvement:		Completely inadequate	Excellent
C10	Identifying problem areas and good practice: The degree to which data feedback helps us to rapidly detect problems and identify instances of excellent care.	1 2 3 4 5 6 7 8	
C11	Benchmarking: The degree to which the data feedback allows us to compare ourselves against similar units and/or national guidelines in a meaningful way.	1 2 3 4 5 6 7 8	
C12	Prioritising action: The degree to which data feedback supports prioritising where we put our efforts to improve care and which specific processes to focus upon.	1 2 3 4 5 6 7 8	
C13	Setting measurable objectives: The degree to which data feedback supports setting quantifiable targets for improvement.	1 2 3 4 5 6 7 8	
C14	Monitoring progress: The degree to which data feedback supports evaluation of our progress towards targets over time and whether any gains are sustained.	1 2 3 4 5 6 7 8	
C15	Overall usefulness for improvement: The degree to which current data feedback is useful in monitoring variations and improving care.	1 2 3 4 5 6 7 8	

D) THE ENVIRONMENT IN WHICH YOU WORK

Please consider the extent to which you would agree with the following statements and circle the appropriate number on the scale provided.

		Strongly Disagree	Strongly Agree
D01	In this department we are proactive in striving to continuously improve standards of care	1 2 3 4 5 6 7 8	
D02	In this department we routinely review data on quality of care outcomes	1 2 3 4 5 6 7 8	
D03	In this department it is clear as to what are acceptable standards of care	1 2 3 4 5 6 7 8	
D04	In this department we monitor compliance with best practice guidelines for clinical care	1 2 3 4 5 6 7 8	

D05	In this department we respond to variation constructively, to improve care, rather than blaming and punishing individuals	1	2	3	4	5	6	7	8
D06	In this department we openly discuss minor failures in care to learn lessons	1	2	3	4	5	6	7	8
D07	In this department we review critical incidents and serious failures to improve systems	1	2	3	4	5	6	7	8
D08	In this department we can demonstrate to senior levels of the organisation the quality of care we are delivering	1	2	3	4	5	6	7	8
D09	In this department we are supported by the organisation in our efforts to collect and use data	1	2	3	4	5	6	7	8
D10	I have adequate knowledge and training on the statistics required to interpret quality of care data	1	2	3	4	5	6	7	8
D11	I have adequate knowledge and training in quality improvement methods	1	2	3	4	5	6	7	8
D12	In this department variations in the quality of care delivered to patients often go undetected	1	2	3	4	5	6	7	8
D13	In this department it is clear who is responsible for taking action to make changes and improve care processes	1	2	3	4	5	6	7	8
D14	In this department we use personal data on the quality of care that individuals deliver in a constructive way	1	2	3	4	5	6	7	8
D15	I am comfortable for the quality of care received by my patients to be monitored and fed back to me	1	2	3	4	5	6	7	8
D16	In this department we are effective in taking action based upon audit and data feedback	1	2	3	4	5	6	7	8

Appendix 4 Raw survey item scores by study epoch and site

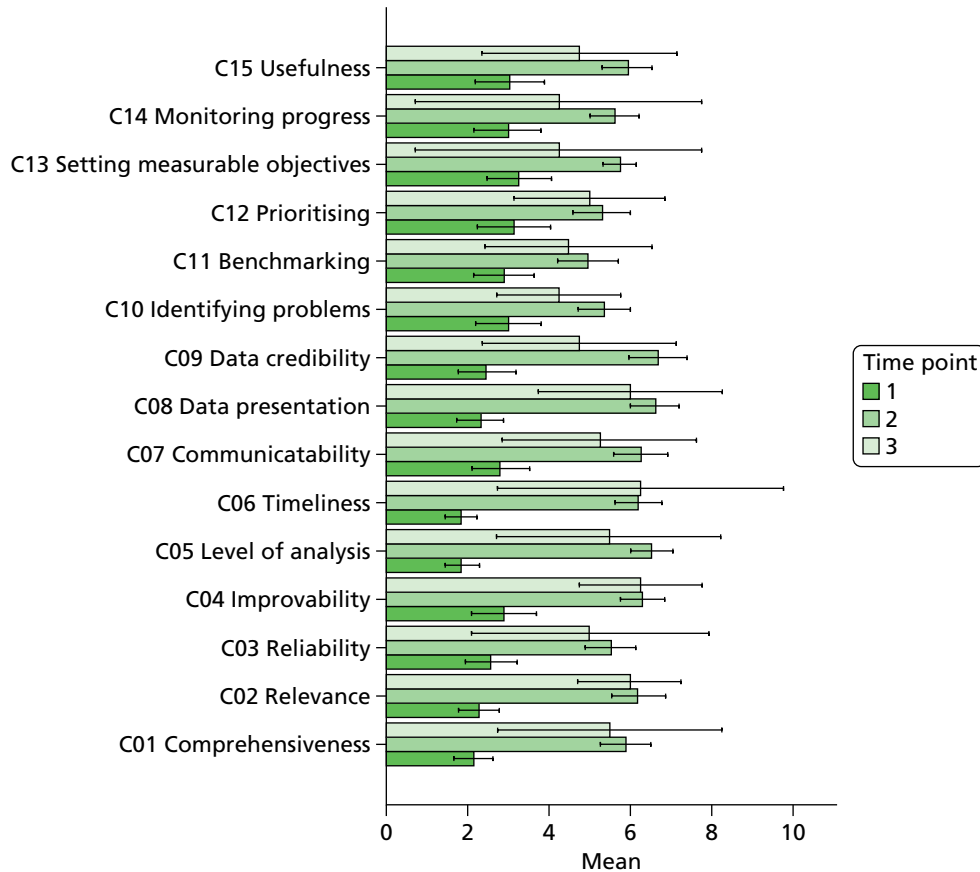


FIGURE 21 Primary site items C1–15 (overall effectiveness). Error bars: 95% confidence intervals.

