

# **A methodological study to compare survey-based and observation-based evaluations of organisational and safety cultures and then compare both approaches with markers of the quality of care**

D Freeth, J Sandall, T Allan, F Warburton,  
EJ Berridge, N Mackintosh, M Rogers and  
S Abbott



May 2012  
10.3310/hta16250

**Health Technology Assessment**  
**NIHR HTA programme**  
[www.hta.ac.uk](http://www.hta.ac.uk)





### **How to obtain copies of this and other HTA programme reports**

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

### **Contact details are as follows:**

Synergie UK (HTA Department)  
Digital House, The Loddon Centre  
Wade Road  
Basingstoke  
Hants RG24 8QW

Email: [orders@hta.ac.uk](mailto:orders@hta.ac.uk)

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’  
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

### **Payment methods**

#### *Paying by cheque*

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

#### *Paying by credit card*

You can order using your credit card by phone, fax or post.

### **Subscriptions**

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

### **How do I get a copy of HTA on DVD?**

Please use the form on the HTA website ([www.hta.ac.uk/htacd/index.shtml](http://www.hta.ac.uk/htacd/index.shtml)). *HTA on DVD* is currently free of charge worldwide.

---

The website also provides information about the HTA programme and lists the membership of the various committees.

# A methodological study to compare survey-based and observation-based evaluations of organisational and safety cultures and then compare both approaches with markers of the quality of care

D Freeth,<sup>1\*</sup> J Sandall,<sup>2</sup> T Allan,<sup>3</sup> F Warburton,<sup>4</sup>  
EJ Berridge,<sup>3</sup> N Mackintosh,<sup>5</sup> M Rogers<sup>3</sup> and S Abbott<sup>3</sup>

<sup>1</sup>Centre for Medical Education, St Bartholomew's and The Royal London School of Medicine and Dentistry, Queen Mary University London, London, UK

<sup>2</sup>Division of Women's Health, School of Medicine, King's College London, London, UK

<sup>3</sup>School of Health Sciences, City University London, London, UK

<sup>4</sup>Department of Psychological Medicine, Institute of Psychiatry, King's College London, London, UK

<sup>5</sup>Patient Safety and Service Quality Research Centre, King's College London, London, UK

\*Corresponding author

**Declared competing interests of authors:** none

Published May 2012

DOI: 10.3310/hta16250

---

This report should be referenced as follows:

Freeth D, Sandall J, Allan T, Warburton F, Berridge EJ, Mackintosh N, *et al.* A methodological study to compare survey-based and observation-based evaluations of organisational and safety cultures and then compare both approaches with markers of the quality of care. *Health Technol Assess* 2012;**16**(25).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

#### Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA programme in April 2007 under the newly established NIHR Methodology Panel. The HTA programme project number is 06/92/06. The contractual start date was in September 2007. The draft report began editorial review in January 2011 and was accepted for publication in July 2011. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE

Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Tom Marshall, Professor John Powell, Dr Rob Riemsma and Professor Ken Stein

Associate Editor: Dr Peter Davidson

Editorial Contact: [edit@southampton.ac.uk](mailto:edit@southampton.ac.uk)

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

ISSN 2046-4932 (DVD)

© Queen's Printer and Controller of HMSO 2012. This work was produced by Freeth *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health.

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

This journal may be freely reproduced for the purposes of private research and study and 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: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://www.prepress-projects.co.uk)), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by Charlesworth Press.

# Abstract

## A methodological study to compare survey-based and observation-based evaluations of organisational and safety cultures and then compare both approaches with markers of the quality of care

D Freeth,<sup>1\*</sup> J Sandall,<sup>2</sup> T Allan,<sup>3</sup> F Warburton,<sup>4</sup> EJ Berridge,<sup>3</sup> N Mackintosh,<sup>5</sup> M Rogers<sup>3</sup> and S Abbott<sup>3</sup>

<sup>1</sup>Centre for Medical Education, St Bartholomew's and The Royal London School of Medicine and Dentistry, Queen Mary University London, London, UK

<sup>2</sup>Division of Women's Health, School of Medicine, King's College London, London, UK

<sup>3</sup>School of Health Sciences, City University London, London, UK

<sup>4</sup>Department of Psychological Medicine, Institute of Psychiatry, King's College London, London, UK

<sup>5</sup>Patient Safety and Service Quality Research Centre, King's College London, London, UK

\*Corresponding author

**Background:** Patient safety concerns have focused attention on organisational and safety cultures, in turn directing attention to the measurement of organisational and safety climates.

**Objectives:** First, to compare levels of agreement between survey- and observation-based measures of organisational and safety climates/cultures and to compare both measures with criterion-based audits of the quality of care, using evidence-based markers drawn from national care standards relating to six common clinical conditions. (This required development of an observation-based instrument.) Second, to examine whether observation-based evaluations could replace or augment survey measurements to mitigate concerns about declining response rates and increasing social desirability bias. Third, to examine mediating factors in safety and organisational climate scores.

**Design:** The study had three strands: (A) a postal questionnaire survey to elicit staff perceptions of organisational and safety climates, using six prevalidated scales; (B) semistructured non-participant observation of clinical teams; and (C) a retrospective criterion-based audit carried out by non-clinical auditors to minimise hindsight bias. Standardised summary scores were created for each strand, and pairs of measurements were compared (strand A with strand B, strand A with strand C, and strand B with strand C) using Bland–Altman plots to evaluate agreement. Correlations were also examined. Multilevel modelling of Strand A scores explored mediating factors.

**Setting and participants:** Eight emergency departments and eight maternity units in England, UK.

**Interventions:** None.

**Main outcome measures:** Examination of feasibility, correlation and agreement.

**Results:** Strand A: the overall response rate was 27.6%, whereas site-specific rates ranged from 9% to 47%. We identified more mediating factors than previous studies; variable response rates had little effect on the results. Organisational and safety climate scores were strongly correlated ( $r=0.845$ ) and exhibited good agreement [standard deviation (SD)

differences 0.449; 14 (88%) within  $\pm 0.5$ ; one large difference]. Two commonly used scales had high levels of positive responses, suggesting positive climates or social desirability bias. Strand B: scoring on a four-point scale was feasible. Observational evaluation of teamwork culture was good but too limited for evaluating organisational culture. Strand C: a total of 359–399 cases were audited per condition. The results varied widely between different markers for the same condition, so selection matters. Each site performed well on some markers but not others, with few consistent patterns. Some national guidelines were contested. Comparisons: the comparison of safety climate (survey) and teamwork culture (observation) revealed a moderately low correlation ( $r=0.316$ ) and good agreement [SD differences 1.082; 7 (44%) within  $\pm 0.5$ ; one large difference]. The comparison of safety climate (survey) and performance (audit) revealed lower correlation ( $r=0.150$ , i.e. relationship not linear) but reasonably good agreement [SD differences 0.992; 9 (56%) within  $\pm 0.5$ ; two large differences]. Comparisons between performance (audit) and both organisational climate (survey) and teamwork culture (observation) showed negligible correlations ( $<0.1$ ) but moderately good agreement [SD differences 1.058 and 1.241; 6 (38%) and 7 (44%) within  $\pm 0.5$ ; each with two large differences (at different sites)]. Field notes illuminated large differences.

**Conclusions:** Climate scores from staff surveys are not unduly affected by survey response rates, but increasing use risks social desirability bias. Safety climate provides a partial indicator of performance, but qualitative data are needed to understand discrepant results. Safety climate (surveys) and, to a lesser degree, teamwork culture (observations) are better indicators of performance than organisational climate (surveys) or attempts to evaluate organisational culture from time-limited observations. Scoring unobtrusive, time-limited observations to evaluate teamwork culture is feasible, but the instrument developed for this study needs further testing. A refined observation-based measure would be useful to augment or replace surveys.

**Funding:** The National Institute for Health Research Health Technology Assessment programme.

# Contents

<b>List of abbreviations</b>	<b>ix</b>
<b>Executive summary</b>	<b>xi</b>
<b>Part I</b> Context and study design	<b>1</b>
<b>1. Introduction</b>	<b>3</b>
Purpose and context	3
Funding for this study	5
Aims and objectives	5
Hypotheses	6
Study design	7
<b>2. Safety and organisational cultures and climates</b>	<b>11</b>
Introduction	11
Nature of the problem	11
Measuring culture	11
Organisational and safety cultures: concepts and definitions	12
Safety culture and climate	15
Selecting organisational levels for research into safety culture and climate	16
Quality: case note review	16
Comparing measurements of culture and evaluations of the quality of care	17
Conclusion	18
<b>3. Methods for data collection and analysis</b>	<b>19</b>
Comparison of data sets: threshold correlation, power calculations and clustering	19
Selecting the sample of research sites	19
Strand A: staff survey	20
Strand B: observation-based holistic evaluation of safety culture	25
Strand C: criterion-based assessment of quality of care	28
Testing the study hypotheses: correlation and agreement	32
<b>Part II</b> Results	<b>37</b>
<b>4. Process data and issues arising during the study</b>	<b>39</b>
Recruiting research sites	39
The sequence of data collection	40
Close-coupling of departments and services	41
Impact of staff changes in research team	41
Research governance	41
Strand A: staff survey	42
Strand B: observation-based assessment of culture	46
Strand C: criterion-based audit of clinical notes	49
Timely feedback of findings for local use	53
Summary	54

<b>5. Strand A: questionnaire survey of staff perceptions of safety and organisational climates</b>	<b>55</b>
Demographics and role-related characteristics	55
Support for training, learning and development	56
Organisational climate	57
Safety climate and team factors	61
Comparing survey-based organisational and safety climate scores	69
Summary	70
<b>6. Strand B: observation-based holistic assessment of culture</b>	<b>73</b>
Data collection summary	73
Scoring	73
Observation-based evaluations of organisational, teamwork and safety cultures	75
Comparing observation-based organisational and teamwork culture scores	75
Summary	76
<b>7. Strand C: retrospective audit of evidence-based markers of the quality of care</b>	<b>79</b>
Condition-specific audit results	79
Summarising site-specific audit results	83
Summary	83
<b>8. Comparison results from strands A–C</b>	<b>85</b>
Comparing survey-based assessments with observation-based assessments	85
Comparing climate and culture assessments with the quality of care	90
Summarising the comparisons	97
<b>Part III Discussion, conclusions and recommendations</b>	<b>99</b>
<b>9. Discussion</b>	<b>101</b>
Revisiting the study's original aims, objectives and hypotheses	102
To what extent might the empirical results be considered predictable?	115
Challenges identified during this study	116
Strengths of this study	121
Limitations of this study	122
<b>10. Conclusions</b>	<b>125</b>
Survey-based assessments of organisational and safety climate	126
Making observation-based assessments of organisational and safety culture	127
Making audit-based assessments of the quality of care	129
Comparing different approaches to measuring climate, culture and the quality of care	130
Summary	131
Future research	133
<b>Acknowledgements</b>	<b>135</b>
<b>References</b>	<b>137</b>



<b>Appendix 1</b> Extracts from the tender specification	<b>151</b>
<b>Appendix 2</b> Study protocol	<b>157</b>
<b>Appendix 3</b> Staff survey	<b>165</b>
<b>Appendix 4</b> New variables created for summary survey-based organisational climate scores	<b>171</b>
<b>Appendix 5</b> Strand B data collection prompt sheet	<b>173</b>
<b>Appendix 6</b> Scoring guide for observation-based evaluation of culture	<b>175</b>
<b>Appendix 7</b> Data extraction sheet for fractured neck of femur (audit)	<b>177</b>
<b>Health Technology Assessment programme</b>	<b>179</b>



## List of abbreviations

ACS	acute coronary syndrome
AHRQ	Agency for Healthcare Research and Quality
ASA	acute severe asthma
CPD	continuing professional development
CRB	Criminal Records Bureau
CVF	Competing Values Framework
df	degrees of freedom
DU	delivery unit
ECS	emergency caesarean section
ED	emergency department
EFM	electronic fetal monitoring
FNoF	fractured neck of femur
ICC	intracluster correlation coefficient
IQR	interquartile range
MaPSaF	Manchester Patient Safety Framework
MI	myocardial infarction
MSL	meconium-stained liquor
ND	normal delivery
NIGB	National Information Governance Board
NPSA	National Patient Safety Agency
P–P	probability–probability (plot)
PCT	primary care trust
QIIS	Quality Improvement Implementation Survey
<i>r</i>	Pearson product–moment correlation coefficient
REC	research ethics committee
SAQ	Safety Attitudes Questionnaire
SD	standard deviation
SE	standard error
VTE	venous thromboembolism

---

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.



# Executive summary

## Background

Concern about avoidable harm during health care has focused attention on actions and values that may promote safety, including the identification and promotion of a safety culture. This is a multifaceted concept, usually understood to be a subset of organisational culture, encompassing norms, attitudes and values that underpin and reinforce behaviours which promote or inhibit safe care. The multifaceted and partly invisible natures of organisational and safety cultures render them awkward to capture and measure.

Ethnographic studies have provided thick description and important insights relating to key interactions within health care. Self-assessment frameworks have been developed to prompt and guide practice development within health-care teams. A range of questionnaires, containing mainly Likert-scale items, have become popular for measuring facets of culture: their popularity may become their downfall as health professionals tire of completing them and knowledge develops about 'correct' (socially desirable) answers, or if use slips from self-assessment and local quality improvement to targets and league tables.

The promotion of positive organisational and safety cultures is aimed at improving the quality and safety of care. However, evidence of any such link is conflicting and generally suggests only a weak link. The methodological challenges of comparing culture with clinical outcomes are substantial. Cultures are multifaceted and vary at different organisational levels, perhaps even from team to team and from shift to shift within clinical environments. Assessments of culture may vary according to the facets that are measured. It is not possible to identify a best measure of the quality of care, and the choice between outcome and process measures is also complicated. Much work remains to be done to develop ways to compare variations in culture with variations in the quality of care.

## Objectives

Our aim was to compare contrasting methods of assessing culture, and to compare each with an assessment of the quality of care by:

- assessing organisational and safety climates, using staff surveys, and exploring mediating factors
- developing quantified evaluations of safety culture and, if possible, organisational culture, using time-limited observations, and examining the coverage and feasibility of assessments
- comparing these two approaches to evaluating facets of organisational and safety cultures
- using criterion-based audit to evaluate the quality of care, and comparing results with survey-based assessments of climate and observation-based assessments of culture.

## Methods

Data were collected in 16 clinical departments, the emergency department (ED) and delivery unit (DU) (labour ward) of eight hospitals within six English strategic health authority areas. The strands of data collection were as follows:

- A. A postal questionnaire survey of staff perceptions of organisational and safety climates. This used prevalidated questions from the annual NHS staff survey and teamwork and safety climate scales developed in the USA. Contextual data about the research sites, demographic and role-related variables were used in multilevel modelling to establish whether site-level or individual-level factors would mediate staff perceptions of organisational climate and safety climate.
- B. Semistructured non-participant observation of work in non-treatment, workload management areas of the clinical departments, for example near a whiteboard used to monitor occupancy and the progress of care. ED observations occurred in the 'major injuries' section, to match the severity of the audited conditions (strand C). To include fluctuating workloads and staffing, observation sampling took place on 6 days of the week (Monday to Saturday) between 6 AM and midnight. Observations were recorded in field notes. Each clinical department was independently observed by two researchers. Service provider coresearchers made observations in eight departments (50%) and a service user coresearcher observed in three. Observations were augmented by an initial orientation interview with a senior member of staff; occasional brief conversations with members of the clinical team when this did not disturb their work; and (for 10 clinical departments which took up an offer made to all departments) feedback discussions with senior staff or the wider team.
- C. Retrospective criterion-based audits of markers (selected from national guidelines) of the quality of care for three commonly encountered conditions which had not recently been the subject of national audit or quality improvement initiatives. Audits were conducted by non-clinical staff to minimise hindsight bias. For DUs the audited conditions were normal labour and delivery (ND), emergency caesarean section (ECS) and birth following detection of grade 2 or 3 meconium-stained liquor during labour (MSL). For EDs the audited conditions were acute coronary syndrome (ACS), acute severe asthma (ASA) and fractured neck of femur (FNoF). Selection of audit markers was influenced by the grade of related evidence and discussion with experienced clinicians: each was considered clinically important. Some markers concerned treatment interventions, whereas others concerned making and recording observations.

First, data from staff surveys, observations and audits were analysed separately. Second, standardised summary scores were created for each strand. Third, pairs of measurements were compared: within strand A; strand A with strand B; strand A with strand C; and, finally, strand B with strand C. A correlation coefficient scatterplot and a Bland–Altman plot were inspected for each comparison.

## Results

### *Research process issues*

As this was a methodological study, research processes were examined, as well as outcomes. These included experiences of recruiting research sites and subsequent organisation-specific research governance processes, which were very variable and caused significant delays; the implications of close-coupling of clinical departments and other services contributing to care pathways; the impact and management of changes in the research team; and processes for each strand of data collection and analysis.

### *Strand A*

We analysed 531 questionnaires. The overall response rate, 27.6%, was in line with similar UK studies. Site-specific response rate (range 9–47%, interquartile range 23–36%) was not a significant explanatory variable in multilevel analyses of scores for five out of six prevalidated

scales, generally allaying concerns about the effect of variable response rates (departments with above average response rates returned lower line management scores). The demographic profile of respondents broadly reflected the wider National Health Service (NHS) workforce for gender and ethnicity. Most (89%) staff survey participants held clinical roles, and 37% managed other staff.

Four scales drawn from the annual NHS staff survey produced normally distributed results. The American Teamwork Climate and Safety Climate scales produced slightly skewed results, owing to high levels of agreement. Their items closely reflect the foci of numerous patient safety improvement initiatives. High levels of agreement may represent the development of positive teamwork and safety climates. However, some degree of social desirability bias may be included in these scores.

Multilevel modelling found several demographic, role-related and professional development variables which mediated perceptions of organisational and safety climates. The mediating variables differed between scales but included, for example, being a manager, gender, age, years of employment with the organisation, professional group, ethnicity and participation in certain types of continuing professional development. At departmental level, DUs perceived higher organisational error wisdom than EDs. We found no previous studies that investigated the mediating effects of a similarly wide range of factors.

Multilevel modelling indicated that scores on the scales selected to elicit perceptions of aspects of organisational climate and safety climate predominantly varied at the level of the clinical department, not at hospital level.

Standardised summary organisational and safety climate scores were strongly correlated ( $r=0.845$ ) and exhibited good agreement (evaluated by inspection of a Bland–Altman plot): it seems that staff found it difficult to view one aspect of culture positively (or negatively) without perceiving the other aspect of culture similarly. Only one site recorded a large difference between organisational and safety climate scores. Field notes illuminated this difference, suggesting the importance of qualitative data for understanding relationships between different measurements.

## **Strand B**

Observation averaged 31 hours per research site, mainly comprising 12 visits by research fellows. The participation of service provider coresearchers was valuable, particularly while new non-clinical researchers were developing expertise for trustworthy observations in clinical departments. The participation of a service user coresearcher helped ensure that researchers' observations and interpretations did not become too closely aligned with clinical perspectives. Onerous research governance procedures inhibited the participation of service user coresearchers.

Reactivity to non-participant observers in clinical departments is minimal, possibly because work in these areas is constantly observed by a range of stakeholders.

Semistructured observations were guided by a data collection prompt sheet: initially this comprised a synthesis of relevant concepts and behaviours identified by earlier research; later it was streamlined through use during the study. Scoring used a four-point scale after observations had been grouped into eight domains. The instrument developed during this study needs to be tested in other contexts and could benefit from development to increase the granularity of scoring, thereby increasing discrimination.

### **Strand C**

The number of cases audited for each condition was as follows: ACS 383, ASA 389, FNoF 384, ND 395, ECS 399 and MSL 359.

The methodological preference for non-clinical auditors caused delays in data collection. Although funding was available, trusts struggled to identify people to undertake audit work. Difficulties were compounded by a climate of research regulation, which strongly discourages research access to clinical notes, although recent guidance aims to increase the use of clinical records in health services research.