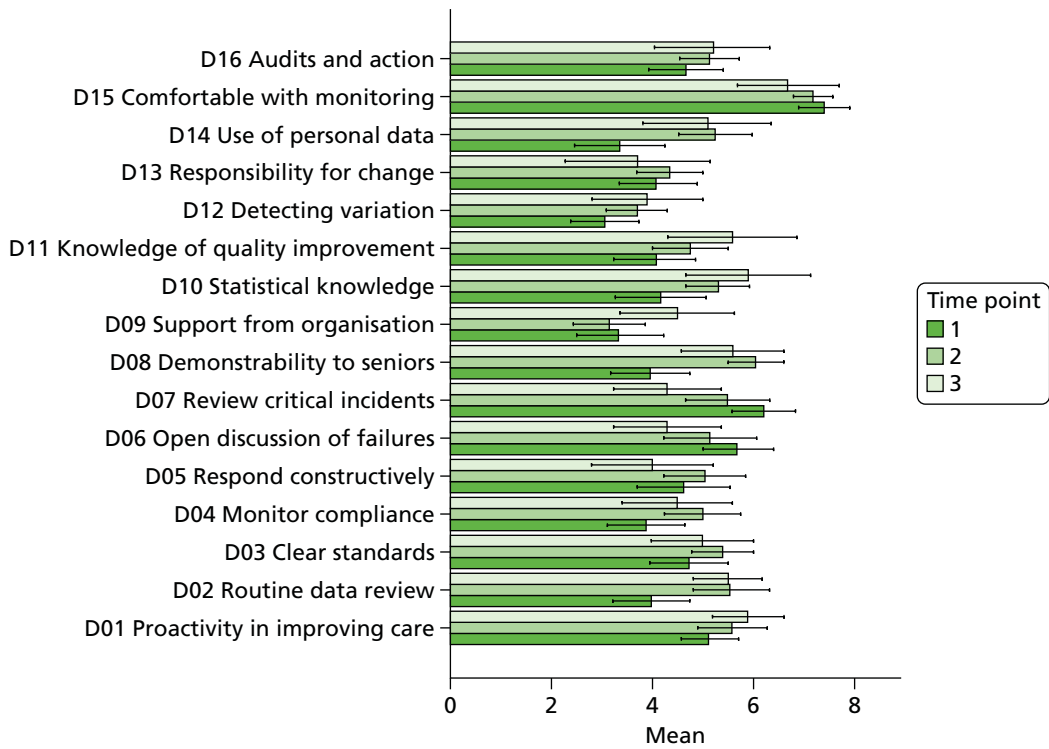


FIGURE 22 Primary site items D1–12 (workplace climate). Error bars: 95% confidence intervals.

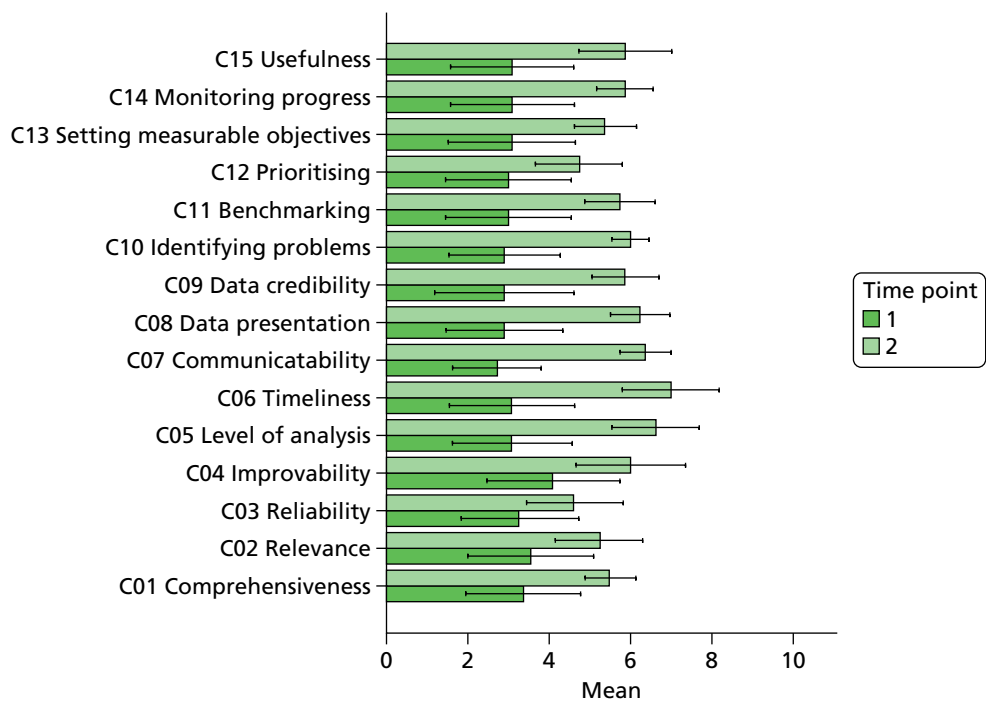


FIGURE 23 Additional site 1 items C1–15 (overall effectiveness). Error bars: 95% confidence intervals.

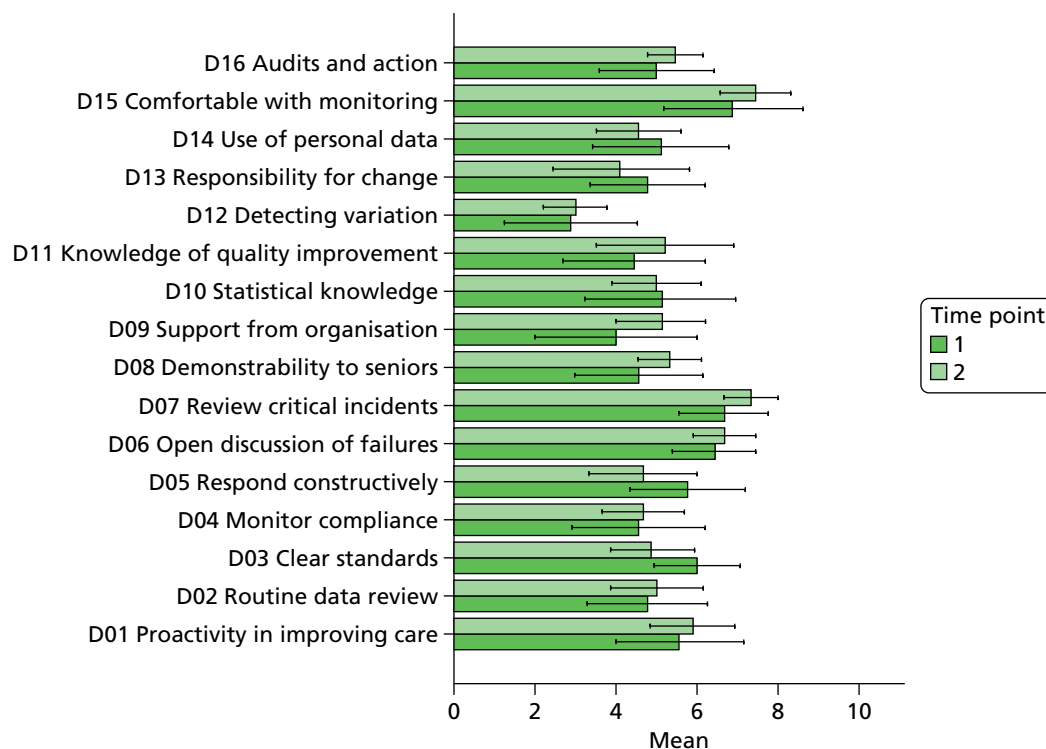


FIGURE 24 Additional site 1 items D1–12 (workplace climate). Error bars: 95% confidence intervals.

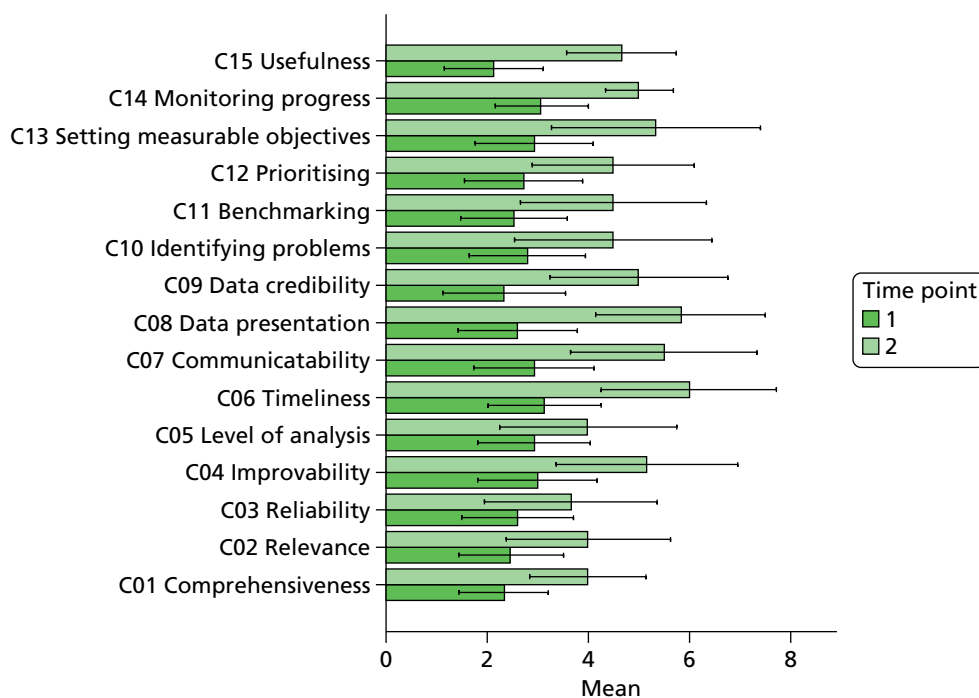


FIGURE 25 Additional site 2 items C1–15 (overall effectiveness). Error bars: 95% confidence intervals.