Some markers are much easier to audit than others, and some research sites' clinical records were much easier to audit than others, creating variable resource requirements. The quality of case notes, although variable, was adequate to assess recorded adherence to markers of clinically important evidence-based practice.

Audit results varied widely between different markers of the quality of care for the same clinical condition, both within and between research sites. This suggests, first, that, for an evaluation of the overall quality of care, audit results from several markers ought to be averaged to smooth variation between markers and, second, it is important not to lose sight of the variation between compliance with different markers of the quality of care because each marker is important and poor adherence could be clinically significant.

Audit markers that addressed repeating important clinical observations exhibited poor compliance, suggesting that this area of clinical practice (or recording clinical practice) may need further support. The design of (paper and electronic) clinical record forms can support or inhibit the recording of key observations or interventions, and may influence clinical actions.

Audit markers that required multiple observations sometimes yielded zero compliance. Other authors have discussed how rules in practice are so numerous that they quickly exceed individuals' ability to act on them. Staff then have to choose which rules to ignore, and may also ignore others without realising that they are doing so. This suggests that multifaceted national care standards may benefit from scrutiny to establish whether they could be streamlined.

### ***Comparisons between measurements arising from different strands of data collection***

Observation-based assessment of teamwork climate exhibited reasonably good agreement with survey-based assessment of safety climate. After further testing and refinement the observation instrument may become a viable alternative to staff surveys of safety climate, where the latter are considered unsatisfactory, for example due to declining response rates or concerns about social desirability bias or cognitive dissonance. At present, the observation instrument and associated field notes are useful to complement surveys and illuminate discrepant results.

Neither relationships between survey-based and observation-based assessments of culture nor relationships between these and clinical audit results are easily interpreted linear relationships. Consequently, separate measurement of culture and care standards is needed, and further research will be required to understand the contribution of one to the other. This study's summary survey-based assessments of safety climate provided the closest levels of agreement with summary audit scores (representing the quality of care), suggesting a continued role for safety climate surveys. When an observation-based assessment of culture is preferred, the instrument developed during this study is practical to use and exhibits moderately good agreement with summary audit scores, which is consistent with earlier research suggesting a weak relationship.



## Conclusions

### *Questionnaire-based assessments of climate*

- Surveys elicit variable response rates. Although it is prudent to be vigilant, this may not negate the usefulness of the measurements of organisational and safety climates.
- Social desirability bias may affect some well-known scales.
- Many factors mediate perceptions of organisational and safety climates. The detail and consequences of this require further research.
- Perceptions of organisational and safety climate are highly correlated and in close agreement: it is not known whether one is antecedent to the other.
- Perceptions of climate vary at the level of clinical departments rather than at hospital level.
- Survey-based assessments of safety climate exhibited close agreement with summary audit scores for most research sites, but large differences were found at two research sites. Observation field notes illuminated the differences, suggesting an important role for qualitative data alongside quantitative assessments.

### *Observation-based assessments of culture*

Based on 12 purposively sampled half-day visits by non-clinical observers:

- Safety-related facets of teamwork can be observed and scored. The observation-based measure of teamwork culture exhibited moderately good agreement with summary audit scores, consistent with the weak relationship described in other studies. The framework and scoring system developed during this study needs to be tested in other contexts, and attention given to greater discrimination in scoring. The current instrument is useful to augment survey assessments of safety climate. Following refinement, the instrument might be a suitable alternative to staff surveys of safety climate, particularly if survey response rates continue to fall, or concerns about social desirability bias in survey responses grow.
- Observation-based assessments of organisational culture proved too limited and offered insufficient differentiation in scores. Alternatives include staff surveys, since the prevalidated scales used to measure facets of organisational climate appear to function well, and organisational ethnographies.
- Non-clinical observers benefit from collaboration with service provider and service user observers, but research governance procedures discouraged the participation of service user observers.

### *Criterion-based audit of the quality of care*

- Current conditions in NHS trusts and research governance processes discourage audit-based research, particularly studies using non-clinical auditors. Recent guidance may mitigate many current problems, but the use of non-clinical auditors to minimise hindsight bias remains poorly supported.
- Markers requiring repeat observations showed poor compliance, suggesting this area as a target for practice development. Improving the design of some clinical record forms may help.
- Markers requiring many observations yielded few cases of full compliance but many cases in which a subset of observations had been made, suggesting that it may be beneficial to examine the evidence for multifaceted care standards and whether any could be streamlined.

- As the selection of audit markers affects performance results, careful choices must be made. It is important to avoid auditing only markers that are simple to assess. Use of markers employed by other studies will support comparisons. However, regularly audited markers will become targets, and so lose their potency as markers of the quality of care.

## Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

# Part I

## Context and study design

The enigmatic concept of ‘safety culture’ ranks high on the list of safety concerns in the safety movement.

(Lilford, 2010, p. 1)<sup>1</sup>



# Chapter 1

## Introduction

### Purpose and context

The purpose of the study was to compare measurements of organisational and safety cultures and the quality of care, employing two contrasting methods to measure culture. The list of detailed aims and objectives can be found in *Aims and objectives*.

Since the key document *An organisation with a memory*,<sup>2</sup> patient safety and the means of promoting it have been high priorities in NHS policy.<sup>3-5</sup> The Department of Health set up the National Patient Safety Agency in 2001, and funded a Patient Safety Research Portfolio, beginning in 2001, with the final tranche of its 50 commissioned studies reporting during 2010.<sup>1,6</sup> In parallel, the Health Foundation commissioned the Patient Safety Initiative (2004–8).<sup>7</sup>

Policy and academic work associated with patient safety draws on a large body of international research and practice development focused on patient safety that began in the early 1990s. Landmark reports<sup>8-10</sup> showed that avoidable harm to patients was occurring frequently in health care. Understanding of patient safety quickly drew in ideas developed from studies of other safety-critical endeavours, for example aviation, offshore drilling, and nuclear power and chemical plants.<sup>11-14</sup> The World Health Organization is supporting patient safety research across the world.<sup>15</sup>

*An organisation with a memory*<sup>2</sup> recognises that organisational factors play a key role in patient safety. Research suggests that significant organisational influence is transmitted not only or mostly through particular processes or directives but through organisational culture:<sup>16</sup> the term 'safety culture' has been used to denote those aspects of organisational culture that have particular relevance to the promotion of patient safety.<sup>17</sup> If a positive safety culture can contribute to safer care (or a negative safety culture can contribute to less safe care) we need to know how to recognise and promote positive safety cultures. *Chapter 2* synthesises a selection of the literature concerning safety culture.

Our definitions of organisational and safety culture are as follows. Organisational culture is the collection of shared beliefs, values and norms of behaviour found in an organisation;<sup>18</sup> safety culture is the subset of those values, beliefs and norms that relate to safety.<sup>19</sup> Climates are distinguished from cultures: organisational climate refers to the aggregate of individual perceptions, practices, policies, procedures and routines in an organisation, while safety climate refers to the aggregate of those perceptions as they relate to safety.<sup>20</sup> Climates are regarded as representing the surface manifestations of underlying cultures.<sup>21</sup>

Evaluating and measuring organisational and safety cultures may allow organisations to identify areas for improving patient safety. Qualitative or quantitative assessments of culture may identify priorities for quality improvement interventions and act as a baseline for subsequent assessments. Questionnaires and other diagnostic frameworks for self-evaluation have been developed for this purpose and are widely used (see *Chapter 2*). In the UK, the National Patient Safety Agency (NPSA) and the 'Patient Safety First!' campaign promoted the use of safety climate questionnaires and safety culture diagnostic frameworks for ongoing quality improvement in a wide range of

health care settings.<sup>22,23</sup> At the beginning of this study, the best-known and most extensively validated safety climate questionnaires were the Safety Attitudes Questionnaire (SAQ),<sup>24</sup> and the subset Teamwork and Safety Climate Survey;<sup>25</sup> we used the latter in this study (see *Chapter 3, Strand A: staff survey*).

Safety climate questionnaires are inexpensive to administer and can be analysed quickly, so if they measure something that promotes high-quality care they are a valuable resource. However, questionnaire scores can be susceptible to ‘social desirability bias’<sup>26</sup> or be invalidated by ‘cognitive dissonance’: the inability of staff to accept evidence that the service that they work in is less than safe.<sup>27</sup> Postal questionnaires, in particular, can elicit very low response rates or variable response rates depending upon the degree of interest recipients have in contributing their views on the topic of the questionnaire<sup>28</sup> and, if there are systematic differences between responders and non-responders, non-response bias influences results. Response rates among health-care professionals appear to be falling<sup>29,30</sup> and, since increased health-care delivery pressures and a rising number of requests to complete questionnaires or similar forms are thought to be probable causes, this trend looks set to continue.

In parallel with the development and promotion of questionnaires and other diagnostic frameworks, there has been investment in ethnographic studies focused on patient safety (see *Chapter 2, Ethnographic methods*, and Dixon-Woods<sup>31</sup>). Ethnographies focus on culture and illuminate microsocial interactions. They are thus well aligned with the concerns of the patient safety movement. However, ethnographies require participant observers to undertake lengthy immersion in the environment which is being studied, which makes them relatively expensive and slow to report. It is also difficult to conduct ethnographies on multiple sites. Consequently, interest in exploring whether less intensive observation can yield useful insights grew, alongside interest in whether any insights could be linked to the quality of care. Strategic observation combined with other data sources were used in clinical governance review and inspection,<sup>32</sup> and highly structured observations have been used to measure the success of specific quality improvement interventions; for example, Catchpole and colleagues<sup>33</sup> examined patient handover from surgery to intensive care. Our study required a quantifiable approach to evaluating culture ‘holistically’, using time-limited observations and yielding results that could be compared with survey-based assessments of climate and indicators of the quality of care. There were no well-established methods or prevalidated instruments for this task, so the iterative development of a suitable observation framework and scoring scheme formed a substantial part of the study. We used semistructured observation and brief key informant interviews, informed by an earlier ethnography conducted by three members of the study team.<sup>34</sup> The observation strand of this study will be described in *Chapter 3, Strand B: observation-based holistic evaluation of safety culture*.

At present, too little is known about how cultures and climates relate to the quality of care. Three studies<sup>16,35,36</sup> found limited evidence to support or to challenge the hypothesis that organisational culture and health-care performance are linked: where found, links tend to be contingent and complex. However, Singer and colleagues<sup>37</sup> suggested that aspects of safety climate and safety performance are related, while Silva and colleagues<sup>38</sup> found organisational climate and safety climate to be inversely correlated with the incidence of accidents.

Bringing together these strands in patient safety research, this study sought to compare organisational and safety cultures as measured by questionnaire and observational tools, and markers of quality of care. Thus, data were collected from eight consultant-led delivery units (DUs) and eight emergency departments (EDs) in England. A DU is one of the specialised clinical areas provided within the continuum of care offered by maternity services. Local terms vary, but DU and delivery suite were the most common terms used at the research sites in this

study. Occasionally, people referred to the labour ward. Some research sites preferred to refer to themselves as accident and emergency (A&E) departments, whereas others preferred ED. More recent literature and advice from practitioners suggested that the use of the term ED is growing and now outstrips the use of A&E, so ED will be used throughout this report.

The research sites were located in 6 out of 10 English strategic health authorities (see *Table 3*). There were three equally important strands of data collection (described in *Chapter 3*), named strands A, B and C for ease of later reference:

- strand A, 'staff survey': a postal questionnaire for staff at each research site to elicit perceptions of organisational and safety climates, and identify some of the factors that may influence these perceptions
- strand B, 'observation-based holistic evaluation': a profile of scores for organisational and team factors representing aspects of organisational and safety culture, derived from semistructured observations and brief key informant interviews
- strand C, 'audit': audits of evidence-based markers of the quality of care for three purposively selected conditions commonly encountered in the research sites.

These data sets allowed examination of different ways of capturing facets of organisational and safety cultures and comparison of findings. In particular, comparisons aimed to establish whether clinical departments with high (or low) scores for the facets of culture captured in strand A also score highly (or poorly) for the facets of culture captured in strand B: in effect asking whether these two approaches to measuring culture would agree. In addition, the study afforded the opportunity to compare evaluations of culture from strands A and B with markers of the quality of care, which were collected in strand C. These comparisons will explore whether clinical departments with high (or low) scores for the facets of culture captured in strand A or strand B also score highly (or poorly) against the selected markers of the quality of care, again exploring whether there is agreement between different approaches to assessment.

## Funding for this study

This study was designed in response to one of the final calls for research proposals from the Patient Safety Research Portfolio (see above) in January 2005. This commissioning round was halted as a result of a review and restructuring of Department of Health funding streams. Subsequently, a similar call was issued by the NHS R&D Methodology Programme in April 2006 (RM05/JH33; see *Appendix 1*). This study was funded to run from September 2007 to August 2010. Further restructuring of Department of Health research programmes led to this study being transferred to the MRC-NIHR portfolio and renumbered as 06/92/06. During the study, excessive delays occurred within research governance approvals processes at some research sites (see *Chapter 4, Research governance*). This was compounded by delays in identifying trust-based auditors for strand C and slower than anticipated progress with audits at most research sites (see *Chapter 4, Engaging non-clinical auditors*). Together, these factors necessitated the request and approval of a 3-month no-cost extension to the spending period for the study. The study was completed in November 2010.

## Aims and objectives

The aims and objectives detailed at the beginning of the study are listed below. Following a convention in the field that will be discussed in *Chapter 2*, this report will use the term 'climate'

rather than ‘culture’ when referring to assessments arising from questionnaires, for example in objective 2 below.

The aims of the study were:

- to compare questionnaire and holistic assessments of organisational and safety culture
- to compare assessments of organisational and safety culture with criterion-based assessment of the quality of care.

The tender specification (see *Appendix 1*) used the term ‘triangulate’, but several types of triangulation are defined in the research methods literature, and different forms of triangulation might be considered to apply to different parts of this study. The central purpose of the study was to *compare* different assessments of culture with each other and with markers of the quality of care, using quantified assessments. For clarity, *compare* is used in this report in preference to triangulate. The tender specification also used the term ‘holistic evaluation’, and this is reflected in the aims. The term ‘holistic evaluation’ is ambiguous and, for clarity, in this report we will reflect the dominant data collection method by substituting the term ‘observation-based assessment’. The tender specification used the term ‘generic’ where we will use organisational.

The objectives of the study were:

1. to work *with* staff in the participating trusts such that their organisational and professional knowledge was respected, the study was understood and supported within participating departments, and prompt feedback to participant departments allowed local development in advance of the study reporting to the wider health and research communities
2. to use questionnaires to obtain quantitative assessments of the organisational and safety climate at each site
3. to generate quantified holistic evaluations of organisational and safety culture for each site using observation
4. to obtain criterion-based measurements for the quality of care at each site
5. to compare levels of agreement between the questionnaire and holistic measurements of culture
6. to compare organisational culture with safety culture
7. to compare culture measurements and criterion-based measurements of the quality of care
8. to collect data such that, where sufficient respondents existed within a category to protect anonymity, data could be explored by stakeholder group (e.g. managers, midwives, nurses, doctors, allied health professionals, support staff) and level (e.g. management responsibility).

## Hypotheses

The initial aims and objectives were linked to six hypotheses to be tested following collection of primary data. The hypotheses were framed to examine whether different approaches to evaluating culture were correlated (had a linear relationship) and agreed (had the same value when measured on the same scale). For this study strong correlation was defined at the outset as 0.7 and moderately strong correlation as 0.4 (see *Chapter 3, Comparison of data sites: threshold correlation, power calculations and clustering*); agreement was evaluated from inspection of Bland–Altman plots (see *Chapter 3, Testing the study hypotheses: correlation and agreement*).

Comparing questionnaire assessments with holistic (observation-based) assessments:

- H1a There will be a strong correlation and good agreement between questionnaire-based and observation-based evaluations of *organisational* culture.



H1b There will be a strong correlation and good agreement between questionnaire-based and observation-based evaluations of *safety* culture.

Testing the relationship between organisational and safety climate/culture:

H2a There will be a strong correlation and good agreement between *questionnaire*-based evaluations of organisational and safety climates (for a discussion of the convention regarding the terms 'culture' and 'climate' see *Chapter 2, Safety culture and climate*).

H2b There will be a strong correlation and good agreement between *holistic* evaluations of organisational and safety cultures.

Comparing culture assessments with the quality of care:

H3a There will be a moderately strong correlation and reasonably good agreement between criterion-based measurements of the quality of care and both (1) questionnaire-based and (2) holistic evaluations of *organisational* climate/culture.

H3b There will be strong correlations and good agreement between criterion-based measurements of the quality of care and both (1) questionnaire-based and (2) holistic observation-based evaluations of *safety* culture.

## Study design

Using questionnaire- and observation-based methods, the study assessed organisational and safety climates/cultures in the DU and ED of eight hospitals (i.e. 16 research sites). Quality of care was assessed from retrospective audits of patients' notes focused on three commonly occurring conditions for each type of clinical setting (DU and ED). These data sets permitted the comparisons required by the study objectives. The details of data collection and analysis can be found in *Chapter 3* and the research protocol in *Appendix 2*.

Hospitals were purposively selected and researched sequentially, each hospital receiving feedback at the end of its data collection period. Sixteen independent research sites were needed to yield 90% power to test the hypotheses listed in the previous section, at the 1% level (see *Chapter 3, Comparison of data sites: threshold correlation, power calculations and clustering*).

Delivery units and emergency departments are high-stakes clinical settings, which were identified as priorities in a number of policy initiatives, such as the Reforming Emergency Care programme,<sup>39,40</sup> the National Service Framework for maternity services,<sup>41</sup> the NPSA Women's Services programme of work<sup>42,43</sup> and NPSA oversight of the Confidential Enquiries into Maternal and Perinatal Deaths. In each service, frontline staff must possess an extensive range of technical and non-technical skills and knowledge, applying these correctly and flexibly in a manner that supports choice. Effective interprofessional teamwork is essential, alongside effective collaboration with other parts of the hospital and community services for health and social care.

These clinical environments were expected to provide good opportunities for observing a range of safety-related processes and issues, including the management of triage; handovers within teams and between the multiple teams that attend these areas according to need; high turnover of patients/clients within each 24-hour period; high footfall through the department (including the multidisciplinary team and relatives); high variability in patient dependency, requiring ongoing

skill mix management; patient/client transfer (within the hospital, to other hospitals and to community services); referral (to primary and secondary care and other agencies); responding to staff shortages or suboptimal skill mix; and, of course, the clinical issues that arise throughout the hospital (e.g. medications safety, infection control and adherence to evidence-based guidelines).

Arguably, our chosen units are untypical of hospital services in general. Hall and colleagues<sup>44</sup> argue that EDs are likely to be the most unsafe of all hospital departments, with one-third of patient visits including a 'non-ideal care event', although these are usually not associated with harm. Woloshynowych and colleagues<sup>45</sup> (p. 9) list the unique features of EDs, and some are also true of DUs, particularly the lack of scope for staff to manage demand. Unmanaged demand may require care teams to be particularly skilled at negotiating flexibly in order to respond appropriately and promptly to high demand. Conversely, one could argue that team cohesion is less likely in such settings, because there is less opportunity than elsewhere for staff meetings and discussion, or team-building. However, this study was designed not to measure the safety culture of a range of hospitals, but to compare contrasting approaches to making such assessments. These busy, high-risk clinical settings provided opportunities to assess culture and compare culture assessments to markers of the quality of care.

No claim is made that the empirical results reflect hospital-wide cultures or quality. However, we would argue that the research processes used in this study are feasible across a very wide range of clinical settings, which was part of our intent in selecting DUs and EDs. A high proportion of patient safety studies are focused on work practices and the environment in operating theatres and other highly bounded situations. Although well-developed approaches to observing well-bounded settings or activities are likely to transfer readily to other well-bounded health-care settings (e.g. intensive care, pharmacies) or bounded activities (e.g. drug rounds and team meetings), they are less well suited to less bounded health-care settings and activities, that is, the majority of hospital work and many aspects of community-based services. The semistructured approach to observation-based evaluation of culture developed during this study would be feasible in both bounded and unbounded contexts, although it needs further refinement and testing in a wider range of contexts.