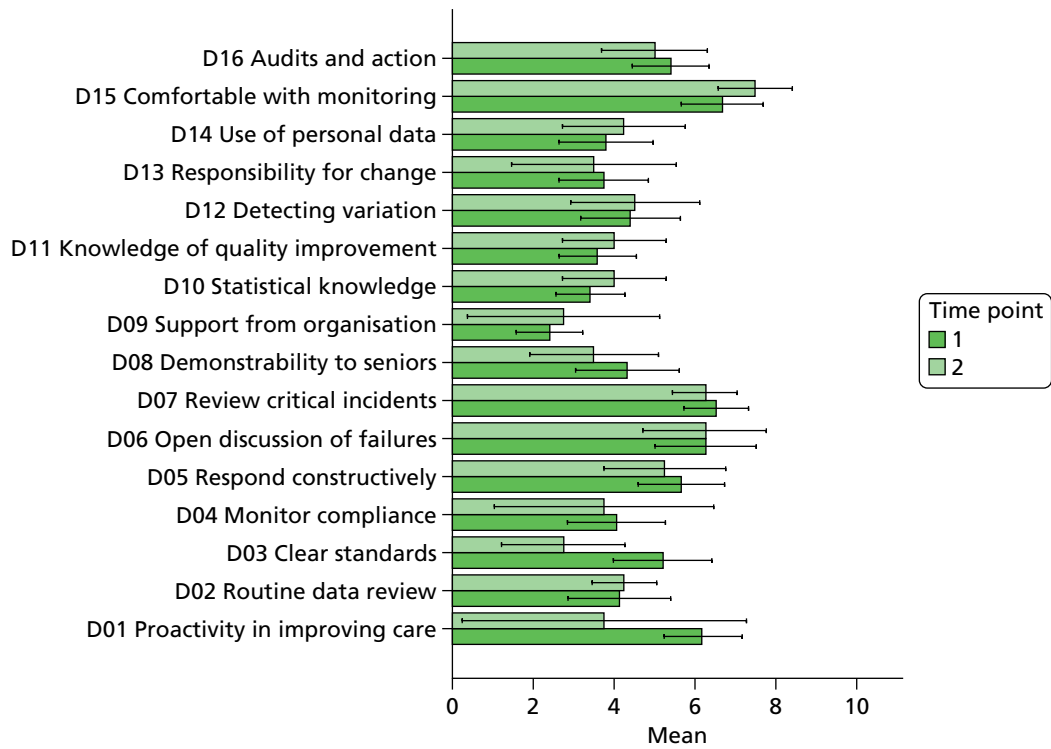


FIGURE 26 Additional site 2 items D1–12 (workplace climate). Error bars: 95% confidence intervals.

Appendix 5 Recovery unit data collection form

Note that the 0–10 pain scale score 'Pain on waking' included below was subsequently recoded into a 5-point pain scale for analytic purposes and to provide clinician feedback.

NETSCC 11/1015/21: Continuous monitoring and feedback in anaesthesia

Recovery Unit Data Collection Form

Please complete a form for every patient not transferred from Day Surgery

Hospital Number (Please do not include patient's name)	
Date of Birth and Sex	DoB: ____/____/____ Male Female
Consultant ANAESTHETIST surname or code (Please do not include the Surgeon – Refer to Rota only if illegible on patient notes)	
Procedure (Need not include laterality and reason for surgery)	
Date and Time of Admission to Recovery	Date: ____/____/2011 Time: ____:____
Temperature at Admission to Recovery (First set of readings immediately upon arrival. Please continue to measure if UTA)	Temp: ____ . ____ °C
Pain on waking (Worst pain within entire 40 minutes of waking from GA, or entire stay if RB or LA)	<p style="text-align: center;">0 1 2 3 4 5 6 7 8 9 10</p> <p style="text-align: center;">[No Pain] [Severe Pain]</p>
Nausea, Retching, Vomiting during stay (Entire stay in recovery until discharged)	Nil <input type="checkbox"/> Nausea <input type="checkbox"/> Vomiting Retching <input type="checkbox"/>
Time when Patient ready for Discharge (When score >= 9, even if no bed assigned)	Time: ____:____
Time bed was assigned (by site manager) (Only if no bed assigned when patient is ready for discharge. Bed need not be ready)	Time: ____:____ (only if applicable)
Time of first call to ward (Or time ward called recovery, whichever is sooner)	Time: ____:____
Time ward nurse arrived (Only if patient not going home or taken to ward by recovery nurse)	Time: ____:____ (only if applicable)
Time patient discharged to ward (Time patient is picked up by ward staff)	Time: ____:____
Discharge Ward (Please check if HDU bed or not)	

Comments (if applicable) <ul style="list-style-type: none"> - Reason for delays (bed, nurse, porter) - Overnight stays - Taken to ward by recovery staff - Change in ward - General comments 	
---	--

Post-anaesthesia patient experience survey

To Patient:

We would like to know how well you feel after your operation and the anaesthesia provided to you in order to improve the quality of our service.

Please read the following statements on the left of the table and tick the box corresponding to the appropriate response for each question, thinking specifically about the anaesthetic care you have received.

To Nurse:

Please complete this form on behalf of the patient, if the patient is unable, forgets or declines to fill it in, when the patient becomes **ready for discharge / meets the discharge score criteria**.

It is important that this form is completed in full. If you are unsure about a question, please use your best judgement to answer it. If the patient experienced pain after 40 minutes of waking or generally had pain, this is the place to make a note of it.

Categorisation:

Please tick the appropriate box below for categorisation purposes only.

Patient completed (direct patient input, even if completed by nurse)

Nurse completed (no direct patient input)

We are very grateful for your help – Thank You.

Have you:	Not at all	Some of the time	Most of the time
Had a general feeling of well-being ¹			
Had support from others			
Been able to understand instructions and not felt confused ²			
Felt able to look after your personal and hygiene needs unaided			
Felt able to comfortably toilet yourself ³			

NETSCC 11/1015/21: Continuous monitoring and feedback in anaesthesia

Been able to breathe easily			
Been free from nausea, dry retching or vomiting			
Been free from any severe pain, or constant moderate pain			

¹ How do you feel psychologically? Are you emotionally comfortable?

² Do you feel light-headed, disoriented or unable to think clearly? Are you aware you are in recovery? [Question unrelated to dementia]

³ If you have not needed to use the toilet/bottle, would you have sufficient muscle control to do so and without discomfort?

Appendix 6 Extended project summary

Background

Research and theory suggests that effective monitoring of the quality of service delivery is central to the capacity of an organisation, unit or individual to maintain and improve standards of care. Effective monitoring and feedback is essential if clinical teams and individuals are to understand variations in care, detect and respond to opportunities to improve standards, and evaluate the impact of changes on services.

Anaesthetists as a professional group have a high degree of patient contact in the perioperative pathway yet receive little routine feedback on patient experience or outcomes, such as pain and postoperative nausea, specific to quality of anaesthetic care. Feedback on postoperative nausea, pain control or perioperative normothermia measures often occurs irregularly in acute care organisations through clinical audit projects, but these information streams are discontinuous and are not geared towards continuous monitoring and improvement. Recent reviews have highlighted the fact that, from the anaesthetist's perspective, current perioperative quality indicators lack sensitivity and specificity. Furthermore, there exists limited evidence concerning the reliability and validity of measures that can be used to monitor patient satisfaction with the quality of anaesthetic care. A clear need exists for the development of routine monitoring and feedback of quality of anaesthetic care, from the postoperative period, to support improvement in anaesthetic practice.

Studies show that providing feedback to clinicians can be an effective improvement intervention and results in generally small to moderate positive effects on professional practice. Studies suggest that process-of-care measures may be more sensitive to data feedback initiatives than outcomes. Initiatives that use feedback are more effective than those that do not, and feedback paired with an educational strategy or implementation plan is more effective than using simple passive feedback alone. Further important characteristics of effective feedback include proximity to the time of decision-making, being timely and continuous, credibility of the source, trust in the data, presence of a sense of ownership of the data by recipients, being non-punitive or non-judgemental in tone, high outcome expectancy of recipients, motivation of recipients and being supported by management and hospital resources.

Objective

The main aim of this research was to conduct a comprehensive, mixed-methods, quasi-experimental evaluation of the impact of a departmental continuous quality monitoring and feedback initiative for quality of anaesthetic care, within a London teaching hospital over a 3-year period. The project was designed to produce both formative and summative data to support the development and evaluation of the feedback initiative.

The intervention

The feedback initiative was developed and implemented as part of the CLAHRC North West London portfolio of quality improvement projects and was given the title IMPAQT. Based on industrial quality improvement models, the initiative was conceived as a continuous quality monitoring and feedback programme for anaesthetists. It comprised continuous measurement of anaesthetic quality indicators in the PACU of the primary study site coupled with continuous monthly feedback of personal-level case data to 44 consultant anaesthetists. Baseline data collection of anaesthetic quality indicators began in March 2010 at St Mary's Hospital main theatre suite. The intervention model was then implemented in two main phases: (1) implementation of basic, passive monthly feedback using a simple summary statistical report

(from October 2010), and (2) implementation of an enhanced feedback protocol (from July 2012 until the end of the project in November 2013).

Basic feedback consisted of the provision of monthly personal data summaries in tabled form for a limited number of summary quality metrics, compared with group-level averages without adjustment. Limited longitudinal and normative comparisons were included in graphical format. In the enhanced phase of the programme, basic data feedback was enhanced with broader professional engagement activities and rapid, responsive development of the feedback model in response to user feature requests, in order to increase the capability of the feedback to stimulate improvement in professional practice, learning from case experience and quality of anaesthetic care. In addition to the basic feedback content, the enhanced reports were more statistically sophisticated and included monthly detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination. During the enhanced feedback phase, engagement with the anaesthetist group was much more active, involving regular presentation of statistical results at meetings, consultative interviews by the research team for formative evaluation of the preferred features of the feedback for end-users and more focused engagement and facilitated peer interaction on specific specialty areas in which potential quality issues were identified (e.g. pain management after gynaecological surgery). During the enhanced feedback phase, the scope of data collection was increased to include multiple sites, which increased the profile of the project within the trust perioperative department, and the initiative was subsequently rolled out and maintained at multiple sites across the health-care trust.

Methods

Qualitative evaluation

The design was a longitudinal, qualitative work stream which ran parallel to the intervention work and took a realist evaluative perspective on the project. The realist position provides a framework for identifying not only what outcomes are produced by an intervention, but how they are produced and how the intervention interacts with varying local conditions to produce the outcomes. Complex, multifaceted interventions are subject to phased implementation and intensive iteration. We also expected levels of engagement and impact of the initiative to vary as a result of ongoing interaction with contextual and organisational preconditions. In total, we conducted interviews lasting between 30 and 60 minutes with 24 consultant anaesthetists, six surgical nursing leads and five perioperative service leads, in two main phases, over the course of the project.

Survey evaluation

In order to quantify the response to the development of the feedback initiative from the perspectives of end-users of the feedback, a longitudinal evaluative survey study was designed, with three time points corresponding to baseline, basic feedback and enhanced feedback conditions at the primary study site. Pre- and post-intervention survey measures were additionally taken at the secondary sites associated with the project, which later received the same enhanced feedback protocol. The survey was developed based on prior research and in consultation with a group of consultant anaesthetists, who later undertook comprehensive piloting in order to refine the measures. The items included scales designed to quantify the effectiveness of current quality indicators, data feedback and the usefulness of the feedback for improvement, along with attitudes to quality improvement within the local working environment. In total, 70 individual anaesthetists completed the survey at one or more time points across the three study sites.

Quasi-experimental evaluation

Evaluation of the impact of the anaesthetics quality monitoring and feedback initiative on anaesthetic quality indicators and perioperative outcomes utilised a single-group longitudinal design, with multiple study epochs. ITSA was used as the primary evaluative model, with interrupts representing multiple intervention time points corresponding to the onset of basic and enhanced feedback protocols. ITSA was chosen in order to model changes in time series level and slope associated with implementation and

escalation of the intervention, representing the stepwise impact of the intervention and effect on rate of change in a parameter over time, while controlling for prior temporal trends. The perioperative indicators modelled included patient temperature on arrival in recovery (two metrics), patient-reported QoR Scale score (two metrics), postoperative pain (two metrics), postoperative nausea (two metrics), SSI rate and 30-day postoperative mortality rate. The primary ITSA investigated the effects of the two-stage intervention model on perioperative indicators recorded for all surgical cases performed by the pilot cohort of anaesthetists at the primary study site. Secondary sensitivity analysis was performed to investigate the robustness of primary findings to variations in case inclusion criteria, control of covariates and intervention timeline. The study anaesthetist cohort comprised 50,235 cases, performed by 44 anaesthetists over the course of the study, with 22,670 cases performed at the primary hospital site.