Wears and colleagues<sup>46</sup> identified 'the distributed nature of A&E clinical work' (p. 698) as a particular challenge for observational research in emergency care. Work in maternity units is also highly distributed across patients, specialised spaces and different professional teams as well as individual service providers. Conducting this study in DUs and EDs was ambitious and challenging, but an important springboard for shifting attention and the development of research methods towards approaches more suitable for less bounded environments and activities, and for environments that are characterised by distributed work.

Multisite research ethics approval was gained from Oxfordshire Research Ethics Committee C (REC reference: 07/H0606/87). This was a straightforward process with committee discussion, minor clarifications and the final approval occurring within a few weeks. Research governance approvals were obtained from each participating trust (eight; see *Table 3*). This was an extremely lengthy and time-consuming process that, at times, threatened the successful completion of the study (see *Chapter 4, Research governance*).

The next chapter reviews a selection of literature focused on safety culture and its measurement. *Chapter 3* describes data collection and analysis for each strand of the study. There are five results chapters in Part II of the report: *Chapter 4* describes process results that arose as the study progressed; *Chapter 5* describes the results from strand A, the staff survey, and tests

hypothesis H2a (*Hypotheses*); *Chapter 6* presents the results from strand B, the observation-based assessment of culture, and tests hypothesis H2b; *Chapter 7* describes the results from strand C, the retrospective audit of evidence-based markers of the quality of care; and, finally, *Chapter 8* compares the results from strands A–C, thus testing hypotheses H1a and b and H3a and b. Part III of the report contains a discussion of the findings and the conclusions of the study.



## Chapter 2

# Safety and organisational cultures and climates

### Introduction

Over the past 40 years, technological catastrophes including Three Mile Island, Chernobyl and the *Challenger* accidents have created growing interest in concepts of safety<sup>47,48</sup> and led to the emergence of safety culture as ‘an explanation for accidents’ and ‘a recipe for improvement in complex sociotechnical systems’<sup>49</sup> (p. 341). Within health care, improving patient safety has become an important aspect of quality improvement. International reports detail that between 2.9% and 16.6% of patients admitted to a hospital suffer some form of unintentional harm.<sup>50–52</sup> Many of these adverse events may be avoidable. Patient safety improvement programmes have called for cultural change to be tackled alongside structural reorganisation and systems reform in order to bring about a culture in which excellence can flourish.<sup>53</sup>

### Nature of the problem

Concerns regarding patient safety within health care point to a number of key factors that make a clinical service less safe: pressure on resources, unwillingness to admit fallibility and difficulties in reporting concerns across professional boundaries and organisational hierarchies.<sup>10,54,55</sup> Taking a lead from other safety-critical environments such as aviation, causes of adverse incidents are perceived to go beyond individual clinical failures, extending to systemic factors such as inadequate training and poor communication, equipment design, management systems and work processes.<sup>10,56,57</sup> Together such factors create not only behaviour patterns, but also organisational cultures, i.e. shared beliefs, norms and values that underpin and reinforce behaviours.<sup>19,55,58,59</sup>

### Measuring culture

Underpinning these understandings of safety culture are those approaches that regard culture as something an organisation *is*, i.e. that which is elusive, emergent and indeterminate.<sup>49</sup> Contrasting views suggest culture is something an organisation *has*, i.e. aspects that can be isolated, described and manipulated.<sup>60</sup> These distinctions lead to different assumptions regarding how much an organisation’s culture is controllable. Within safety scholarship, both approaches acknowledge the value of describing and evaluating cultural characteristics. However, viewing culture as something an organisation *has* is more likely to sustain an interest in *measuring* culture (or at least facets of culture), as this can support the diagnosis of excellence and problems, and both guide and monitor change efforts. Research on NHS reforms embodies this view,<sup>18</sup> as does the burgeoning activity around developing measurement instruments (see *Structured questionnaire research instruments*). Davies and colleagues<sup>18</sup> caution against going ‘too far down this road’, suggesting (p. 112) a view of organisational culture as:

an emergent property of that organisation’s constituent parts – that is, the culture may emerge somewhat unpredictably from the organisation’s constituents (making it not

necessarily controllable), but nonetheless characteristics of that culture may be described and assessed in terms of their functionality vis-à-vis the organisation's goals.

The tender specification (see *Appendix 1*) placed the requirements for this study firmly in a view of culture as something an organisation *has*, stating:

There is increasing interest in the idea that one way of improving healthcare performance factors such as quality, efficiency and patient safety could be through influencing professional and organizational culture ... In order to test the validity of culture as a marker for quality/safety, it is necessary to have access to reliable measurements of both quality and culture.

This study focused on comparing different approaches to evaluating culture with each other and with criterion-based assessment of the quality of care. Nevertheless, our underlying conception of culture is close to that of Davies and colleagues<sup>18</sup> (see above).

Because culture is a fusion of values, attitudes, perceptions, competencies and behaviour, it is difficult to measure, although many attempts have been made in health-care organisations.<sup>61</sup> A commonly used method is the survey, using questionnaires that elicit individual perceptions of the organisational and safety culture, which are then aggregated to indicate group perceptions. A more invasive alternative is temporary immersion in the work environment: either prolonged engagement as in traditional ethnography<sup>62</sup> or shorter strategic immersion, such as clinical governance review and inspection<sup>32</sup> and highly focused assessments of the efficacy of quality improvement interventions.<sup>33</sup> There is a need to compare these methods, which this study begins to address.

## Organisational and safety cultures: concepts and definitions

Safety culture is often framed as a dimension or subset of organisational culture. Although there is little agreement on a precise definition of organisational culture, numerous components or attributes can be said to contribute to an organisation's character and norms: dress, language, behaviour, beliefs, values, assumptions, symbols of status and authority, myths, ceremonies and rituals, and modes of deference and subversion.<sup>63</sup> As a subset of organisational culture, safety culture comprises these attributes as they relate to patient safety. Facets of a positive safety culture include:

- norms and rules for handling hazards, attitudes towards safety; reflexivity on safety practice<sup>64</sup>
- positive attitudes to safety behaviours and role modelling of such behaviours to peers and juniors<sup>65</sup>
- recognition of the inevitability of error and proactively seeking to identify latent threats<sup>55,66</sup>
- non-punitive reporting systems, analysis of errors and near-misses, feedback to frontline staff, sharing learning<sup>3,57,67</sup>
- openness, fairness and accountability at all organisational levels<sup>68</sup>
- well-functioning teams<sup>11,12,55,69-73</sup>
- maintenance of situational awareness among team members<sup>34,74,75</sup>
- non-hierarchical teams in which roles are flexible<sup>76,77</sup> and staff at all levels feel empowered<sup>78</sup> and
- attention to staff development,<sup>67,79</sup> with consequent reduction in staff stress and burnout.<sup>80,81</sup>

Wide variation in the framing of organisational and safety cultures and climates makes it inevitable that there will be little agreement on how they should be observed or measured. The

most common method is self-evaluation by staff. Self-evaluations need to be structured, and one useful framework is the Manchester Patient Safety Framework (MaPSaF),<sup>82,83</sup> which has been adapted for a range of clinical settings.<sup>84</sup> More commonly a structured questionnaire is used. The next section provides examples of the wide range of structured questionnaire assessment tools. In research studies the main alternative to structured self-evaluations by staff is observation and in *Ethnographic methods* we will provide examples of observation-based studies that have drawn from the ethnographic tradition. In *Strategic immersion in the environment* we summarise a different approach to observation-based assessments of health care, which we have termed strategic immersion.

### Structured questionnaire research instruments

Structured questionnaire research instruments typically adopt a typological or a dimensional approach, but vary in terms of their theoretical or conceptual underpinnings, their scope and their depth.<sup>63</sup> Several organisational climate and culture measures have been applied in health-care settings, for example:

- Harrison's<sup>85</sup> Organizational Ideology Questionnaire assesses the 'ideology' of the organisation and its relationship to the interests of its members and to the external environment. The tool has been used in the UK to examine the effect of NHS reforms and the monitoring of culture over time.<sup>86</sup>
- The Competing Values Framework (CVF)<sup>87-89</sup> (originated and is better known outside health care) uses a typology of four dominant culture types based on core values that characterise an organisation, arranged on two axes: internal-external focus and flexibility/individuality-stability/control. It has a strong theoretical basis and has been used in UK health care.<sup>16,90</sup>
- The Organizational Culture Inventory<sup>91</sup> assesses 12 sets of normative beliefs related to 12 different cultural styles; these reduce to three general types of culture: constructive, passive-defensive and aggressive-defensive. The tool has been widely used including in health care<sup>92,93</sup> but is too lengthy to be completed by busy clinical staff.
- The Quality Improvement Implementation Survey<sup>94</sup> (QIIS) uses four culture concepts associated with the CVF and adds an extra dimension ('rewards').

As interest in health care and health services research in the 1990s and early 2000s shifted towards patient safety, attention turned towards developing measures of those aspects of organisational culture and climate that related to safety. In the USA, several high-profile safety climate and culture measures were developed and validated for use in audit and research in health-care settings. Pronovost and Sexton<sup>95</sup> provided a useful review and guidance. Well-developed research instruments include:

- multiple versions of the SAQ, including the Teamwork and Safety Climate Survey<sup>25,96</sup> and the Institute for Health Care Improvement Safety Climate Survey,<sup>24,96,97</sup> all of which were developed and extensively tested by the University of Texas Center of Excellence for Patient Safety Research and Practice
- Nieva and Sorra's Hospital Survey on Patient Safety Culture for the United States Agency for Healthcare Research and Quality (AHRQ)<sup>55,98-100</sup>
- the Patient Safety Climate in Healthcare Organizations questionnaire (PSCHOQ) (Stanford University with the Palo Alto Veterans' Affairs Health Care System), also funded by AHRQ under its Systems Related Best Practices initiative,<sup>20</sup> and based on several existing surveys including the Operating Room Management Attitudes Questionnaire,<sup>101</sup> an earlier iteration of the SAQ; the PSCHOQ has been extensively used<sup>137,59,102-106</sup>
- Gershon and colleagues<sup>107</sup> Hospital Safety Climate Questionnaire which, in contrast with the more generic tools above, focuses on universal precautions for blood-borne pathogens; Turnberg and Daniell<sup>108</sup> adapted it for use in respiratory care.

Each of these instruments takes a dimensional approach to evaluating safety culture by aggregating individuals' levels of agreement with statements on Likert scales. The dimensions of safety culture that are explored include teamwork climate, safety climate, job satisfaction, stress recognition, perceptions of management, working conditions, absence of barriers to safe working practices, minimal conflict and good communication among staff, safety-related feedback, and many more.

In the UK, there are three notable strands of development of instruments drawing together organisational and safety climate:

- The MaPSaF is based on Westrum's theory of organisational safety (reprinted in Westrum, 2004<sup>109</sup>). It is a typological tool initially developed for use in primary care trusts (PCTs) to assess safety culture maturity. It is intended primarily for team-based reflection and development<sup>82</sup> and has been tailored for use in acute, ambulance and mental health settings.<sup>84</sup>
- The NPSA sponsored the development of organisational and safety culture instruments for use across the NHS.<sup>110</sup> Aston University has developed the TCAM (Team Climate Assessment Measure).<sup>111</sup> The dimensions addressed include task reflexivity, team stability, leadership, participative trust and safety, open exchange and interprofessional exchange.
- The Healthcare Commission (subsequently Care Quality Commission) National Staff survey was derived from a model developed by Aston University that links work context (including organisational climate) to the management of people, psychological consequences for staff, staff behaviour and experiences, and errors and near-misses that ultimately affect patient care.<sup>112</sup>

Elsewhere in Europe, Silva and colleagues<sup>38</sup> developed an organisational and safety climate inventory, based on the CVE,<sup>87</sup> while, during the course of this study, the development of other instruments has begun.<sup>113–117</sup>

### **Ethnographic methods**

Because culture is a fusion of values, attitudes, perceptions, competencies and behaviours, it is very difficult to measure. However, evaluations of culture can be obtained from temporary immersion in the environment: either prolonged engagement, as in traditional ethnography,<sup>62</sup> or shorter strategic immersion, as employed in clinical governance review and inspection.<sup>32</sup> Ethnographies provide rich, multilevel understandings of researched environments, but they are time-consuming, and comparison of multiple environments is difficult. A wide range of safety culture studies have adopted the ethnographic approach.<sup>34,45,118–125</sup>

### **Strategic immersion in the environment**

A major advantage of direct observation, in contrast with the questionnaire response, is that it enables researchers to see what people *do* and *say* rather than just what they *say they do*.<sup>126</sup> It can uncover how complex jobs are routinised together with 'the tacit skills, the decision rules, the complexities and the discretion' utilised in routine and marginal work (Smith,<sup>127</sup> p. 221). Although traditional ethnography requires lengthy immersion in the researched environment by participant observers, accomplishing high-quality observational research during relatively brief periods of immersion is potentially achievable providing researchers restrict their studies to a topic or 'lens' through which to view the group they are studying. Willis<sup>128</sup> (p. 557) refers to 'focused (limited gaze) ethnography' and 'rapid (quick time) ethnography', where speed is a virtue and a necessity. Strategic immersion represents a more structured approach to observation and supplementary data collection, to provide defined coverage at reasonable expense and to permit comparison between different organisations. For example, studies of nursing care have used checklists of features of good practice to structure strategic immersion,<sup>129–131</sup> whereas



observations to develop or evaluate quality improvement interventions are structured by the focus of the intervention itself.<sup>33,118</sup>

Strategic immersion is unsuitable for settings or issues for which there is inadequate earlier work to define a framework for observation. However, previous work in the field of organisational and safety culture renders evaluation through strategic immersion possible in this study. In particular, the observation strand of this study benefited from an earlier 3-year ethnographic study in delivery units, which three of the study team completed,<sup>34</sup> and which linked experience of setting ethnographic observations alongside highly structured observations.<sup>132</sup>

During strategic immersion to observe health-care teams at work there is limited or no observer participation in the observed environment, as the limited time period does not allow the observer time to negotiate an active role in the team being observed. The method may therefore be described as primarily that of 'passive participation'.<sup>133</sup> However, passive observers may ask brief questions for instruction and clarification, although in health-care settings this may not always be possible when staff are busy. The scope for immersion may be limited by the difficulties associated with the qualifications required to enter into many of the roles of medical work. Smith<sup>127</sup> (p. 227) argues that diversity of observational approaches results not from 'methodological sloppiness', but from 'real constraints governing the conditions under which researchers can and cannot conduct qualitative field research'.

Hammersley<sup>134</sup> (p. 26), discussing observation, argues 'there are always multiple, non-contradictory, true accounts possible of any scene and we need to know the basis for particular selections'. Actions of individuals are observable, but the meaning attached to the physical actions is divorced from the actions themselves. Researchers may apply their own meanings to these actions, which may be different to the meanings of the actors.<sup>135</sup> The degree of inference and, therefore, potential for error is likely to be greater when observing some features of health-care practice than others.

## Safety culture and climate

Whereas culture is generally taken to indicate the collection of values, beliefs and assumptions that guide behaviours<sup>19</sup> and that are shared by group members,<sup>35</sup> climate refers to the aggregate of individual perceptions of practices, policies, procedures and routines about safety in an organisation.<sup>14,136</sup> Being more concerned with health-care practitioners' conscious perceptions and attitudes, safety climate is thus easier to measure than more deep-seated and less overt beliefs and values,<sup>20</sup> and it is often argued that staff surveys capture 'safety climate' rather than 'safety culture'.<sup>25</sup> In this study we have treated the staff survey (see *Chapter 3, Strand A: staff survey*) as reflecting some important aspects of safety climate (aggregated espoused attitudes and values, aggregate perceptions of situated work practices) and observation-based evaluations (see *Chapter 3, Strand B: observation-based holistic evaluation of safety culture*) as reflecting some important aspects of safety culture (situated practices and artefacts supporting the enactment of shared norms and meanings).

The theoretical distinction between culture and climate is not always upheld in practice: the meanings of each are contested<sup>137,138</sup> and confused. For example, some authors use the terms interchangeably<sup>58,139</sup> while one review of safety climate measures<sup>140</sup> includes four measures that use the term 'culture'. In this report, we follow Sexton and colleagues<sup>25</sup> by acknowledging the terminological confusion while using safety climate (measured by questionnaire) to designate the surface indications of an underlying safety culture.

## Selecting organisational levels for research into safety culture and climate

Although the literature typically refers to organisational culture, there is nevertheless substantial evidence that, at any rate in relation to safety culture, the department or ward may be as important. Mohr and Batalden<sup>141</sup> argue the need to examine patient safety at the level of the 'clinical microsystem' (a small organised group of staff caring for a defined population of patients). Zohar and Luria<sup>142</sup> found meaningful variations in safety climate scores within as well as between organisations, and Smits and colleagues<sup>143</sup> found that responses to a patient safety survey clustered more strongly at unit than at hospital level. Huang and colleagues<sup>137</sup> found that perceptions of safety culture varied even across similar units within the same hospital (in this case, intensive care units), and Gaba and colleagues<sup>139</sup> commenting on that study emphasised that research needs to be carried out at both clinical unit and organisational levels. However, unit-level research brings its own problems: the boundaries of the unit are porous when units are 'highly coupled', so that decisions in one constrain choices in another (e.g. ED cannot move a fracture patient until an orthopaedic ward is willing to admit).<sup>144</sup>

In this study, for example, our staff survey was designed to distinguish between the wider organisational climate and the climate in the immediate clinical department by, on the one hand, asking questions about trust-level communication, error wisdom and support for continuing professional development (CPD) and, on the other hand, asking questions about respondents' individual role contexts, line management and teamwork and safety climates in the immediate clinical department (further details in *Chapter 3*). However, some staff perceptions may arise from a fusion of organisational-level and department-level factors; many health-care professionals work in more than one team or department, and departments are closely coupled to other departments and services. We will return to this discussion in *Chapter 4, Close-coupling of departments and services*.

## Quality: case note review

The use of clinical process measures to monitor the quality of health care has many advantages,<sup>145</sup> although the concept of quality itself is multidimensional and contested. Case note review (comparing recorded activity with evidence-based standards of care) is often used in quality research, for example to detect adverse events<sup>146</sup> or to assess compliance with predetermined evidence-based standards.<sup>147-149</sup>

The methodology of case record review has considerable strengths. It has been noted to provide a more complete indication of the incidence of adverse events or critical incidents than reporting systems.<sup>150</sup> Case review forms provide a standardised method of recording and data collection which is robust when used on a random sample of case records.<sup>150</sup> The epidemiological data obtained are potentially useful for comparative studies, although any comparisons need to take account of variations in methodology, particularly with the definition and inclusion criteria.

Criterion-based review has been proposed as an effective alternative to semistructured holistic review methods that rely on professional judgement to determine standards of care and quality, which have been associated with concerns about inter-rater reliability,<sup>151</sup> consistency,<sup>152</sup> hindsight bias<sup>153</sup> and reviewer idiosyncrasy.<sup>154</sup> Clinical audit in the UK has adopted criterion-based review to identify substantial variations in organisation and clinical care between hospitals.<sup>155</sup> However, adoption of the criterion-based approach may fail to identify the nuances of health-care variation.<sup>156</sup> Lilford *et al.*<sup>157</sup> note that, although certain clear-cut violations, such as failure to

check a patient's blood potassium level when indicated, can be detected from case note review, other factors, such as the quality of communication with patients or surgical skill, remain largely in the tacit domain. Hospitals also tend to vary considerably in adherence to quality standards.<sup>158</sup> A particular hospital may not, for example, do well in getting patients with hip fractures to surgery within 24 hours, but it may have high adherence to drug therapy after a heart attack.<sup>157</sup> Process-based monitoring is subject to potential bias as the opportunity for error varies by case mix:<sup>159</sup> sicker patients need more care, which gives greater opportunity for errors of commission and omission.<sup>160</sup>

Case note audit is underpinned by 'the assumption that good quality recording reflects good quality care'<sup>161</sup> (p. 134) or, at any rate, that good-quality recording is an important and tangible aspect of good quality care. Zegers and colleagues<sup>162</sup> distinguished between poor-quality information within case notes, which was linked to an increase in adverse events, and missing information, which underestimated adverse events. Case note audit is a challenging method,<sup>163</sup> as well as being limited: only a proportion of the whole narrative of care is recorded,<sup>164</sup> and the unmeasured and unaudited may constitute an important aspect of care quality.<sup>165,166</sup> It is wholly dependent on the accuracy, completeness and legibility of patient records. Incomplete records do not necessarily provide evidence that an event did not occur.<sup>149</sup> Hindsight bias is an additional challenge.<sup>157</sup> Documentation provides only a 'partial' representation of events. Studies have demonstrated considerable discrepancy between 'objective' assessment of adverse events reported by patients and 'objective' accounts constructed via retrospective auditing of patient records. This can partly be explained by differences in professional and patient perceptions of errors, but may also be due to incomplete documentation in medical records.<sup>167,168</sup>

Criterion-based review focuses on adherence to or violation of agreed evidence- or logic-based standards, suggesting that what is written is a direct measure of performance. Although the record can be seen as a form of 'organisational memory',<sup>169</sup> it is not a *neutral* repository of information. It mediates medical work and there is selectivity in what ends up in the record.<sup>170</sup> What is written represents the production of hierarchical relations and socialisation processes that constitute medical work.<sup>170</sup> Scientific evidence is not clear, accepted and bounded.<sup>171</sup> The evidence base for particular health-care technologies and practices is often contested and continually redefined to fit the local context.<sup>172</sup> Documentation provides a reconstruction of events and may be subject to normative expectations to comply with evidence-based standards that may not be reflected by the reality of practice.

In this study, a retrospective review of case notes was used to evaluate recorded compliance with a spectrum of evidence-based markers of the quality of care. The methods used are described in *Chapter 3*, whereas challenges arising during the audit process are discussed in *Chapter 4, Strand C: criterion-based audit of clinical notes*.

## Comparing measurements of culture and evaluations of the quality of care

The various methods of measuring safety climate and culture outlined above have not been systematically compared, a gap that this study goes some way to filling. Attempts have been made to investigate the link between organisational culture and the quality of health care, although reviews<sup>16,35-36</sup> have found that the evidence available is sometimes contradictory and suggested links are contingent and complex. However, there is some evidence of a link between safety climate/culture and health-care performance, although it is not known whether a causal relationship exists, and the direction of any causality would need to be established.<sup>37,38,173</sup>

## Conclusion

Aspects of culture such as poor communication and a climate that discourages speaking up to ask questions or alert others to potential problems have been linked to accidents, suboptimal processes and outcomes and avoidable harm in a wide range of contexts, including health care. Logically, this has spurred interest in assessing and, where necessary, improving certain aspects of culture with the aim of preventing avoidable harm and using resources (human, physical and financial) as productively as possible. However, culture is difficult to measure because it is a fusion of values, attitudes, perceptions, competencies and behaviour. Nevertheless, a range of questionnaires and frameworks have been developed to evaluate aspects of culture at different levels within organisations and ethnographies have provided complementary insights. Less immersive observation-based studies have also captured targeted facets of culture or the quality of care in the observed health-care environment.

In parallel, there has been greater attention to monitoring the quality of care to identify areas of success and areas for improvement. Monitoring usually focuses on key markers which are important for safety, the patient experience or for good management of limited resources. Few studies have compared quantitative or qualitative evaluations of culture with assessments of the safety and quality of care. The evidence to date is equivocal. This study provides an additional set of comparisons for the slowly growing evidence base and, further, examines the processes and challenges of obtaining estimates of culture or climate and the quality of care.

No prior studies were found which compared quantitative and qualitative assessments of culture. Although some would argue ‘Why would one wish to compare apples with pears?’, this study was charged with responsibility for examining agreement between very different approaches to assessing culture. Quantitative data about facets of organisational and safety climates were collected using a questionnaire distributed to staff in 16 clinical departments. Semistructured observations were made by researchers who made multiple visits to these departments. To enable the required comparisons, the observations had to be banded, i.e. made quantitative. The process is described in *Chapter 3*.

## Chapter 3

# Methods for data collection and analysis

### Comparison of data sets: threshold correlation, power calculations and clustering

Silva and colleagues<sup>38</sup> obtained a strong correlation ( $r=0.72$ ) between organisational and safety climate, and showed organisational climate and safety climate to be inversely correlated with the incidence of accidents ( $-0.955 \leq \rho \leq -0.865$ ). Setting a similar threshold correlation level for this study is reasonable. It acknowledges that not only *can* these measures be highly correlated but indeed they *should* be if safety climate measures are meaningful. In the power calculations for this study the threshold correlation was set at 0.7. For independent research sites, 80% power at 5% significance requires 12 pairs of measurements<sup>174</sup> and 90% power at 1% significance requires 16 pairs of measurements.

Sixteen research sites were recruited to this study but, for efficiency, the study was carried out in DUs and EDs at eight NHS hospitals in England so a degree of clustering was present at the organisational level, although these clinical departments have virtually no interaction at team level. When the study began there was no prior work from which intraclass correlation coefficients (ICCs) could be estimated to allow for clustering in the power calculations. However, multilevel modelling has been used to allow for clustering in the data (see *Multilevel modelling*).

### Selecting the sample of research sites

The reasons for selecting DUs and EDs were outlined in *Chapter 1, Rational for focusing on delivery units and emergency departments*. For efficiency, the study was restricted to hospitals that have both types of clinical provision. Hospitals in England were purposively selected using a number of criteria:

- *Geographical spread*. The research sites were situated in 6 out of the 10 strategic health authorities in England (East Midlands, East of England, London, South West, West Midlands, and Yorkshire and Humberside).
- *Type of NHS trust*. The selected research sites included general hospitals and major tertiary centres; single-site and split-site provision; and foundation and non-foundation trusts.
- *Size of NHS trust and participating departments*. This was indicated by the approximate trust income and patient throughput statistics for the relevant clinical areas. The hospital reporting year 2008–9 was selected for the comparisons shown in *Chapter 4* (see *Table 3*). This was the midpoint of the study. Patient and birth statistics for England allowed us calculate the boundaries for quartiles and place each study site in the relevant quartile: Q1 (smallest 25% of departments), Q2, Q3 or Q4.

We also sought to include variation in relation to the populations served by the hospitals (rural and urban; ethnic diversity; mobility; indices of social deprivation; complexity of case mix; and the ease of recruiting and retaining a well-qualified workforce). The sample of research sites is described in *Chapter 4, Recruiting research sites*.

## Strand A: staff survey

This strand relates to objective 2: to use questionnaires to obtain quantitative assessments of the organisational and safety climate at each site.

### *Developing the organisational and safety climate questionnaire*

We wished to use prevalidated questions and scales within the organisational and safety climate questionnaire for staff because the aims of the study (see *Chapter 1, Aims and objectives*) centred on comparing different approaches to evaluating culture rather than developing and validating new research instruments. A review was conducted at the beginning of the study (2007) to establish the availability and psychometric properties of organisational climate and safety climate data collection instruments. The instruments identified were examined more closely by reading the associated websites and manuals. We also discussed (in person or by e-mail) the development, reliability and availability of a range of climate instruments with people engaged in their initial or continuing development. A small number of climate instruments developed in languages other than English were considered, but translation and revalidation were not feasible within our study funding and timetable.

The final selection of instruments from which to draw questions to evaluate organisational climate and safety climates used the following criteria:

- availability of material describing the development and testing of the instrument, scores from psychometric testing and benchmarking data
- stability and reliability
- extensive prior use in England and
- cost (no fees were levied for use of the scales and questions selected for this research study).

The staff survey questionnaire (see *Appendix 3*) was developed by combining groups of questions from the NHS national staff survey<sup>175</sup> and the Teamwork and Safety Climate Survey.<sup>25</sup> We received permission to use questions from the NHS national staff survey in any month except October. This was to minimise any possible confusion with the annual NHS staff survey, which is distributed each October. The data collection timetable for this study was adjusted to satisfy this restriction.

The study questionnaire contained four sections of questions from the NHS national staff survey which were intended to capture facets of organisational climate, namely:

- Q6 a six-item scale evaluating perceptions of the organisation, which in subsequent analyses we have termed 'organisation'
- Q7 two questions about whistle-blowing
- Q8 a question about reporting errors, near-misses and incidents
- Q9 a seven-item scale concerning organisational responses to errors, near-misses and incidents, which we have termed 'error wisdom' in subsequent analyses.

At research sites 1 and 2, the staff survey contained additional questions relating to other aspects of organisational culture, but response rates from these sites that were lower than we had hoped for prompted shortening of the questionnaire (see *Sampling, delivery and maximising returns*).

A further five sections of the questionnaire were intended to capture perceptions of safety climate and team factors or local work environment factors that are thought to be related to safety climate (see *Chapter 2*), namely:

- Q2a–c a three-item scale, which we have labelled ‘overload’
- Q2d a five-point Likert-scale item called ‘Relationships at work are strained’
- Q3 a five-item scale evaluating supportive behaviours from the respondent’s immediate manager, which in subsequent analyses we have termed ‘line management’
- Q4 one question each concerning working closely with other team members and meeting regularly to discuss effectiveness and improvement
- Q10 teamwork climate scale (six items)
- Q11 safety climate scale (seven items).

Questions 2–9 were reproduced from the NHS national staff survey,<sup>175</sup> whereas questions 10 and 11 were Teamwork Climate and Safety Climate scales.<sup>25</sup> A minimal change to question 10 was required to reflect professional titles in study sites: in the questionnaires distributed to DU staff the word ‘nurse’ was replaced with ‘midwife’.

One question (Q5, *Appendix 3*) asked about trust-provided or trust-supported training, learning or development during the past year. This was included because CPD of individuals and teams is argued to be important for developing and maintaining positive organisational and safety cultures<sup>67</sup> and to maintain job satisfaction and motivation.<sup>81</sup> Information on common forms of CPD was elicited in the form of closed questions (see *Table 9*) and space provided for additional responses. The listed categories of CPD overlapped, but all were included because previous research has shown that people find it surprisingly hard to call to mind examples of workplace or work-related learning.<sup>176</sup> Providing a variety of prompts, including the most popular approaches to CPD, was intended to help respondents recall CPD and to encourage them to recognise learning within a wide variety of activities.

Demographic and role-related questions were also included to permit analysis of whether these factors mediated responses to the organisational climate and safety climate questions and to address objective 8 (see *Chapter 1, Aims and objectives*). This study used response categories from the NHS national staff survey to maximise opportunities for comparison between the study results and trust-wide or national results from the same survey questions.

Piloting with clinicians associated with the university where most members of the study team were based established the usability and face validity of the questionnaire for clinicians from different professions working in DUs or EDs, and that it took < 30 minutes to complete.

### **Sampling, delivery and maximising returns**

Clinical leads at the research sites were asked to identify staff who should be invited to complete the staff survey. Thus, midwives, nurses, doctors, support workers, administrators and managers (and, where appropriate, allied health professionals) were invited to participate in the survey, the exact selection being determined by team leaders’ definitions of team membership. However, clinical leads were asked to exclude certain groups whom we felt they would have insufficient consistent experience of the culture of the particular clinical area: students and members of staff who had joined the department < 4 weeks before the survey was distributed, and in particular Foundation Programme doctors (doctors in training in their first 2 years after graduation from medical school). Comparison of the national rotation schedules for these junior doctors and the data collection timetable for the sequentially researched study sites revealed that these junior doctors would be eligible for inclusion at some sites but not at others due to very recent rotations: we therefore asked that they be excluded from staff lists at all sites. It is a moot point whether these exclusions were necessary and advisable, as newcomers may evaluate the culture of their work environment very quickly and may be better able than more established members of clinical teams to identify safety concerns. Further research into the differing perspectives of newcomers and established members of staff may be useful.

Seeking a locally defined definition of the ED or DU team was part of the commitment stated in objective 1 to work *with* staff in participating trusts, respecting their organisational and professional knowledge. The definition of the departmental team was thus conceptually coherent for senior clinicians at each research site but differed between research sites. We accept that our design in this respect is vulnerable to the criticism that data between sites were not directly comparable. However, health-care teams do vary between departments and hospitals, according to local need, resources and preferences; so strict comparability between staff survey samples from multiple research sites would be an unrealistic ideal for any study. In this study it was deemed more important to examine how team compositions and perceptions of the team varied. For example, some sites had dedicated allied health professionals, administrative and clerical staff, porters and cleaners, and one (an ED at two geographically separated locations) had dedicated ambulance staff; others had few or none of these. The use of agency and/or locum staff for long periods (months or even years) varied substantially between sites: these may be key team members but do not appear on trust staff lists. We would suggest that the variability we have documented is a feature of all multisite studies of teams, but rarely made explicit.

Hand delivery and collection of questionnaires at 16 research sites was not feasible for research fellows, and preliminary work with research sites established that local clerical or clinical staff would not be able to undertake this work, so a postal survey was conducted. Questionnaires were sent to all designated members of staff along with a covering letter and a project information sheet. A stamped, addressed envelope was provided for the return of the questionnaire. One reminder was sent to non-respondents after 2 weeks. A unique identifier, which could be removed by respondents if desired, was placed on each questionnaire to prevent unnecessary reminders. The key to the identifiers was stored separately from the returned questionnaires and destroyed when no longer required. Each distribution of the questionnaire required minor updating to reflect the specific clinical area, contact details for queries and requested return dates. The data collection timetable of sequentially researched sites involved a risk of late returns from one site being allocated to the next site. To avoid this, we alternated the named contact for questionnaire responses at the first 10 sites. This strategy was insufficient at research sites 11–16, where data collection overlapped to compensate for earlier delays in the data collection timetable. For these sites different coloured paper was used for the questionnaires to prevent misallocation of responses.

Because most staff surveyed were nurses or midwives, we anticipated that the response rate would be no higher than 50–55%,<sup>30</sup> even with the inclusion of elements that are known to support better response rates in the general population: sending from a university address, including a stamped return envelope, sending a reminder and offering an incentive for returning the questionnaire<sup>177</sup> (entry into a prize draw with the opportunity to win one of 10 gift vouchers for a popular high-street store to the value £10). Salience is an important factor influencing response rates from health-care practitioners and the general population.<sup>29,177</sup> Patient safety was expected to be a salient issue in DUs and EDs, since this study coincided with a peak of safety-focused policy and practice development initiatives. However, it is possible that, particularly by the time of data collection at the later research sites, staff may have begun to be jaded by multiple waves of safety-focused activity. Response rates were lower than hoped for at the first two research sites (20% and 27%; see *Table 5*), and two steps were taken with a view to improving response rates. First, the questionnaire was shortened,<sup>177</sup> largely by omitting questions about matters with a less immediate relationship to patient safety culture than those that remained (e.g. work–life balance; physical attacks on staff). In the final version in *Appendix 3*, 52 separate responses are requested, whereas the original questionnaire requested 107 separate responses. Secondly, changes were made to the use of the prize draw as an incentive for staff to complete and return the questionnaire promptly. Initially, the draw (for 10 gift vouchers) was made after 2 weeks but, to encourage responses to



the reminder letters, a draw for four of the prizes was held 2 weeks after the second mailing at site 3 onwards. The unique identifiers were used to trace prize winners.

Response rates were variable and remained lower than hoped for, although similar to comparable studies.<sup>100,178</sup> A final attempt to increase response rates was trialled for research sites 11–14. This will be discussed in *Chapter 4, Examining influences on response rates*.

## Analysis

### Cleaning the data file and gaining an overview of results

Staff survey questionnaire data were coded within a SPSS (SPSS Inc., Chicago, IL, USA) v. 17.0 data file, which was cleaned by exploring each variable using a variety of descriptive statistics and single variable analyses. This process allowed the researchers to become very familiar with the survey data and gain an intuitive sense of patterns within it. Because response rates varied widely between research sites (*Table 5*), site-specific response rates were added to the data file to make this fluctuating level available for subsequent analyses.

### Missing data

Each analysis in *Chapter 5* included all cases for which responses to the questions under analysis were present. Percentages have been calculated after removing cases with missing data from the denominator. Sections within *Chapter 5* indicate the number of missing cases for the analyses in that section.

### Testing for normality

Throughout the data set continuous variables were examined for normality using the Kolmogorov–Smirnov test, comparison of the mean and median and inspection of histograms and P–P plots (probability–probability plots, which are used to check how well two distributions agree).

### Descriptive statistics and correlations

The descriptive statistics and correlations reported in *Chapter 5* were calculated using SPSS v. 17.0. Exact *p*-values were calculated using an online Statistical Toolbox.<sup>179</sup> We found no definitive reference for the interpretation of correlation coefficients but consulted a panel of statisticians with experience of health services research. In this report correlation coefficients will be interpreted as follows:

- < 0.1, negligible
- 0.1 to < 0.4, low
- 0.4 to < 0.7, moderate
- 0.7 and above, high.

### Calculating scores on prevalidated scales

In line with instructions in the associated manuals, mean scores were calculated for the items that formed prevalidated scales. Where required, scores from negatively worded items (questions 2a–c, 9c, 10b and 11g) were reversed before calculating the mean scores. All scores contributing to the overload scale were reversed to achieve consistency across all scales such that high scores represent positive outcomes.

### Multilevel modelling

As the 16 research sites were clustered within eight hospitals, multilevel modelling supported by MLwiN 2.9 software (Centre for Multilevel Modelling, Bristol, UK) was used to analyse the survey data. First, influences on site-specific response rates were investigated by fitting a two-level

model (hospital and research site) and adding the three site-level characteristics that were known in this study: the type of service (ED or DU), the size of the hospital department represented by the quartile recorded (see *Table 3*) and the number of people to whom questionnaires were sent (this was centred on the mean number of questionnaires distributed to ease interpretation of the parameter estimate) to indicate how inclusively the definition of team was drawn. An alternative indicator of inclusiveness could have been the number of professional groups included in the local definition of team. These characteristics were included in the model as fixed functions. The results can be found in *Chapter 4, Examining influences on response rates*.

Secondly, influences on scores for each of the prevalidated scales reported in *Chapter 5, Organisational climate and Safety climate and team factors*, were investigated using a three-level multilevel model of individual responses nested within research sites nested within hospitals. Site-level and individual-level demographic and role-related variables were added one by one to investigate whether any had a significant coefficient and provided a better-fitting model, as indicated by the reduction in the value of  $-2 \times \log\text{-likelihood}$  with degrees of freedom (df) equal to the number of new parameters entering the model.<sup>180</sup> This included variables such as service (ED or DU), survey response rate, number of questionnaires distributed, gender, age, ethnicity, profession, hours worked, CPD profile over the past 12 months, years worked for the trust and whether the current role includes managing staff.

### Combining indicators of organisational climate

To facilitate the study's central aim of comparing different approaches to evaluating safety culture (see *Chapter 1, Aims and objectives*), it was necessary to create a summary survey-based measure of organisational climate to set alongside summary measures from *Chapter 6* (observations) and *Chapter 7* (audit) in the comparisons made in *Chapter 8*. A weighted average was used: the responses to the well-established organisation and error wisdom scales (see *Appendix 3*, Q6 and Q9) were weighted equally (one-third), whereas the individual questions (Q7 and Q8) were weighted one-ninth each to form the other third of the summary organisational climate score. To avoid the weighted average being distorted by combining values from different normal distributions, before averaging each component was scaled to fit the standard normal distribution [*z*-scores, centred on mean 0 and with standard deviation (SD) 1]. Before this, the responses to the questions 7 and 8 had to be converted from categorical variables (yes, no, don't know) to continuous variables. The conversion was made by calculating the proportion of respondents at each site answering 'yes' to each question (see table in *Appendix 4*). The three newly created variables were inspected and found to be normally distributed, paving the way for conversion to a common scale and the calculation of a weighted average.