Productivity analysis

A mixed-methods evaluation of the effects of the feedback intervention on productivity was undertaken. Interviews were conducted with the perioperative service manager, the lead nurse for the post-anaesthetic care unit and six surgical nursing leads from the primary site to identify and interpret themes related to productivity. WWT was defined as the interval between the receiving ward being contacted after the patient was deemed ready for discharge from the PACU unit and handover of the patient to the ward nurse. We used *t*-tests to identify significant differences in WWT pre and post feedback both for the aggregated hospital level data and for individual surgical wards.

Results

Qualitative evaluation

The results from the qualitative work provided a rich understanding of the causal mechanisms of effectiveness for monitoring performance and making improvements to practice based on quality indicators, along with a developmental perspective on acceptability and engagement over time. Clinicians clearly agreed with the rationale for the initiative, recognising the existence of a problem and the need for a solution. Clinicians emphasised that the right quality indicators needed to be selected with the right characteristics (i.e. they must be specific, relevant and meaningful) in order to promote the necessary level of trust in the data and demonstrate fitness for purpose. The interviewees explored the translation of information into action at two levels of the health-care system: the departmental level and the individual clinician level. Crucially, the mechanisms of effective data use were different at each level.

The issue of anonymity was important to end-users and appeared to demonstrate a process of maturity that was longitudinally dependent and tied to end-users growing confidence in the intent of the feedback system. At the individual level, with a desire to receive normative feedback, people wanted to identify and contact high performers in order to obtain support and ideas for behaviour change. In this sense anonymity could actually be viewed as a barrier to the effective translation of information into improvements.

Dealing with case-mix variations (and the intraprofessional issues it gave rise to) was identified as a critical success factor for initiatives of this type, too. Careful selection of the right indicators can reduce the need for case-mix adjustment. Arguably, all patients should arrive in recovery warm and have well-managed pain and nausea postoperatively. If there are differences in risk then anaesthetists should be compensating through their practice. Realistically, though, pain/nausea is more difficult to control for in certain procedures. This goes some way towards explaining why within-specialty group work was so welcomed by end-users in the later stages of the project. This enabled fair comparison alongside an opportunity to learn from peer experts in an open and constructive way. Our analysis suggests that a combination of normative comparison (i.e. genuine peer benchmarking) and individual trends over time may have the greatest effect. The need to transform hard data into usable information and the experience of health-care professionals in doing so can be viewed as a powerful message emerging from this study.

Interviewees clearly identified a role for this initiative in revalidation and participating in quality monitoring and acting on the results is an identified dimension of good medical practice. In fact, the conceptual connection of the initiative with revalidation and appraisal appeared to significantly increase levels of engagement throughout the evolution of the project.

Survey evaluation

The results of the survey study suggest that anaesthetists perceive a range of factors as important in determining the usefulness of feedback. Specifically, the local departmental context and its support of quality improvement is an important determinant of how instrumental feedback from monitoring quality indicators is likely to be. Furthermore, feedback that is tailored to be relevant to the personal professional practice of the individual clinician is an important predictor of usefulness. In terms of the feedback content and design characteristics that anaesthetists value most, the perceived credibility of the data and the local relevance of the quality indicators are paramount.

In the longitudinal evaluation, the survey data from the primary site demonstrated a significant improvement in perceptions of quality indicators, feedback, data use and overall effectiveness of quality monitoring between baseline and implementation of basic feedback. Furthermore, for the majority of the survey measures, there was a significant improvement at the secondary study sites between the baseline condition and implementation of the enhanced feedback protocol.

Interrupted time series analysis

Hypothesis 1 was concerned with the effect of implementing basic feedback where no prior feedback existed. It stated that implementation of routine departmental anaesthetic quality monitoring and basic feedback at the level of the individual consultant anaesthetist would improve professional practice and improve anaesthetic care. Improvement would be detectable by a positive change in perioperative normothermia, postoperative nausea and pain, patient-reported QoR, SSI rates and 30-day surgical mortality indicators.

The observed response to the implementation of basic feedback in the quality indicators assessed was, on the whole, limited and the hypothesised benefits of implementing basic feedback were generally not observed in the data. While the average weekly temperature of patients arriving in recovery increased by 0.082 °C in response to the onset of basic feedback, in contrast, the weekly proportion of patients arriving in recovery with temperature < 36 °C increased between the baseline and basic feedback condition, too, by 3.3%. Patient-reported QoR Scale scores showed no response to the implementation of basic feedback, with evidence to suggest that the proportion of patients reporting high scores may actually have decreased following the implementation of basic feedback. Patient-reported freedom from severe postoperative pain did not respond to the onset of basic feedback and the proportion of patients reporting freedom from postoperative nausea appeared to decline between the baseline and basic feedback epochs. No significant effects of the implementation of basic feedback were observed for either SSI rate or 30-day postoperative mortality rate. For the vast majority of measures, hypothesis 1 is, therefore, rejected, with the exception of the apparent effect of introduction of basic feedback on the average patient temperature on arrival in recovery. Examination of longitudinal variation in the baseline data set, however, offered a potential alternative explanation for the largely negative results of implementing basic feedback, in terms of issues with the reliability of data collection and data administration in the earlier phases of the study.

Hypothesis 2 was concerned with the effect of implementing enhanced feedback in a group that had been receiving routine basic feedback previously. It stated that provision of enhanced feedback, including detailed case category breakdown, specialty-specific information, deviant case details, enhanced comparative and longitudinal data, and institution-wide dissemination would facilitate personal professional learning from case experience and stimulate further improvement in quality of anaesthetic care.