### Combining indicators of safety climate

To facilitate the comparison among strands A–C, a summary score for safety climate was created. This combined questionnaire elements relating to the immediate work context that are thought to contribute to the local safety culture (see *Chapter 2*). As described in *Developing the organisational and safety climate questionnaire*, the questions included were prevalidated scales from the NHS annual staff survey,<sup>175</sup> which we termed 'overload' and 'line management'; three individual questions from the same source (questions 2d, 4a and b; see *Appendix 3*); and Sexton and colleagues' teamwork and safety climate scales.<sup>25</sup> Mirroring the procedure described above, the four well-established scales were weighted equally (one-fifth each) and the individual questions were accorded less weight (collectively constituting one-fifth). As noted in *Chapter 5, Safety climate and team factors*, 90% responded positively to question 4a, 'Do you have to work closely with other team members to achieve the team's objectives?', rendering this question less able to differentiate between sites than the other two stand-alone questions, 2d and 4b. Question 4a was, therefore, allocated half of the weight of questions 2d and 4b. Thus the indicators were combined as shown in equation 1 so that, as required, the weights total 1. As with the calculation

of the summary survey-based organisational climate score, before combining each element of the survey-based safety climate score was scaled to fit the standard normal distribution and so z-scores are shown in the equation and subsequent results.

$$\begin{aligned} \text{Summary survey-based safety climate} = & 0.2z_{\text{overload}} + 0.2z_{\text{line management}} + 0.2z_{\text{Sexton teamwork climate}} + \\ & 0.2z_{\text{Sexton safety climate}} + 0.08z_{\text{relationships at work are strained}} + \\ & 0.04z_{\text{proportion responding yes to Q4a}} + \\ & 0.08z_{\text{proportion responding yes to question 4b}} \end{aligned} \quad [\text{Equation 1}]$$

## Strand B: observation-based holistic evaluation of safety culture

This strand relates to objective 3 (see *Chapter 1, Aims and objectives*): to generate quantified holistic evaluations of organisational and safety culture for each site using observation. There was no well-established method or prevalidated instrument available for this strand. The closest fitting well-developed approach was MaPSaF,<sup>84</sup> but pilot work with this framework convinced us that it would not be suitable for this study and, furthermore, to use it in this study would require an inappropriate subversion of its intended usage as a developmental tool based on self-assessment and reflexivity. A specifically tailored approach had to be developed as part of this study.

The holistic evaluations were derived from strategic immersion, primarily direct observation of staff at work and their working environment, supplemented by brief conversations with key informants. This method was derived from previous work by team members in delivery units,<sup>34,132</sup> and its application to EDs was piloted at a local hospital not included in the study.

Use of the word ‘holistic’ derives from its use in the commissioning brief (see *Appendix 1*): because our method was primarily observational, we generally use that term instead in the text, though we continue to use the term ‘holistic’ when referring directly to the study’s objectives, so as to preserve the structure of the report in line with the commissioners’ specification and the research protocol.

### Data collection

Six 1-day visits were undertaken at each hospital by at least two observers, which allowed sustained non-participant observation to see how staff behaved over a number of hours, and sampled sufficient time points to gauge the range of work and activity levels. In recognition of fluctuating workloads and staffing levels, the visits were distributed throughout the week. Data were, therefore, collected at all periods of the day excluding midnight to 6 AM, and on all days except Sunday.

Researchers recorded their observations in hand-written field notes. These were made during observation periods, then reviewed and annotated within 24 hours. Field notes contained a minimum of identifying information, for example labelling participants in any documented interactions by their profession and level or role rather than by name or initials. No identifiable patient information was recorded, although patients were anonymously included in notes if their interactions with staff were of relevance. See *Developing an observation prompt list* for description of developing prompts to guide data collection.

Observers positioned themselves near the functional centre of the department, either the whiteboard where bed occupancy and the progress of care are recorded or the desk or counter where computer monitors were sited (these are often, but not always, colocated). The distributed nature of ED work<sup>46</sup> (*Rationale for focusing on delivery units and emergency units*) necessitated the selection of one functional centre within EDs: 'major injuries'. Observers interacted with staff (provided this did not mean interrupting their work) if observers wanted information to help interpret what they observed or staff wanted to talk to us about our research or their work. Information sheets were provided for use as leaflets or posters or both, and we offered to attend briefing sessions for staff, as well as pre-meetings with local clinical leads. All of these information sources pointed out that staff could ask observers to leave the vicinity, or themselves move more than 3 metres away, if they wished what they were saying or doing to be excluded from our data. But observers noted events and conversations that were clearly visible or audible to anyone, including patients and visitors, in the vicinity, regardless of distance.

Observation data were supplemented by an orientation meeting with a senior nurse or midwife at each site. This enabled observers to contextualise and interpret what they observed, and included questions about safety-related matters that we were unlikely to be able to answer from observation alone: staffing establishment, maintenance of equipment, and updating of protocols. The timing of such meetings varied, depending on the availability of staff; some took place before observations, some afterwards. We believe that such variation did not significantly affect our observations: when orientation interviews occurred later rather than earlier, we tried wherever possible to ask staff during observations about factors that seemed directly relevant to what we were observing; when the orientation meeting took place earlier rather than later, we sometimes contacted the senior nurse or midwife concerned at a later date by telephone or e-mail to ask questions arising later. Observers met between observation sessions to debrief and exchange information and observations: these meetings served also to ensure that each observer was keeping all the items on the topic guide in mind.

Within EDs observation was planned within major injuries and minor injuries sections but not in the resuscitation and paediatric sections. This decision was mainly guided by the additional ethical concerns raised by observing in the latter areas and a strong desire among the researchers not to risk distracting staff in the normal conduct of their work, which seemed more likely in these areas.

At some sites the project research fellows were joined by a service user or an experienced service provider who did not work at the research site (see *Chapter 6, Data collection summary*). The purpose of this was to ensure that the perspectives and interpretations of the research fellows were supplemented and if necessary challenged by clinical and lay viewpoints. Service provider coresearchers were able to offer contextual and explanatory information, whereas the role of the service user coresearcher was to ensure that the research fellows were not too closely aligned with clinical perspectives at the cost of critical distance. These coresearchers received training from one of the project research fellows prior to visiting any of the research sites. They observed from the same vantage points as the project research fellows, and using the same prompt list, though not at the same time. They also made hand-written field notes, which were subsequently passed to the research fellows to be included in data analysis. Each coresearcher met with the research fellows to discuss and elaborate upon field notes and subsequent reflections.

Each site was offered early feedback on data collected in the three strands, A–C. Where this opportunity was taken up, discussion of the observation occasionally corrected minor misunderstandings or added new material. Field notes made during and immediately after the feedback meetings also became part of the strand B data set.

### Developing the observation prompt list

Mindful that the commissioning brief and study protocol (see *Appendices 1* and *2*) required quantified observations, the research team initially planned to develop a checklist that would be scored in situ. A provisional detailed checklist was developed to pilot interview and observation data collection focused on indicators of organisational and safety culture drawn from our previous work<sup>34</sup> and published literature (see *Chapter 2*). It included some of the factors that influence safe clinical practice developed by Vincent and colleagues:<sup>36</sup> organisational and management factors, work environment, staffing levels and skills mix, team factors, individual (staff) factors and task factors. The aim was to identify a sufficient but manageable range of items to support evaluations of organisational and safety cultures. The provisional checklist was piloted in an ED that is not included in the main study and it was anticipated that the checklist would undergo further refinement as the study progressed.

Provisionally, it was planned that items in the final checklist would be rated by the researchers who had conducted observations and supplementary interviews, using the following ordinal scale, which would permit numerical scoring: C, evidence that this facet of organisational or safety culture is consistently in place; P, partially in place/dependent on individuals; N, not in place. There was provision for rating DK, don't know/can't assess, which needed to be used sparingly to prevent missing scores from undermining the overall scoring. Space was provided in pilot checklists to note items of evidence to support the rating given. While piloting the checklist, researchers (experienced ethnographers) also wrote extensive field notes to inform and evidence their provisional ratings. Rating of sites was undertaken provisionally in the field and final ratings for each item allocated on the return from the field, based on the two researchers' provisional ratings and discussion and comparison of field notes and observations. The use of field notes permitted the accumulation of evidence to inform ratings, and permitted researchers to cross-reference with one another's observations, offering the opportunity to check inter-observer reliability. The initial checklist for the study was always anticipated to be too wide-ranging for real-time use in clinical areas, but the items continued to be useful prompts for researchers reviewing the comprehensiveness and comparability of the observation and interview data gathered at each research site so it was retained as a prompt list (see *Appendix 5*).

During piloting researchers found that some items in the initial checklist overlapped and it was difficult to gather evidence relating to some items. Researchers were sometimes uncertain about assigning ratings quickly in situ, and preferred recourse to detailed field notes. They were also concerned to make provision for recording unanticipated facets of the local environment or workplace practices that influenced workplace and safety cultures. Consequently, semistructured field notes were made during observation periods and annotated soon afterwards. The field notes were then coded with reference to the prompt list, which later evolved into the scoring guide for holistic assessment of culture (see *Appendix 6*).

While all items on the detailed prompt list (see *Appendix 5*) were useful to sensitise observers to relevant features of the clinical setting, and to structure field notes, assigning ratings for every item was not always possible, even with discussion between observers and reference to field notes. Overlap remained a problem, so the decision was taken to collapse items into fewer higher-level categories that could be scored consistently (see *Appendix 6*). We will discuss the scoring of the instrument in *Appendix 6* in *Chapter 4, quantifying semistructured observations*.

### Analysis

Extensive field notes, guided by sensitising prompts in *Appendix 5*, formed an assessment of culture in each of the research sites. (A separate inductive qualitative analysis of the field notes will be possible but does not form part of the commissioned study reported here.) Quantification

of the results was required in order to perform the comparisons intrinsic to this study and its commissioning specification. To achieve this the observations were categorised using eight domains (see *Appendix 6*), each of which was scored separately following the period of strategic immersion. Three of the categories related to safety culture at organisational level, and five at team level. Two research fellows (SA and MR) scored each site separately, after reading all field notes (from preliminary meetings, observations, brief conversations during observation periods, coresearchers and feedback meetings). They then met to agree a final score.

The organisation and work environment factors scores reported in *Chapter 6* (see *Table 27*) capture safety-related facets of the organisational culture in each of the research sites, namely the adequacy of staffing and premises, the availability of sufficient well-functioning equipment, and support from managers, administrative staff and other support roles. The mean of these scores was calculated to form a summary observation-based organisational culture score for use in subsequent analyses.

The team factors scores in *Table 27* (undertaking informal training and supervision, particularly of juniors and students; leadership offered by senior clinicians, team members taking responsibility for their own work, levels of individual and team situational awareness; evidence of respect, warmth and collegiality within communication and actions; the quality of information exchange within the team and evidence of mutual support within and between professions) capture several facets of the local culture of teamwork that have been linked to safety culture (see *Chapter 2*). The mean of team factors scores became a summary, observation-based teamwork culture score.

The grand totals in *Table 27* include organisational, work environment and teamwork factors that relate to safety. These form a more holistic evaluation of safety culture and the mean of the scores for all eight items in *Table 27* are the observation-based holistic safety climate score for each research site. However, this assessment is not distinct from the observation-based organisational culture assessment comprising organisation and work environment factors. Consequently, it was not possible to test hypothesis 2b as written. Instead, the observation-based assessments of organisational culture and teamwork culture are compared in *Chapter 6*.

## Strand C: criterion-based assessment of quality of care

This strand relates to objective 4 (see *Chapter 1, Aims and objectives*): to obtain criterion-based measurements for the quality of care at each research site. The methodology selected was a retrospective audit of clinical notes, conducted by non-clinical auditors.

The technique of explicit criterion-based reviews (the objective application of predetermined standards) was used.<sup>181</sup> The aim of maximising consistency of data extraction between research sites made the use of explicit criteria essential. One strength of the method is that non-clinical auditors obtain results more similar to clinical auditors in criterion-based review than in holistic review, i.e. assessment of the process of care without specified criteria.<sup>149</sup>

Separately for DUs and EDs, common conditions were identified for which there was a strong evidence base and national clinical guidelines on appropriate actions and care processes. Based on prestudy power calculations, we aimed to select three audit standards per condition. Our choice of conditions and standards was guided by the strength of the evidence base for each aspect of the relevant clinical guidance (see sources at foot of *Box 1*): we produced short-lists of conditions and standards drawn from the literature and reached our final choice after extensive

dialogue with clinicians (project advisors, service provider coresearchers and clinicians from the research sites recruited to the study before data collection began at site 1).

Some conditions commonly received in EDs were removed from the prestudy shortlists because, despite strong evidence and clear guidelines, we were advised that facilities and local practices varied widely. Stroke was the most notable example of this. Despite considerable efforts to include this condition in the study, we were unable to find sufficient well-evidenced markers of the quality of care for which reasonable consensus emerged regarding acceptability as markers of quality and both feasibility and validity of measurement. We discussed our choice with the study's senior clinical advisors (see *Acknowledgements*), clinicians from our pilot sites (DU and ED in a hospital not involved in the main study) and research sites 1, 2, 4, 7 and 8. These discussions checked the acceptability of the selected markers to clinicians: to help us recruit sites, we thought it wise to focus on conditions and markers that clinicians viewed as acceptable measures of clinical care within their departments.

The conditions selected for the criterion-based audit of care within emergency departments were:

- acute coronary syndrome (ACS)
- acute severe asthma (ASA) and
- fractured neck of femur (FNoF).

For labour wards the selected conditions were:

- normal delivery (ND)
- emergency caesarean section (ECS) and
- delivery after the detection of grade 2 or grade 3 meconium-stained liquor (MSL) before delivery.

Inclusion and exclusion criteria for each condition are listed in *Table 1*. Sources for the evidence-based markers appear at the foot of *Box 1*.

For two conditions, ASA and FNoF, clinicians argued strongly that a meaningful audit of the quality of care would require additional markers so these have four and five audit standards, respectively (see *Box 1*). ECS audit standard 2, prophylactic antibiotics, was considered vital by clinicians but they anticipated that compliance would be so near to 100% that there would be too little 'headroom' for variation between sites to be evident. Thus, ECS was allocated four audit standards.

Audit data extraction using the evidence-based markers was piloted at a hospital that did not participate in the main study.

### **Audit data collection**

At each hospital we aimed to audit 50 consecutive cases for each condition, working backwards for up to 6 months from the date the researchers first became noticeable in the clinical areas by commencing observations. The audit worked backwards to avoid any concerns about a possible Hawthorne effect emerging after researchers were encountered in clinical environments, leading to more than usually diligent record-keeping. The time limit of no more than 6 months was set so that the audit data could be regarded as contemporaneous with the observation data and staff survey responses. For MSL the audited markers spanned mothers' and babies' clinical records. Mothers' records had to be audited first to obtain required audit data and identify the hospital record number for each associated baby, then the babies' notes could be ordered and audited.

**TABLE 1** Inclusion and exclusion criteria for audited conditions

<i>Emergency departments</i>	
ACS	Exclusion: suspected MI with ST segment elevation on ECG
ASA	Inclusion: one of peak flow rate 33–50% of best or predicted respiratory rate $\geq$ 25/minute; heart rate $\geq$ 110/minute inability to complete sentences in one breath
FNoF	No exclusions
<i>Delivery units</i>	
ND	Inclusion: all the following must be true: spontaneous onset of labour and delivery no spinal or epidural or general anaesthesia gestation > 37 weeks no medical condition such as diabetes or hypertension
ECS	Inclusion: grades 1–3 ECS only
MSL	Inclusion: both must be present: meconium grades 2 or 3, detected before delivery baby in good condition at birth (Apgar score > 6)

ECG, electrocardiogram; MI, myocardial infarction.

Data extraction sheets were developed to structure the audits. One example, the data extraction sheet for FNoF, is provided in *Appendix 7*.

Non-clinical auditors were preferred, as earlier research found this reduced hindsight bias,<sup>186</sup> and more recent research has found that non-clinical auditors are no less reliable than clinical.<sup>145,156</sup> The study aimed to recruit local audit clerks and provide study-specific training and ongoing support. The study budget made provision to reimburse trusts for the audit clerks' time. Audit clerks sent data extraction sheets to the research fellows, who checked the extracted data and made enquiries about any anomalies they found (e.g. suspected duplicates, apparent inclusion of cases that did not meet inclusion criteria, suspected recording errors and patterns within missing data that suggested that certain sections or types of clinical record were being overlooked). Once enquiries were satisfactorily resolved, data from the extraction sheets were coded and data entry was checked before statistical analysis. In some sites, the research fellows carried out the audits (see *Chapter 4, Engaging non-clinical auditors*).

Aligned with objective 1 (see *Chapter 1, Aims and objectives*), which sought to work supportively with staff in the research sites and offer prompt feedback to allow local development, we undertook some finer-grained data extraction. In addition to the minimum data required to assess whether the audit standards had been met, we collected additional contextual data to enable richer feedback to research sites. For example, rather than record simply whether or not aspirin was given immediately, we also recorded how long it took to give aspirin in non-immediate cases. Thus, research sites could be given information about how far short of the standard they fell during the sampled period.

## Analysis

In order to calculate each site's scores in relation to each standard, we calculated the number of cases where the standard was met as a percentage of the total number of cases to which it applied (i.e. removing cases from the denominator such as, for standard ACS2, when aspirin had already been given by the ambulance service, or the patient had taken regular daily aspirin, or the patient



**BOX 1** List of audit standards***Acute coronary syndrome (excluding MI)***

1. Electrocardiography performed immediately
2. Aspirin should be given immediately if not already given by ambulance service
3. If patient in pain, times from arrival to the administration of pain relief:
  - in cases of severe pain, 50% in 20 minutes, 75% in 30 minutes, 98% in 60 minutes
  - in cases of moderate pain, 75% in 30 minutes, 98% in 60 minutes

***Acute severe asthma***

1. Measure and record oxygen saturation on arrival in 98% of cases
2. Salbutamol or terbutaline given within 10 minutes of arrival
3. Intravenous hydrocortisone or oral prednisolone given within 30 minutes of arrival in 90% of cases
4. Repeat measurement of oxygen saturation within 60 minutes in 75% of cases

***Fractured neck of femur***

1. Pain score recorded on arrival
2. Time from arrival to receive analgesia:
  - (a) if severe pain, 50% in 20 minutes, 75% in 30 minutes, 98% in 60 minutes
  - (b) if moderate, 75% in 30 minutes, 98% in 60 minutes
3. Radiography performed within 60 minutes of arrival in 75% of cases
4. If in pain, evidence of re-evaluation of pain (90% of those with severe pain re-evaluated within 30 minutes; 75% of those in moderate pain within 60 minutes)
5. Admitted within 4 hours of arrival

***Normal deliveries***

1. Initial assessment should include temperature, pulse, blood pressure, urinalysis, length, strength and frequency of contractions, fundal height, lie, presentation, position and station, show, liquor, blood, pain and fetal heart rate
2. Assessment during first stage should include fetal heart rate every 15 minutes, frequency of contractions every 30 minutes, pulse hourly, temperature and blood pressure 4-hourly, vaginal examination offered 4-hourly, frequency of emptying bladder
3. Assessment during second stage should include fetal heart rate every 5 minutes, frequency of contractions every 30 minutes, pulse and blood pressure hourly, vaginal examination offered hourly, frequency of emptying bladder, woman's emotional/psychological needs

***Emergency caesarean section***

1. Documentary evidence of consultant obstetrician involvement in the decision to perform ECS
2. Women having an ECS should be offered prophylactic antibiotics at the time of ECS
3. All women undergoing ECS must receive thromboprophylaxis for VTE
4. Women undergoing ECS should be offered regional anaesthesia (spinal or epidural)

***Meconium-stained liquor***

1. Continuous EFM should be advised
2. Health-care professionals trained in advanced neonatal life support should be available for the birth
3. Baby assessment should include at 1 and 2 hours and then 2-hourly for 12 hours:
  - general well-being
  - chest movements and nasal flare
  - skin colour including perfusion
  - feeding
  - muscle tone
  - temperature
  - heart rate, and
  - respiratory rate

EFM, electronic fetal monitoring; MI, myocardial infarction; VTE, venous thromboembolism.

Sources: College of Emergency Medicine (2008),<sup>182</sup> Scottish Intercollegiate Guidelines Network (2002),<sup>183</sup> National Institute for Health and Clinical Excellence (2007),<sup>184</sup> NHS Information Centre for Health and Social Care (2004).<sup>185</sup>

was allergic to aspirin). We then calculated the mean level of compliance for each condition at each site so that, regardless of the number of audit standards per condition, each condition would contribute equally to the summary audit score for each site. It is appreciated that this method of calculating one summary score for each condition may result in a good overall score being obtained for a condition even when some of the individual standards had very low values. It was felt, however, that each standard should be given equal weighting within each condition. Possible pitfalls in this approach will be discussed in *Chapter 9, Objectives*. The mean scores were inspected for normality (see *Chapter 3, Testing for normality*) and then converted to the standard normal distribution. Finally, summary audit scores were calculated as the mean of the condition-specific  $z$ -scores for the three conditions measured in each research site (ED conditions or DU conditions, as appropriate).

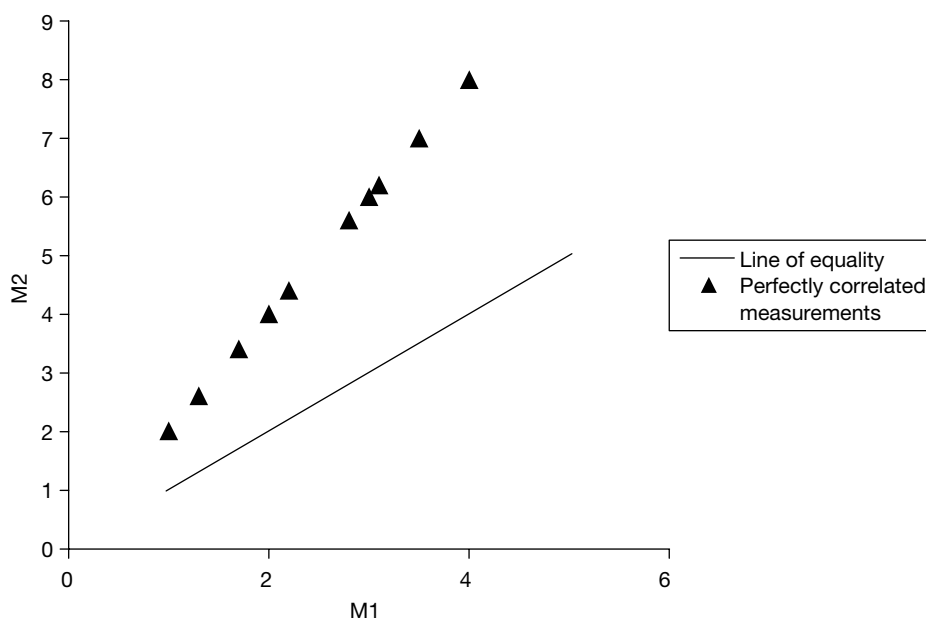
### Testing the study hypotheses: correlation and agreement

This study was commissioned to compare contrasting approaches to evaluating safety culture and to compare each of these with a criterion-based assessment of the quality of care (see *Appendix 1*). There are epistemological and practical difficulties in this, which will be examined in *Chapter 9, Objectives*. This section describes the comparisons that were made, which will be reported in *Chapters 5 and 8*.

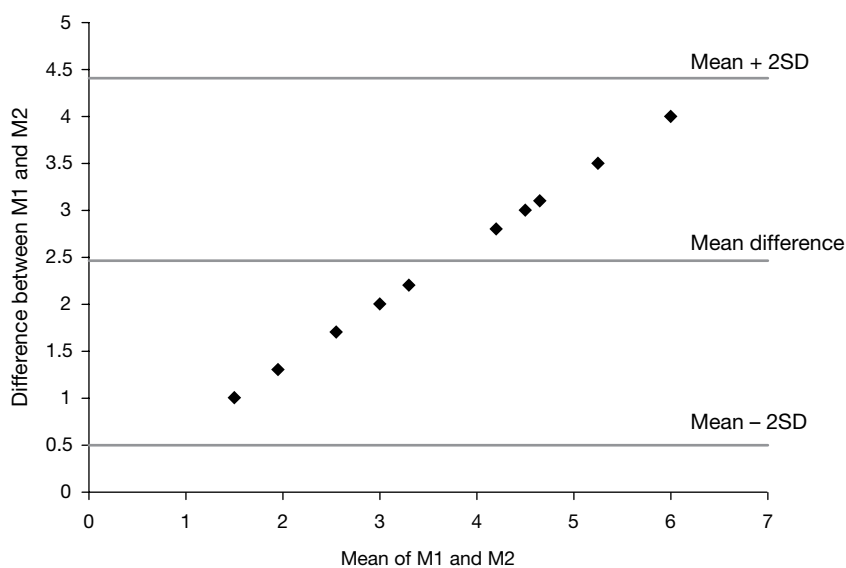
The Pearson product–moment correlation coefficient ( $r$ ) measures the extent to which two sets of measurements lie on a straight line, when plotted on a graph. For example, the measurements M1 and M2 in *Table 2* perfectly fit a straight line (*Figure 1*) because M2 is simply twice M1; there is perfect positive correlation between M1 and M2 ( $r = 1$ ). However, if these two sets of measurements agreed (were the same), they would lie along the line of equality in *Figure 1*; clearly they do not. Bland and Altman<sup>187</sup> argued that correlation was not the correct approach for assessing agreement, instead recommending visual inspection of the type of graph shown in *Figure 2*. This plots the mean of pairs of measurements along the  $x$ -axis and the difference between pairs of measurements on the  $y$ -axis. In the case of measurements made on the same scale, if the measurements agree (or approximately agree), the difference between them will be zero (or close to zero) and the mean of differences for all pairs of measurements will be zero. However, if one measure is consistently higher than the other, the mean of differences will be higher than zero, which is the situation we see in *Figure 2*, in which the line marked ‘mean difference’ lies at 2.46, whereas those marked ‘mean  $\pm 2SD$ ’ lie at 4.416 and 0.504. The lines marked ‘mean  $\pm 2SD$ ’ provide an approximate 95% confidence interval, and most points on the graph should lie between these lines (in fact, all points do in *Figure 2*).

**TABLE 2** Fictitious measurements generated to discuss correlation and agreement

M1	M2	M3	M4
1.0	2.0	0.9	1.5
1.3	2.6	1.3	0.7
1.7	3.4	1.0	2.6
2.0	4.0	1.5	3.2
2.2	4.4	2.3	3.3
2.8	5.6	2.9	3.0
3.0	6.0	3.0	1.5
3.1	6.2	3.1	2.5
3.5	7.0	2.6	0.9
4.0	8.0	2	0.8



**FIGURE 1** Scatterplot illustrating perfect correlation.



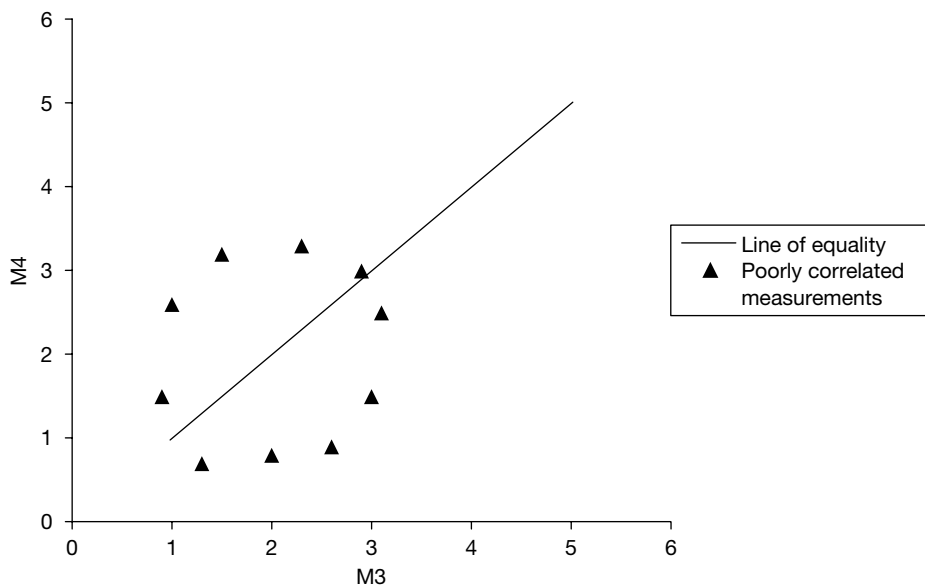
**FIGURE 2** Bland-Altman plot illustrating poor agreement.

Figure 2 is commonly termed a Bland-Altman plot in health research, although known as a Tukey mean-difference plot in other fields. In Figure 2, the line of equality is the  $x$ -axis ( $y=0$ , no difference between the two measurements). If measurements M1 and M2 were the same apart from small random fluctuations in the readings, we would see points lying mainly close to the  $x$ -axis, above and below, scattered with no discernible pattern. However, in Figure 2 all the points lie above the  $x$ -axis and there is a clearly discernible pattern. In this case, the mean difference is sufficiently large that the 95% confidence interval also lies above the  $x$ -axis, but matters are not always so clear-cut. However, for the sets of measurements M1 and M2, which were made on the same scale, we can confidently say that, while perfectly correlated, they do not agree.

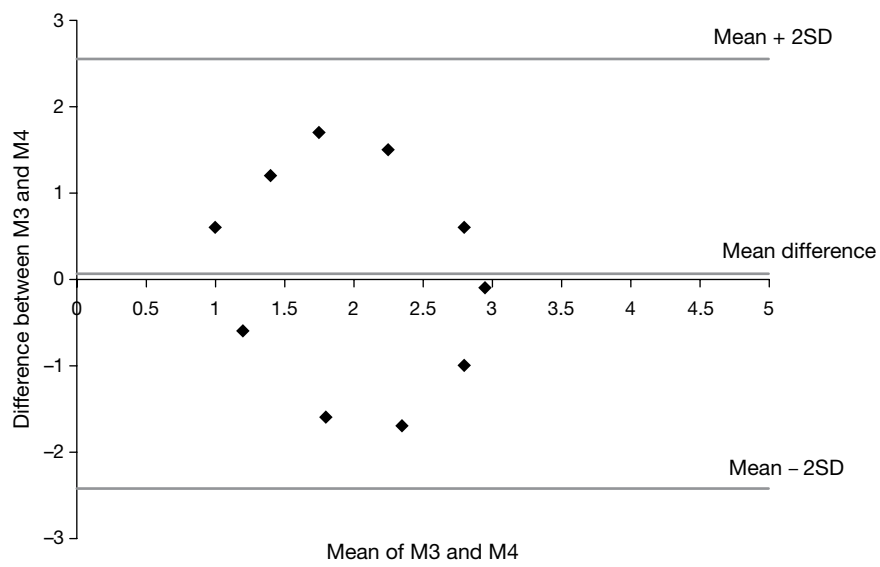
It is also possible to have agreement with poor correlation, simply because measurements do not lie on a straight line. We will illustrate this with a second artificial example; measurements M3 and M4 in *Table 2* are poorly correlated ( $r=0.120$ ), because they more or less form a circle (see scatterplot, *Figure 3*). However, the mean difference between the measurements is very small (0.06; see Bland–Altman plot, *Figure 4*) and the SD of differences is 1.25. All points on this graph lie within the interval  $\text{mean} \pm 2\text{SD}$  and the largest difference is 1.34 SDs from the mean. Measurements M3 and M4 lie close together, suggesting good agreement. But this is an artificial example in which the measurements were constrained to form a circle. It is the presence of a discernible pattern within *Figure 3* that alerts us to a systematic relationship between measurements M3 and M4 rather than agreement within the limits of normal random errors. However, in real examples we should not expect the discernible pattern to ‘shout so loudly’ from the graph as it does in *Figure 3*: careful visual inspection is needed on each occasion.

In the analyses that follow, a scatter diagram and a Bland–Altman plot will be drawn for each comparison to permit interpretation of both correlation and agreement. This requires that the same scale is used for each measurement and so normally distributed raw scores for different types of measurement will be standardised, that is converted to scores on the standard normal distribution (mean 0, SD 1), known as  $z$ -scores. The expected mean difference between sets of standardised scores is zero, therefore the Bland–Altman plots in *Chapter 8* are expected to have mean differences of zero. It is the distribution of the points that will be inspected, first for discernible patterns and outliers and, second, for size. The graphs will be plotted over the same area to ease comparison.

The acceptability of the level of agreement between two measures depends on professional judgement of the practical significance of the size of differences. This interpretation can be straightforward when differences are measured in units such as kg, ml/minute or mmol/l and differences have clinical significance. However, in this study, the measurements are unit free and how large a difference between measurements is acceptable is a moot point. Nevertheless, the smaller the SD of differences, the closer the agreement between the different approaches to measurement.



**FIGURE 3** Scatterplot illustrating low correlation.



**FIGURE 4** Bland-Altman plot for poorly correlated measurements.



## Part II

### Results

There are five results chapters. First, *Chapter 4, Process data and issues arising during the study*, summarises process results arising during data collection, including such matters as the challenges associated with each of the three data collection methods (and more general issues such as site recruitment and research governance). *Chapters 5 (Strand A: questionnaire survey of staff perceptions of safety and organisational climates)*, *6 (Strand B: observation-based holistic assessment of culture)* and *7 (Strand C: retrospective audit of evidence-based markers of the quality of care)* present the results from strands A–C separately, namely the staff survey, observation-based assessments and the criterion-based audit. *Chapter 5* includes testing of hypothesis 2a, and hypothesis 2b is tested in *Chapter 6, Strand B: observation-based holistic assessment of culture*. Finally, in *Chapter 8, Comparison results from strands A–C*, the results from strands A–C are compared; this includes testing of hypotheses 1a and b and 3a and b.





## Chapter 4

# Process data and issues arising during the study

As this is a methodological study, it is important to report and discuss research processes as well as outcomes, as these will help inform future use of the research methods employed. The three methods posed different problems, some of which (e.g. the survey response rate; see *Samples and response rates*) were predicted, and others (e.g. identifying case notes for audit; see *Cases audited*) not anticipated.

### Recruiting research sites

A total of 13 NHS trusts were approached sequentially until eight hospitals had agreed to participate in the study. During this process five hospitals declined to participate; where reasons were given these related to high workloads for staff. Naturally, hospitals needed time to consider whether to participate in the study, during which some key stakeholders in the relevant clinical areas asked the research team for clarification, by phone or e-mail and, in one case, by personal visit to the hospital by a member of the research team. The recruitment of each hospital normally took a few weeks, but at some hospitals there were delays of several months before the decision to join the study was made. This caused substantial delays in the data collection timetable.

Purposive sampling, using the criteria outlined in *Chapter 3, Selecting the sample of research sites*, used published statistics to identify potential research sites, but the professional knowledge and networks of the study team and study advisors were also very important. By integrating publicly available information and personally held professional knowledge, we were able to fill gaps in published data and we were better informed about ongoing or planned major organisational restructuring which would render particular hospitals unable to participate in the study. We were also alerted to parallel studies in DUs and EDs making it inappropriate to ask these departments to participate in this study too.

Personally held professional knowledge was also very important for identifying appropriate senior clinicians and managers with whom to make initial contact to introduce the study. Trust websites proved unsuitable for this task. Patient advice and liaison services were able to offer well-targeted assistance at some hospitals, but not at others.

The sample of hospitals from which delivery units and emergency departments were recruited matches the planned purposive selection well (see *Chapter 3, Selecting the sample of research sites*), as illustrated in *Table 3*. The sample also contained diversity in relation to the populations served by the hospitals and the ease of recruiting and retaining a well-qualified workforce, but we have chosen not to extend *Table 3* to reduce the risk of compromising anonymity. All the DUs included in this study were consultant led, and levels of midwife-led care varied across them.

Three DU research sites (8, 12 and 14) had previously participated in a study focused on simulation-based education and safety-related workplace practices: the Multidisciplinary Obstetric Simulated Scenarios (MOSES) study.<sup>34,132,188</sup> They were invited to join this study because they met the purposive sampling criteria (a fourth site from the MOSES study did not meet

**TABLE 3** Main sampling characteristics of research sites

Sites: hospital	Strategic Health Authority region	Type of hospital and NHS trust	2008–9			
			Trust income (£millions)	Patients ('000s)		Quartiles
				ED	DU	ED, DU
1, 2: H1	A	Tertiary centre, foundation hospital trust	500	141.9	5.0	Q4, Q3
3, 4: H2	B	Tertiary centre, foundation hospital trust	500	86.2	5.9	Q3, Q4
5, 6: H3	C	General hospital, hospital trust, primary care trust (site 6) <sup>a</sup>	200	66.8	4.6	Q2, Q3
7, 8: H4	A	General hospital, hospital trust	150	90.3	5.2	Q3, Q3
9, 10: H5	D	Tertiary centre, hospital trust	250	99.9	3.7	Q3, Q2
11,12: H6	E <sup>b</sup>	General hospitals, foundation hospital trust	300	86.8	3.9	Q3, Q2
13,14: H7				38.0	1.4	Q1, Q1
15,16: H8	F	Tertiary centre, hospital trust	650	156.0	6.3 <sup>c</sup>	Q4, Q4 <sup>c</sup>

a Provides midwifery services; hospital trust provides medical services.

b Sites 11–14 were in two hospitals that are part of the same NHS foundation trust, but some distance apart, with minimal overlap of staff and mainly serving separate communities.

c Data obtained from hospital, not available on BirthChoice.

Sources: Trust-specific annual reports (not referenced to preserve anonymity of research site), NHS Hospital Episode Statistics,<sup>189</sup> BirthChoice.<sup>190</sup>

the sampling criteria). All three DUs accepted the invitation to participate. These decisions may have been influenced by memories of good working relationships established during the previous study, although at one site nearly all senior staff had changed in the interim. SA and MR negotiated the participation of these three units and the data collection arrangements. The research fellows from the MOSES study (EJB and NM) did not participate in these processes. The research team decided to place these research sites late in the data collection timetable to ensure a break of around 2 years since the MOSES study's final phase of data collection in these units.

## The sequence of data collection

The core staffing for this study was two 0.5 full-time equivalent research fellows, making it impossible to work in parallel at a large number of research sites. We intended to work at each hospital (including both departments) for 12-week periods in sequence, with periodic 1-month breaks in the data collection timetable for provisional analyses and as a contingency for data collection over-running at any site resulting from the complexity and limited predictability of 'real-world' research.<sup>126</sup> For example, delays in recruiting research sites were mentioned in the previous section. Delays in research governance processes (*Research governance*) and the recruitment of trust-based auditors (*Engaging non-clinical auditors*) put further pressure on the study timetable. More importantly, it became clear that audit data collection would take longer than the other methods, because of problems encountered in identifying and obtaining records for data extraction, which varied from site to site, and because some auditors took much longer than others to complete the task (*Unpredictable resource requirements*). Data collection was begun in the order shown in *Table 3*, but varied in duration. It was of no methodological consequence if completing retrospective audits over-ran into data collection at the subsequent site, but efforts were made to complete observations at one hospital before commencing observations at another hospital to help observers maintain a coherent perception of the culture that they were assessing.