Escalating the intensity of feedback through implementation of an enhanced feedback protocol had a positive effect across a greater range of measures than implementation of basic feedback. After implementation of enhanced feedback, patients were on average warmer on arrival in recovery by 0.064 °C, though the rate of change in mean patient temperature got slightly worse. Again, the mean patient temperature and proportion of patients with temperature < 36 °C metrics did not show consistent results for the effects of enhanced feedback, with the latter demonstrating a detrimental response in the rate of change over time. Both the mean patient-reported QoR Scale score and the proportion of patients reporting high-quality recovery showed small improvements in the rate of change between basic and enhanced feedback conditions (change in trend for mean scale score = 0.009 on a 17-point scale and change in trend for proportion of patients = 0.001%).

The two measures of postoperative pain demonstrated consistent positive responses to the implementation of the enhanced feedback protocol, with significant improvement in both level and rate of change in the proportion of patients reporting freedom from severe pain (change in level = 7.2% of patients; change in trend = 0.004%) and those reporting no or mild pain on arrival in recovery (change in level = 12% of patients; change in trend = 0.003%). The proportion of patients with nurse-reported absence of nausea similarly increased by 5.8% in response to the implementation of enhanced feedback, coupled with an improvement in the rate of change in this measure (change in trend = 0.001%), though no significant effect was detected for patient-reported freedom from postoperative nausea. No significant effect of the implementation of enhanced feedback on SSI rate was detected. Thirty-day postoperative mortality appeared to show a complex response to the implementation of enhanced feedback, with a modest increase in level (0.8%) coupled with a modest improvement in the rate of change over time (change in trend = -0.001%).

Hypothesis 3 stated that implementation and escalation of performance feedback at the level of the individual consultant anaesthetist would improve professional practice in anaesthetists ranked lowest at baseline and that any beneficial effects of feedback will be strongest for this subgroup, compared with the full cohort. Based on the study findings, the third study hypothesis is supported for the effects of enhanced feedback on anaesthetic quality indicator scores for lower-ranked anaesthetists, in the areas of postoperative pain management, control of postoperative nausea and overall patient-reported QoR.

Taken as a whole, the implementation of enhanced feedback, as an escalation of the intensity of a basic feedback programme, was found to have a significant positive impact on a broad range of perioperative quality indicators, including mean patient temperature and QoR Scale score. The strongest apparent positive impact of enhanced feedback, however, was demonstrated for postoperative pain measures, both self-reported for duration in recovery and on arrival in recovery, and for nurse-recorded freedom from postoperative nausea. The observed positive effects of enhanced feedback on patient temperature data, mean QoR score, and measures of both postoperative pain and postoperative nausea were robust after covariate analysis in which longitudinal variation in disease severity, patient age and gender were controlled. When further statistical models were fitted based on stricter case inclusion criteria, limited to elective GA cases, despite the loss in statistical power, the implementation of enhanced feedback was still associated with significant improvement in level and slope in both proportion of patients with nurse-reported freedom from nausea and freedom from pain on arrival in recovery, lending further confidence in these findings.

TABLE 77 Summary of findings from ITSA structured by anaesthetic quality indicator and main evaluative hypothesis

Anaesthetic quality indicator	Dependent measure used in analysis (and unit)	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
Patient temperature	Mean patient temperature on arrival in recovery (°C)	Improvement in level observed After controlling for pre-intervention trends, patients were on average 0.082 °C warmer following implementation of basic feedback	Complex effect observed After controlling for pre-intervention trends, patients were on average 0.064 degrees warmer following implementation of enhanced feedback, but the rate of change in temperature over time got worse (change in trend = -0.002 °C)
	Proportion of patients arriving in recovery < 36 °C (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 3.3% fewer patients arrived in recovery with a temperature above the recommended 36 °C following implementation of basic feedback	Detrimental shift in rate of change observed After controlling for pre-intervention trends, the rate of change in the proportion of patients arriving in recovery with a temperature above the recommended 36 °C got worse following implementation of enhanced feedback (change in trend < 0.1%)
QoR	Mean QoR Scale score (0– 6; low- to high-quality recovery)	No significant effect observed After controlling for pre-intervention trends, no change in mean patient QoR Scale score was observed following implementation of basic feedback	Improvement in rate of change observed After controlling for pre-intervention trends, the rate of change in mean patient QoR Scale score improved following implementation of enhanced feedback (change in trend = 0.009 scale score)
	Proportion of patients with QoR score above 14 (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 9.4% fewer patients reported high QoR Scale scores (above 14) following implementation of basic feedback	Improvement in rate of change observed After controlling for pre-intervention trends, the rate of change in the proportion of patients reporting high QoR Scale scores (above 14) improved following implementation of enhanced feedback (change in trend = 0.001%)

TABLE 77 Summary of findings from ITSA structured by anaesthetic quality indicator and main evaluative hypothesis (*continued*)

Anaesthetic quality indicator	Dependent measure used in analysis (and unit)	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
Postoperative pain	Proportion of patients reporting freedom from severe pain during recovery (QoR item) (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in the proportion of patients reporting freedom from severe pain or constant moderate pain during recovery was observed following implementation of basic feedback	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 7.2% more patients reported freedom from severe or constant moderate pain during recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients reporting freedom from pain improved following implementation of enhanced feedback (change in trend = 0.004%)
	Proportion of patients reporting no or mild pain on arrival in recovery (excluding baseline period) (% proportion)	Not evaluated owing to absence of baseline data	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 12% more patients reported no or mild pain on arrival in recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients reporting no or mild pain improved following implementation of enhanced feedback (change in trend = 0.003%)
Postoperative nausea	Proportion of patients reporting freedom from PONV during recovery (QoR item) (% proportion)	Detrimental change in level observed After controlling for pre-intervention trends, on average, 4.6% fewer patients reported freedom from PONV following implementation of basic feedback	No significant effect observed After controlling for pre-intervention trends, no change in the proportion of patients reporting freedom from PONV was observed following implementation of enhanced feedback
	Proportion of patients with nurse-recorded absence of nausea during stay in recovery (excluding baseline period) (% proportion)	Not evaluated owing to absence of baseline data	Improvement in level and in rate of change observed After controlling for pre-intervention trends, on average, 5.8% more patients had nurse-recorded absence of nausea during their stay in recovery following implementation of enhanced feedback. Furthermore, the rate of change in the proportion of patients with nurse-recorded absence of nausea improved following implementation of enhanced feedback (change in trend = 0.001%)

continued

TABLE 77 Summary of findings from ITSA structured by anaesthetic quality indicator and main evaluative hypothesis (*continued*)