## Close-coupling of departments and services

The clinical departments in this study were closely coupled to other departments and services. For example, EDs are closely coupled to pre-hospital care from ambulance trusts and general practice and, within the hospital, to the medical assessment unit and wards, which receive patients from the ED and service departments such as the radiography department. The DUs in this study were closely coupled to community-based care, antenatal and postnatal wards, neonatal intensive care, and sometimes maternity assessment units and midwife-run birthing centres. Some aspects of the performance of the clinical departments in the study were, therefore, linked to the performance of other departments in the continuum of care. In addition, EDs and maternity services deploy staff flexibly over a range of specialised clinical areas. This study was focused on DUs and the major injuries sections of EDs. Staff working in these areas normally worked in related clinical areas on a flexible or rotational basis. Although the observation and audit strands of this study could be closely focused on the selected specialised clinical areas, this was more difficult for the staff survey although distribution was targeted at those currently working in the selected specialised area. Questionnaire responses may well reflect the totality of respondents' working experiences rather than just the areas we observed and audited.

## Impact of staff changes in research team

The staff team changed during the study. Two research fellows, authors NM and EJB, helped DF and JS design and set up the study. NM and EJB conducted the majority of data collection at sites 1 and 2, then both left the team (one temporarily, on maternity leave). SA and MR were appointed to replace the departing fellows; SA worked alongside the original two for a few days. EJB worked briefly on the project on returning from maternity leave, contributing to discussions about methods and interpretation, and doing observations at sites 7 and 8. Weekly meetings between the principal investigator (DF) and the project research fellows reviewed the data collection and helped smooth the transition from the first pair of research fellow observers to the second pair. When the observations were scored (see *Chapter 6, Scoring*), EJB assisted with scoring for research sites 1 and 2.

## Research governance

Research governance for this project proved particularly arduous and time-consuming. The literature contains accounts of the challenges of research ethics committee (REC) processes,<sup>191,192</sup> but our experience of applying for REC was relatively straightforward. It was approved by the Oxfordshire Research Ethics Committee C (REC reference: 07/H0606/87).

Much more demanding were the research and development (R&D) governance requirements of the eight NHS organisations where we carried out the research (seven NHS hospital trusts and one primary care trust). This accords with the experience of other researchers.<sup>193–195</sup>

Requirements varied significantly between sites and were not always clear (in two sites, significant changes were made in what was required during the process). For example, some sites required their own occupational health departments to clear us while others did not; some accepted the Criminal Records Bureau (CRB) check carried out at the request of our own employer, others applied for another or required that we did so.

In many sites, we had to deal in parallel with an R&D governance officer for approval of the research as a whole, and a human resources officer processing our applications for honorary contracts or letters of access. These parallel processes were not always streamlined (we learnt to prompt officers to check on progress with the other, though even then this was not always done).

Several sites required us to fill in the site-specific information part of the REC application process even though, according to REC definitions, and the committee that approved our study, this was neither applicable nor required. This posed a particular problem for sites 15 and 16 because the research ethics service website hosting our original application had been closed: although documents were still available for consultation and printing, they could not be amended with local details.

Although in theory the new Research Passport system was introduced as the study began, sites where we worked for the first year were not ready to recognise the system, and required us to have honorary contracts instead.

*Table 4* summarises some of the inconsistencies and anomalies that we found in our applications to carry out the research. The cumulative effect of these slow processes was that data gathering was delayed and analysis time significantly curtailed. The need for CRB checks at the instigation of the trust reduced our ability to deploy service user coresearchers in the observation strand. Both already had recent CRB clearance: one was alienated by what he perceived to be needless bureaucracy and withdrew, and the help of the second was severely limited by administrative delays in further CRB clearance.

Our interpretation of the reasons for delays and cumbersome processes can only be conjectural, but we believe that:

- R&D departments found it difficult to process applications for non-clinical mixed methods research (hence confusions about occupational health clearance, even though we had no contact with patients, and the requirement at the eighth hospital to pass an online examination in running clinical trials).
- Applications are sometimes handled by junior officers without the expertise or authority to make exceptions for non-clinical research.

It was unfortunate that this study coincided with a period of crisis in research governance arrangements, leading to widespread criticism that the establishment and timely progress of research studies were being threatened. The ensuing national review and new arrangements should streamline research governance processes,<sup>196</sup> although they will favour recruitment to clinical trials over other types of research, so the benefit for research similar to this study may be limited.

## Strand A: staff survey

### *Samples and response rates*

Some clinical leads identified team membership themselves (see *Chapter 3, Sampling, delivery and maximising returns*); others delegated the task to junior colleagues or, most often, administrative staff. No site had a consolidated staff list, so different lists for different professional groups were provided. Generally, departments found lists of nurse or midwife names easy to provide, though some did not provide roles or bands, and some were out of date (excluding more recent arrivals whose names we learnt during observation: where possible, these were added). Lists of doctors proved more challenging (in one case, no list was provided at all, despite multiple requests, and

**TABLE 4** Summary of R&D governance at eight NHS organisations

Site	Duration of application process in months	SSI needed?	Trust occupational health clearance needed?	Needed trust's own criminal records check?	Accepted research passport?	HC or LA	Requirements changed during process?	Notes
<b>1, 2</b>	1	Y	Y	N	N	HC	Y	Needed additional letter for medical records.
<b>3, 4</b>	4	Y	Y	Y	N	HC	Y	Department uncertain about how to apply its own requirements
<b>5</b>	2	Y	Y/N	Y	N	HC	Y	Occupational health requirements varied between researchers
<b>6</b>	2	N	N	N	N	Neither	N	Midwifery services provided on hospital site by primary care trust
<b>7, 8</b>	4	Y	Y then N	N	Y	HC	Y	Officer consistently unclear and unhelpful
<b>9, 10</b>	5/10 <sup>a</sup>	N	N	Y	Y	LA	Y	Would not permit audit without NIGB support. Approval of audit delayed by a further 5 months because of NIGB application
<b>11, 12, 13, 14</b>	5	Y	N	N	Y	LA	N	A researcher had to visit in person to receive letters of access (110 miles distance)
<b>15, 16</b>	6	Y	N	N	Y	LA	Y	Very straightforward, although the process took a long time Initially, clinical trials training and examination was required of the research fellows. This was waived only if researchers went in person for general research training (100 miles)
								Local contact had to register as part of research team
								All information required to be on hospital's own headed notepaper though the hospital was not responsible for the research (eventually agreed to logo only)
								MREC documents had to be submitted, some in three formats (paper, PDF and XML files). Original MREC application for SSI purposes no longer accessible on-line

HC, honorary contract; LA, letter of access; MREC, Multicentre Research Ethics Committee; NIGB, National Information Governance Board; SSI, site-specific information form from Research Ethics Committee application.

<sup>a</sup> Five months initially, with a further 5 months for application to NIGB and agreement to provide trust staff to do audit.

we had to rely on the hospital website, which listed only consultants). These added to factors already identified (see *Chapter 3, Sampling, delivery and maximising returns*) meant that samples were not comparable across sites. In theory, we could have tried to gather further information in order to edit and harmonise lists, but we were convinced that we would have encountered considerable resistance from overstretched departments that would have been sufficient to prevent us achieving our objective.

It became evident that some databases were not quite up to date because 21 questionnaires were returned with a message saying that the individuals to whom they were addressed no longer worked in the department. These have been subtracted from the figures shown in the 'surveys sent' column of *Table 5*. We cannot know how many other questionnaires could not be delivered to the named recipients but were not returned. Turnover of staff may mean that a small number of new staff who could have participated in the survey were not invited to do so. This was occasionally apparent when observations for strand B identified a member of staff from an eligible staff group who was not listed for inclusion in the staff survey. Some of the lists we received contained excluded categories (see *Chapter 3, Sampling, delivery and maximising returns*), mainly Foundation Programme doctors. We tried to clean the lists before questionnaires were sent, within the limits of the information fields provided in each list. Many sites did not provide sufficient information to allow us to augment *Table 5* with a breakdown of the surveys sent by professional group. However, where the professions of respondents are known they are summarised in *Chapter 5, Demographics and role-related characteristics*. Interactions with key contacts suggested that requests to improve lists were likely to be perceived as problematic, so we chose to accept what we were given, even when imperfect. This simply reflects the reality of conducting research in increasingly busy, pressurised health services: pragmatic decisions must be made or research would halt.<sup>126</sup> Nevertheless, the pragmatic decisions should be noted to allow reflection on their consequences. Published reflection on pragmatic decisions and known or likely consequences occurs less frequently than that which might be considered warranted by those who are familiar with the day-to-day imperfections and predicaments of research. This is likely to be attributable to publication bias and also, in Goffman's terms,<sup>197</sup> to a degree of 'idealization' in the 'front stage performances' of researchers.

Decisions about which groups of staff to include in the survey varied, for example site 3 was particularly interested in eliciting the perceptions of support staff and included a wider range of support roles than was the case at other sites. Because site 3 adopted a broad definition of team membership, we distributed twice as many questionnaires to staff at site 3 as at site 11 (see *Table 5*), although, as shown in *Table 3*, the two sites had a very similar patient throughput. Varying local decisions mean that there is no simple relationship between the size of the clinical department, as indicated by patient throughput, and the number of people designated team members. Furthermore, we would also expect the number of people designated team members to be affected by the local prevalence of part-time working, data which we do not have.

The overall response rate of 27.6% was very similar to the maximum response rate of 27% achieved by Catchpole and colleagues,<sup>178</sup> who distributed safety climate questionnaires to staff in two surgical departments. It was a little higher than the 22% estimated response rate achieved by Sarac and colleagues,<sup>100</sup> who distributed Hospital Patient Safety Culture (HPSC) questionnaires to clinical staff in seven acute hospitals in Scotland. However, the response rates for this study were consistently lower than the trust-wide response rates published for the NHS annual staff survey (*Table 6*). Comparing *Tables 5* and *6*, it is interesting to note that both data collection exercises yielded noticeably lower response rates at the trust in which research sites 7 and 8 were located. Ideally, we would have wished to conduct a non-response analysis to discover any systematic differences between survey respondents and non-respondents, based on sample characteristics provided by research sites with staff lists for the survey (e.g. profession, grade and demographic

**TABLE 5** Staff survey questionnaires sent and returned

Hospital	Site	Surveys sent	Surveys returned	Response rate (%)
H1	1 ED	132	36	27
	2 DU	133	27	20
H2	3 ED	266	83	31
	4 DU	109	38	35
H3	5 ED	111	35	32
	6 DU	60	28	47
H4	7 ED	119	17	14
	8 DU	76	7	9
H5	9 ED	94	37	39
	10 DU	100	30	30
H6	11 ED	111	25	23
	12 DU	74	27	36
H7	13 ED	55	21	38
	14 DU	59	25	42
H8	15 ED	249	40	16
	16 DU	176	55	31
	All sites	1924	531	27.6

variables). However, as noted above, it was difficult for research sites to provide us with up-to-date staff lists and we did not obtain sample characteristics beyond those that routinely appeared on staff lists in each research site, mainly, but not always, profession, grade and gender. Missing and inconsistent information on sample characteristics precluded non-response analysis.

### Examining influences on response rates

Multilevel modelling estimated the overall site within hospital response rate to be 29.5% [standard error (SE) 3.2%]. Variation in response rates was partitioned as follows:

- 60% at hospital level (estimated variance 0.006, SE 0.005,  $\chi^2 = 1.820$ , 1 df,  $p = 0.177$ )
- 40% at research site level (estimated variance 0.005, SE 0.003,  $\chi^2 = 4.000$ , 1 df,  $p = 0.046$ ).

Variation in response rates associated with differences between hospitals was not statistically significant but variation due to differences between research sites was significant at the 5% level. Adding site-level characteristics to the model, as described in *Chapter 3, Multilevel modelling*, showed that the type of service (ED or DU) and the quartile in which the department lay (see *Table 3*) did not have significant effects of the response rates. There was a very small, statistically significant, effect for the number of questionnaires sent (estimated parameter  $-0.001$ , SE  $< 0.001$ ,  $\chi^2 = 4.806$ , 1 df,  $p = 0.028$ ). This means that response rates are predicted to fall by 0.0032% ( $0.001 \times \text{SD}$  of 3.2%) for every additional 62 (1 SD) people included in the staff survey. This effect is too small to have any practical effect linked to large departments or departments defining the team very broadly.

Adding the number of questionnaires sent as a fixed factor in the multilevel model reduced the residual site-level variance slightly to 0.004 (SE 0.002,  $\chi^2 = 3.993$ , 1 df,  $p = 0.046$ ). There remains significant site-level variation that cannot be explained by the site-level variables we collected in this study. The residual site-level variation may be partly due to site-level factors not recorded in this study. It may also be partly due to the characteristics of individuals working within the research sites. The analyses in *Chapter 5* include individual-level factors.

**TABLE 6** National staff survey response rates (%)

Sites <sup>a</sup>	2009	2008
1, 2	43	58
3, 4	55	56
5	57	58
6 <sup>b</sup>	66	68
7, 8	33	28
9, 10	Not available	Not available
11, 12, 13, 14	51	54
15, 16	60	65

a National response rates are trust based and cannot be disaggregated by research site.

b Data here are for the PCT that employs the midwives. Medical staff are employed by the same trust as those at site 5.

Source: Healthcare Commission (2006).<sup>175</sup>

In retrospect, it seems that the lower-than-hoped for response rates at sites 1 and 2 are not adequately explained by the length of the questionnaire, as response rates were lower at some subsequent sites where the shorter questionnaire was used. A further attempt to increase response rates was made by providing an online option at sites 11–14. An online survey was not offered at the beginning of the study because we knew that many front-line staff had only limited access to computers. However, web access using hand-held devices and wireless networks became popular during the course of this study and it was thought that respondents might find this more convenient and attractive. The online questionnaire was developed using Smart-Survey™ online survey software<sup>198</sup> and hosted on the Smart-Survey website. However, we were still reliant on conventional post to alert staff to the survey and to provide the web address, as research sites were not asked to supply e-mail addresses for individual staff. Consequently, we could not provide the convenience of clicking on a hyperlink embedded in an e-mail. No-one at research sites 11–14 submitted an online survey. For this reason, it was not offered at sites 15 and 16.

## Strand B: observation-based assessment of culture

At most research sites, data collection began with preliminary meetings with senior staff designated by the clinical department. At some sites, this was supplemented by attendance at routine meetings to explain the project to more junior staff. At one site, a planned meeting had to be cancelled, and the first observation visit was made after e-mail contact only.

### Observable activities

Within EDs, the observations were piloted within ‘major injuries’ and ‘minor injuries’ areas, but proved unsatisfactory in minor injuries areas, where cases tended to be dealt with by individual staff in closed rooms or curtained cubicles. Observers at the record-keeping and workload management hub of minor injuries had insufficient opportunity to observe interactions and work practices that would allow scoring in the eight domains of the observation prompt list (see *Appendix 5*). The conditions that had been selected for the criterion-based audit (see *Chapter 3, Audit data collection*) were all treated within major injuries, making it logical to restrict observation to the major injuries sections of EDs.

There was considerable variety in what could be observed from each vantage point (see *Table 7* for a summary of this). In the list that follows, ‘fully observable’ means that the specified activities were as a general rule both visible and audible from our vantage point. ‘Partly observable’ means that activities were either only visible, or only audible, or partly visible and/or partly audible.



- At eight research sites the whiteboard was fully observable; in four, partly observable; and in four (all EDs), there was no whiteboard (computer monitors were used instead, not observable).
- At four research sites (all DUs), medical and nursing handovers were fully observable; in 11 sites, some or all handovers were partly observable; in one site, handovers took place off the ward.
- At one research site the main area for doctors to meet, review care and write up clinical records was fully observable; in four sites it was partly observable; in 11 sites it was not observable.
- At six research sites the main area(s) for nurses to meet, review care and write up clinical records were fully observable; in nine sites, partly observable; in one site, not observable.
- In 14 research sites the shift coordinator's base was fully observable; in two sites, partly observable.
- In eight research sites (all EDs), some interactions with patients were partly or fully observable; in eight sites (all DUs), none or almost none were.

These differences presented some challenges to observers, though it was possible to meet these by sensitisation to different sources of evidence. For example, where we could not observe handovers directly, we listened carefully to discussions about patients for evidence that staff were or were not sufficiently informed. Where computer screens had replaced whiteboards, we carefully observed how staff used them (e.g. we did not need to be able to read patient details to assess staff interactions and decision-making while looking at the screen). Furthermore, greater visibility was not always an advantage; in site 15, doctors worked primarily at the central hub, but the amount of activity meant that conversations though visible were often inaudible.

The organisation and work factors scores capture some facets of the organisational culture in each of the research sites. However, some important aspects of safety culture could not be observed, for example:

- the existence and effectiveness of appropriate systems for ensuring CPD
- the existence and effectiveness of appropriate systems for reporting and acting on errors and 'near-misses'
- routine checking and maintenance of equipment (which is often periodic and infrequent)
- the availability of information to support evidence-based clinical decision-making where this was in computerised rather than in paper form.

Though in some sites, these might, by chance, be observed, or mentioned in passing by staff, we could not assume that no data reliably indicated an organisational deficit.

### **Recruiting service user and service provider coresearchers**

In addition to the research fellows, other observers were recruited where possible: a service user observer, a midwife, a midwifery lecturer and an A&E lecturer-practitioner visited some sites. These coresearchers made observations on one day at each site they visited. Details of observers at each research site can be found in *Chapter 6* (see *Table 26*), which shows that a service user observer could be present at only three research sites. This highlights that the intention of working with service user coresearchers at each research site was far from realised. At the first four sites, this was because we had not yet succeeded in recruiting service user coresearchers. We attempted to do this by approaching NHS service user panels in London and using the researchers' personal networks of potential service user coresearchers. We explained the study and the funding available to compensate coresearchers for their time and expenses. Two potential service user coresearchers from an NHS service user panel expressed an interest in joining the study. However, one did not respond to communications after an initial meeting, while the

**TABLE 7** Variations between what could be observed at different sites

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Whiteboard use observable	Yes	Yes	No <sup>a</sup>	Yes	Yes	Yes	No <sup>a</sup>	Partly	No <sup>a</sup>	No <sup>a</sup>	No <sup>a</sup>	Yes	Yes	Partly	Partly	Yes
Handovers observable	Partly	Yes	Partly	Yes	Partly	Yes	Partly	Partly (2)	Partly (1)	Partly	Partly	Yes	No	Partly (2)	Partly (3)	Partly (4)
Main medical area observable	Partly	No	No	No	Partly (1)	No	No	No	No	No	Partly	No	Partly	No	Yes	No
Main nurse area observable	Partly	Yes	No	Yes	Yes	Partly	Partly	Partly	Partly	Partly	Partly	Yes	Yes	Yes	Partly (3)	Partly (5)
Shift coordinator base observable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Partly (1)	Yes
Some interactions with patients observable	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

<sup>a</sup> Whiteboard function on computer screens.

1, Visible but not audible; 2, audible not visible; 3, partly visible, partly audible; 4, handover at whiteboard visible and audible, but followed by physical ward round (we did not join as this site was very worried about patient confidentiality); 5, 'top desk' at other end of ward not visible or audible.

bureaucratic process of securing research governance approvals separately for each of the study hospitals deterred the other, who withdrew.

Efforts to recruit service user coresearchers recommenced, and one (DP) was recruited before the data collection began at hospital 3 (research sites 5 and 6). Supported by the project research fellows, DP completed the (different) forms required by subsequent research sites and provided the required references. This allowed CRB screening, occupational health screening and any further local procedures antecedent to each hospital issuing permission for DP to observe. At some hospitals this process took many weeks and permission could not be secured before completion of that phase of data collection. Consequently, the service user co-observer was able to work only at research sites 6, 15 and 16.

Identifying potential service provider co-observers was straightforward, although their clinical and teaching duties often meant that they were unavailable to accompany the research fellows (sites preferred to be introduced to additional observers by researchers they already knew). Service provider co-observers worked at eight research sites: 1, 6, and 11–16. They also provided a clinical perspective during discussions of observations made by research fellows at all sites.