Anaesthetic quality indicator	Dependent measure used in analysis (and unit)	Effect of implementation of basic feedback (H1)	Effect of implementation of enhanced feedback (H2)
SSI rate	Proportion of patients with a SSI-specific diagnostic code (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in SSI rate was observed following implementation of basic feedback	No significant effect observed After controlling for pre-intervention trends, no change in SSI rate was observed following implementation of enhanced feedback
Mortality	Proportion of patients with 30-day mortality following surgery (% proportion)	No significant effect observed After controlling for pre-intervention trends, no change in 30-day postoperative mortality rate was observed following implementation of basic feedback	Complex effect observed After controlling for pre-intervention trends, on average, 30-day postoperative mortality rate increased in level (0.8%) following implementation of enhanced feedback, but the rate of change in mortality rate over time improved slightly (change in trend < -0.001%)

H, hypothesis.

Beneficial effects of feedback are shaded dark green, detrimental effects are shaded blue, unambiguous or complex effects are shaded light green and those not evaluated are shaded grey.

Productivity analysis

Although most perioperative ward leads described the initiative as useful and recognised the importance of active clinical engagement, significant improvements in WWT were reflected in only three of the eight wards studied. Qualitative analysis of stakeholder interviews suggested that although they valued the feedback initiative, further organisational changes would be needed to progress improvement in ward transfer efficiency, including improvements to the local bed allocation and discharge process. Although there was strong consensus of support for the initiative, the interviews highlighted the need for an infrastructure to support change and shared goals targeted through a system-wide approach. A strong theme required to achieve this was that of culture change: proactivity, forward thinking and understanding of hospital dynamics. The overall benefits of this initiative were perceived to influence patient satisfaction, patient flow and patient experience as well as staff dynamics and the organisation as a whole.

Conclusions and implications for health care

Taken as a whole, the findings from this evaluation provide rich information concerning the effects of a comprehensive, long-term anaesthetic quality monitoring and feedback initiative on multiple dimensions of service performance. Furthermore, they provide insight into the process of development that took place within this initiative, of interactions between context, intervention and user, and document the experiences and perceptions of the anaesthetists that participated as end-users and co-designers of the feedback.

Although the evaluation was unable to demonstrate positive impact of implementing a simple, passive monthly feedback report on baseline perioperative indicators, convincing longitudinal benefits were demonstrated where investment in refinement and enhancement of the feedback protocol was made in collaboration with end-users and within a broader framework of active engagement and constructive dialogue centred around the data. Findings from the qualitative research and evaluative survey demonstrate the positive response of clinicians to this type of initiative and their willingness to interact with a sustained and comprehensive information system in order to understand variations in care delivered personally and in collaboration with colleagues.

From a practical perspective, in the development of this intervention, what was surprising was the number of routinely collected perioperative administrative data residing in various hospital systems and databases that were not effectively integrated and which could potentially have been harnessed sooner, to provide real-time intelligence on variations in care within a clinically facing information system. Just as the inputs and processes that contribute to variations in the performance of care delivery systems are complex and multivariate in nature, the monitoring and control systems that are put in place to regulate these outputs must demonstrate similar requisite variety. While the challenges of reliable measurement of patient satisfaction with anaesthesia and the experience of postoperative pain and recovery are well established in the research literature, our research has highlighted the considerable scope and opportunity that exists for future research and development to address the questions of how we can best use the data resulting from quality monitoring initiatives and how we transform data into intelligence for specific professional groups.

The research findings give rise to the following implications and recommendations for development in health care:

1. The research suggests that quality monitoring and feedback interventions of the type implemented in this study represent an important quality improvement mechanism, especially where investment is made in their long-term development and sustainment.
2. The findings suggest that the design of feedback, its perceived intent, fitness for purpose and context of use are all important considerations for success.
3. It is essential to not only involve end-users in the development of the feedback system at conception, but to foster an ongoing sense of ownership of the data and a willingness to interact with them.
4. It is important to pair passive data dissemination with support, active engagement and opportunities for intra- and interprofessional dialogue, concerning how to respond to evidence of variations and opportunities for improvement.
5. Continuous feedback can make the natural variation inherent in human-intensive processes, such as health care, visible to improvement efforts. In so doing, subjective and intangible phenomena, such as patient satisfaction, may be objectified for more constructive conversations, enhanced shared decision-making and better control.
6. In the development of monitoring and feedback systems, appropriate attention must be given to how data are used and converted into information for specific user groups, such as clinicians, rather than simply focusing on what to measure and how reliable those measures are.
7. The success of quality improvement interventions such as the data feedback initiative studied in this research should be evaluated using multiple dimensions, including social, organisational and professional outcomes, in addition to clinical end-points.
8. While downstream postoperative outcomes may be insensitive to the effects of an anaesthetic quality feedback intervention, process-of-care measures, such as those associated with postoperative pain management, nausea and perioperative normothermia, are more receptive.
9. Within the health informatics field, considerable scope exists beyond this project to further test evolving theory and practice from improvement science and industrial process control related to how data can be used to support continuous improvement in process-based operations.
10. The trend towards a shift away from intermittent, snapshot audits of practice in favour of a continuous monitoring and continuous improvement model within health-care organisations should be the subject of further investigation in terms of its implications for patients and quality of care.

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library