### **Quantifying semistructured observations**

A number of ways of grading observations were trialled and discussed (see *Chapter 9, Tensions and challenges in quantifying observational data*). The scoring system around which consensus emerged is summarised in *Appendix 6*. The notes of observations were analysed and grouped in eight categories, and then the data within each category were scored as follows:

- 0 consistent lack of features of a safety culture
- 1 frequent lack of features of a safety culture
- 2 frequent presence of features of a safety culture
- 3 consistent presence of features of a safety culture.

It did not prove possible to make confident, consistent decisions with greater differentiation than this four-point scale.

Grouping observations into categories before scoring differed from the approach advocated by others,<sup>199,200</sup> who emphasise the scoring of constituent items rather than overarching domains. Trials of the item-based approach did not work well in this study, which was seeking a quantifiable *holistic* assessment. Following independent trials of item-based scoring by research fellows, debates to achieve agreed final scores were protracted but did not yield results in which the researchers had confidence; at this point, the eight categories were decided.

## **Strand C: criterion-based audit of clinical notes**

### **Engaging non-clinical auditors**

We intended to engage trust-based audit clerks (or equivalent non-clinical personnel) to extract anonymous data from patient records (see *Chapter 3, Strand C: criterion-based assessment of quality of care*), reimbursing the trust for associated costs. Pre-study negotiations indicated that staff members might be interested in carrying out the work as overtime. During the recruitment phase for each site the need for a local auditor was highlighted and discussed. One site, which declined to participate, felt that staff would not be available for this work. All recruited sites indicated confidence that local auditors could be identified to support the study: either within their contracted working hours (with reimbursement for the trust) or as paid overtime. For example, one trust used this work to extend the employment of an administrator whose previous

project was finishing; another viewed the audit work as suitable 'light duties' for a staff member returning from a period of sickness. However, the confidence of other trust key contacts proved unfounded. The study coincided with a period of increasing financial pressures in health care: staff workloads were rising and vacant posts were being scrutinised with a view to cost saving. Some individuals whom senior staff expected to undertake this study's audit work were unable to do so within contracted hours and unwilling to undertake audit work as paid overtime. Some could begin the audit work only after a substantial delay to complete other work, or conducted the audit work over an extended period at the margins of their main roles. Researchers negotiating the recruitment of study sites could anticipate problems with staffing audit work, ask pertinent questions and encourage contingency planning, but could not really know the local situation until the negotiation of data collection with nominated individuals.

At the first research site, a member of the administrative staff expressed an interest, though in the event was not available. At the second, the clinical and central administrative departments were unable to identify any staff for this role. Two research fellows (SA and MR, both non-clinical health services researchers) had already been awarded honorary contracts with the trust for the research, and were permitted to carry out the audit. The researchers found the audit process very complex and time-consuming owing to limitations in administrative systems (see *Chapter 9, Limitations of administrative systems*) but also highly informative. Consequently, they chose to continue conducting audits at the next six sites, after obtaining all necessary permissions and honorary contracts: they checked each other's work, sensitised by the process of checking described in *Chapter 3, Audit data collection*. This direct engagement proved extremely valuable in informing the training and support of local audit clerks at later research sites. Furthermore, it can be argued that non-local auditors are ethically preferable as they are very unlikely to know patients whose notes they are scrutinising.

The study coincided with a period during which National Information Governance Board (NIGB) was established. One hospital required the study team to apply to the NIGB for a ruling on the acceptability of research fellows extracting data from clinical records. The NIGB ruled that individual patient consent would be required, unless members of the care team carried out the audit. It is a moot point whether trust-based audit clerks are members of the care team, in which case this NIGB ruling appears to rule out a wide range of quality improvement audit work currently conducted by non-clinical auditors employed by NHS trusts.

Seeking individual patient consent was impractical for this methodological study for a variety of design and resource reasons. We therefore asked the eight remaining research sites and associated trust administrative departments to identify non-clinical staff to extract anonymous audit data. Staff were not always able to begin work immediately or to devote much time to the work, and this led to delays of several months, necessitating an extension to this study (see *Chapter 1, Funding for this study*), and severely reducing the time available for analysing audit results and the comparisons that lie at the heart of this methodological study (see *Chapter 8*). Eventually, all but three of the remaining research sites identified non-clinical staff who could undertake data collection for this study, either within their normal working week or as paid overtime. Despite considerable efforts, research site 12 was unable to identify a non-clinical auditor and we accepted data extraction by a health-care assistant. At sites 13 and 14, no trust-based auditors could be found. The Research Office offered to issue the research fellows with honorary contracts in order that they could be regarded as trust staff, and thus allowed to conduct audits.

Each trust-based auditor was visited by a research fellow (SA) to discuss the data extraction required and varying local practicalities for identifying 50 consecutive cases for each condition, ordering clinical notes and identifying local conventions with respect to recording the data required by our audit (e.g. in some research sites a mixture of paper-based and electronic records

had to be consulted). The use of the data extraction sheets was explained. SA provided ongoing support by telephone or e-mail until each audit was complete. This training and support helped to ensure the quality, completeness and consistency of audit data extraction by different auditors across multiple sites. It would have been difficult to provide if the research fellow had not had first-hand experience of auditing patient notes.

### Cases audited

The study aimed to audit 400 cases for each selected condition: taking 50 consecutive cases from each emergency department or delivery unit ( $50 \times 8 = 400$ ), working backwards for up to 6 months from the date the researchers became noticeable to the whole clinical team by commencing observations. Inclusion and exclusion criteria for the audited cases were listed in *Table 1*. *Table 8a* and *b* shows the degree to which the audit target was achieved for each condition.

The main reason for shortfalls was human error (miscounting audited records) or the discovery during university-based data entry for analysis that the data for a particular case, which had been extracted anonymously from hospital records, were in some way too flawed to include. In a very small number of cases, the same record had been audited twice. As *Table 8a* shows, the audit of FNoF cases at site 1 failed to identify sufficient cases within the 6-month period (see *Chapter 3, Audit data collection*). The audited number of MSL cases at research site 14 fell short of the target because this small DU experienced only 16 cases that were eligible for inclusion in study during the relevant 6-month period and for which case notes could be found.

### Case note data quality

Data relevant to our needs were frequently missing from the case records. Some records that were present were self-evidently unreliable (illegible, or inconsistent, particularly with respect to timings). Site staff at a number of sites told us that it is normal practice when auditing their own service to exclude notes in which relevant data are missing or unreliable. This is clearly appropriate where the object of the exercise is to begin the audit cycle of service improvement: inclusion of records with missing data would weaken the evidence base for service improvement.

**TABLE 8a** Cases audited for each condition: emergency departments

	Site								Totals
	1	3	5	7	9	11	13	15	
ACS	45	48	50	50	44	49	47	50	383
ASA	39	51	50	50	50	50	50	49	389
FNoF	31 <sup>a</sup>	50	51	50	55	48	49	50	384

<sup>a</sup> This large department failed to identify sufficient cases due to inadequate labelling in admissions and discharge databases (see *Chapter 9, Limitations of administrative systems*).

**TABLE 8b** Cases audited for each condition: delivery units

	Site								Totals
	2	4	6	8	10	12	14	16	
ND	50	51	50	48	50	47	50	49	395
ECS	50	50	50	50	50	49	50	50	399
MSL	50	50	50	50	48	45	16 <sup>a</sup>	50	359

<sup>a</sup> This small unit only had 16 eligible cases during the audited period.

However, it is important not to lose sight of the proportion of case notes with missing data: an indicator of quality in its own right. Because ours was a comparative methodological study, and we were studying the validity and practicability of case note audit as a method, we included notes with missing data. The assumption that recording and activity correlate reasonably well is intrinsic to this method. Ideally, we would have included 50 cases for which data were complete while also noting details of incomplete cases: time and resources did not allow this, however.

At feedback sessions, clinicians often argued that unrecorded observations and interventions would, nevertheless, generally have been carried out. Such explanations have been offered for many years,<sup>201,202</sup> and, though they may be true, they are hard to substantiate retrospectively.

Clinicians also pointed out that, in some cases, monitoring would be continuous, with only occasional notes being made of readings: for example, oxygen saturation may be measured continuously in asthma patients, most of which measurements would be unrecorded. Staff said also that some observations would be recorded only if problematic; for example, midwives at several DUs told us that they would record show and blood at initial assessment only if there were grounds for concern.

It also became clear during feedback that not all sites aimed to meet some standards. For example, at site 5, nurses do not prescribe to ACS and FNoF patients, and it is unlikely that patients will see a doctor 'immediately' (which we interpreted as within 10 minutes) so drugs cannot be prescribed straight away. At site 10, midwives monitor fetal heart rate in the second stage of labour only after each contraction, rather than every 5 minutes as recommended in national standards. At site 12, midwives do meconium observations 3-hourly rather than 2-hourly.

Staff also queried the internal logic of some standards; for example, one ND standard is to observe maternal temperature every 4 hours during the second stage of labour; but if this stage lasts 4 hours, this would not be a normal delivery.

### **Unpredictable resource requirements**

The audit strand of this study was very time-consuming and, therefore, resource intensive. The main problem for those planning future studies is that resource requirements varied unpredictably between research sites. At some sites it was very difficult for local administrators or clinicians to identify which clinical records should be audited owing to deficiencies in admissions or discharge databases. In addition, ordering clinical records was more straightforward at some sites than at others. The administrators undertaking this work did so at the margins of their main role and frequently found it difficult to allocate time to seeking help to resolve difficulties, causing delay in the audit data collection. Although, as part of the process of negotiating recruitment of research sites, researchers tried to understand as much as possible about local contexts that would affect audit work, site-based informants rarely anticipated the difficulties auditors subsequently encountered. In most cases site-based informants were clinicians who frequently used and contributed to clinical records but, quite reasonably, lacked detailed knowledge of some administrative processes relating to clinical records. Advice from central audit departments was normally given by managers. It was difficult for researchers to interact with front-line auditors prior to data collection.

Clinical records at some sites were predominantly electronic and could be audited quickly; at other sites they were wholly paper based. For some conditions at some sites, the case notes that had to be accessed spanned two or three separate systems. At some sites, auditors had to retrieve and re-shelve their own notes from medical records departments; at others, they were retrieved

for them. These factors greatly alter the time required for audits but are difficult to predict before an audit is under way, even though researchers always initiated pre-audit discussion of these matters with local clinicians and administrators. Trust-based auditors from different trusts needed strikingly different lengths of time to conduct audits of the same conditions. We cannot know the degree to which these differences reflected the efficiency and experience of the auditors, other workload demands faced by auditors, or site-specific challenges.

These unpredictable resource requirements contrast with more predictable resource differences relating to the nature of the audit standards. The standards we audited varied in their ease of extraction. For example, those relating to MSL required data extraction from both the mothers' notes and the babies' notes, which would be stored separately. It was necessary to audit the mothers' notes first to identify the required baby notes. Another example is that certain clinical observations occur once and are easy to identify within case notes: this was the case with several ED audit standards. In contrast, the midwifery notes made during normal labour can form an extended narrative in which many clinical observations are embedded: data extraction requires extensive reading. These are largely predictable variations in resource requirements that can be built into study designs.

## Timely feedback of findings for local use

In line with objective 1 (see *Chapter 1, aims and objectives*) and in order to provide a potential benefit in exchange for sites' generosity in allowing us access, we offered to feed back site-specific findings at each site. All sites received a written summary of the site-specific findings and a researcher offered to present these to any interested staff groups. This offer was accepted at 10 sites (five of each type). The feedback enabled sites to consider and act on findings if they wished. Providing feedback to research sites also acted as an opportunity for member checking,<sup>203</sup> which helped ensure the veracity of findings. For example, at one site early in the study, where we had been told that all data were recorded electronically, it became evident at feedback that this was not in fact the case, and we had then to consult paper records. Feedback also enabled a more complete contextualisation of the data, for example in a site where missing observations of mothers in labour were the result not of negligence but of a clinical decision; however, this information could not alter the score, as national rather than local standards were being used.

It is not known why the remaining six sites declined: they did not give reasons but simply failed to respond to our offers. In some cases, informed guesses can be made as to why feedback was not sought. At site 8, the audit took particularly long to complete (because records had to be ordered from off-site with frequent delays), and by the time data were collected and analysed, all senior staff had changed: they may have felt that the data lacked salience at that point. Site 9 had been a reluctant participant in the study, having been instructed to take part by a senior manager, and so it was less likely that opportunities for feedback would be made.

Although it is regrettable that member checking was not possible in these sites, it is unlikely that *major* misunderstandings about recording, such as that concerning electronic records, will have occurred, as we learnt to check at an early stage any unexpected or inexplicable patterns of audit data. From the start of the study, we sought explanations for observed data that seemed inexplicable while still on site.

Those attending feedback sessions were typically most interested in the audit scores, as these had a clinical focus and often related to audit work of their own. Where our findings were compared with the results of local audits, they were always found to be congruent. Sometimes actions had

already been taken to address identified deficits (e.g. site 3 had improved pain relief). In one site (site 12), most discussion arose about the survey findings, as these were surprisingly negative given the other results in our study and the DU's self-perception as a centre of excellence.

## Summary

Conducting this multistrand methodological study yielded process-based insights and highlighted a number of challenges with the contrasting approaches to data collection. The questionnaire surveys yielded variable response rates, while observation-based assessment was incomplete in ways that varied between sites. Criterion-based case notes audit provided to be resource intensive and administratively challenging, in ways that could not be anticipated prior to fieldwork commencing.



## Chapter 5

# Strand A: questionnaire survey of staff perceptions of safety and organisational climates

The findings in this chapter relate to objective 2: to use questionnaires to obtain quantitative assessments of the organisational and safety climate at each site. In some sections it has also been possible to address objective 8, which concerned exploring responses by professional group and level of management responsibility. The chapter will conclude with testing hypothesis 2a (see *Chapter 1, Hypotheses*): There will be a strong correlation between *questionnaire*-based evaluations of organisational and safety climate.

The following section summarises demographic and role-related characteristics for the 531 survey respondents, and *Support for training, learning and development* summarises participation in trust-supported CPD. The next section, *Organisational climate*, summarises responses to questions that were intended to elicit perceptions of certain aspects of organisational climate, shows which site-level and individual characteristics influence responses to these questions and develops a summary survey-based assessment of organisational climate. *Safety climate and team factors* mirrors *Organisational climate* for the questions that were intended to elicit perceptions of certain aspects of safety climate. In *Hypothesis 2a: comparing survey-based organisational and safety climate scores*, the summary survey-based assessments of organisational and safety climate are compared: there is strong correlation and reasonably good agreement between the two assessments.

### Demographics and role-related characteristics

The staff survey (see *Appendix 3*) collected data on several demographic and role-related factors, which will be investigated as mediators of perceptions of organisational climate and safety climate in *Factors influencing the indicators of organisational climate* and *Factors influencing the indicators of safety climate*.

Eighty per cent of respondents were female. This mirrored the NHS overall,<sup>204</sup> in which 81% of staff are female (though medical staff are excluded from this data set), although the match was less perfect when we drilled down into individual staff categories. One notable difference was in the category 'qualified clinical staff' (non-medical): 92% of this study's respondents were female, compared with 84% in the wider NHS. This is due to selecting delivery units for half this study's research sites: nearly all midwives are female (100% in our sample).

The age profile of respondents was as follows:

- 20% (108) aged 16–30 years
- 30% (158) aged 31–40 years
- 28% (146) aged 41–50 years
- 22% (114) aged 51–65 years and
- no respondents identified themselves as > 65 years.

Thus, half of this study's respondents were aged  $\leq 40$  years and 78% were aged  $\leq 50$  years. National statistics<sup>204</sup> show that 40% of NHS non-medical staff are aged  $\leq 39$  years and 70% are  $\leq 49$  years, suggesting that the age profile in this study may be a little younger than in the wider NHS.

Many of the respondents had worked for the NHS trust in which the study site was located for several years. The distribution was as follows:

- 9% (50) < 1 year
- 13% (67) 1–2 years
- 23% (124) 3–5 years
- 24% (127) 6–10 years
- 11% (58) 11–15 years and
- 19% (101) > 15 years.

For the variables reported above, the proportion of missing data was < 1%, and for the other variables (contracted hours, management responsibility and profession) it was < 2%. Most respondents (78%) were contracted to work  $\geq 30$  hours per week. A little over one-third of respondents (37%,  $n = 197$ ) indicated that they managed other employees within the trust. We do not have data on how many people each managed or an estimate of seniority within the organisational hierarchy. Most respondents (322, 61%) were nurses in the ED research sites or midwives in the DU research sites; 92 (17%) were doctors. The remaining staff were placed in two groups, those providing direct care (mostly health-care assistants or midwifery assistants, but also including allied health professionals and social workers) and those whose roles did not include giving direct care (mainly administrative and clerical staff but also including general managers and porters). There were 51 respondents (10%) in the former group and 56 respondents (11%) in the latter. Our respondent group was disproportionately composed of clinical staff when compared with the NHS non-medical workforce as a whole.<sup>204</sup>

People were asked to indicate their ethnic group by choosing among standard categories used by the NHS, including in the annual staff survey. Seventeen respondents (3%) selected the 'prefer not to say' option, and seven (1%) skipped the question. A substantial majority of respondents identified themselves as White British (389, 73%). The next most commonly selected categories were Asian or Asian British (36 respondents, 7%), White not-British (34, 6%) and Black or Black British (28, 5%). These proportions are similar to those found in the NHS census of non-medical staff.<sup>204</sup> As the 15 categories, excluding White British, were generally sparsely populated, they have been grouped for analyses in subsequent sections, using the NHS group headings shown within the questionnaire (see *Appendix 3*).

## Support for training, learning and development

Respondents were asked: 'In the past 12 months, have you taken part in any of the following types of training, learning or development, paid for or provided by your Trust?' The results are shown in *Table 9*. The modal response ( $n = 147$ , 28%) was participation in three of the categories of CPD shown in the table. Very few respondents identified unprompted forms of training, learning or development, but examples included 'discussing with colleagues', 'make educational videos', 'organising away days', 'teaching others' and 'visiting other departments'. Twenty respondents (4%) skipped this question and 25 respondents (5%) indicated that they had not participated in trust-supported training, learning or development during the past year.

**TABLE 9** Proportion of respondents receiving trust-supported CPD in past year

(a) Taught courses (internal or external)	77%
(b) Any supervised on-the-job training	40%
(c) Having a mentor	26%
(d) Shadowing someone	23%
(e) E-learning/online training	59%
(f) Keeping up to date with development in your type of work (e.g. by reading books or journals, or by attending seminars or workshops)	74%

## Organisational climate

This section summarises responses to the survey questions, described in *Chapter 3, Developing the organisational safety climate questionnaire*, which were intended to elicit perceptions of certain aspects of organisational climate. First, responses to the six-item organisation scale were normally distributed and varied significantly between hospitals ( $F = 10.022$ ;  $df 7, 501$ ;  $p < 0.0001$ ). Research sites 3–6 were located in the hospitals with the highest organisation scores, mean 3.31, whereas research sites 11 and 12 were located in the hospital with the lowest mean score for organisation, 2.50 (*Table 10*). Higher scores are indicative of a more positive organisational climate. Respondents at research site 3 had the most positive views of their organisation and staff at research site 9 the least positive views.

Most respondents (462, 87%) said they would know how to report concerns about negligence or wrongdoing. This ranged from 67% at research site 2 to 100% at site 8. Fewer respondents (374, 70%) thought that a confidential reporting system was available [range 52% (site 2) to 86% (site 8)]. Overall, 22% (118) were unsure and 6% of respondents (29) thought that there was no confidential reporting system. The latter group included 18% of respondents at research site 7 and 20% at site 11, whereas levels of ‘no’ responses were lower elsewhere. Nearly all respondents (495, 93%) said that they knew how to report errors, near-misses or incidents. This ranged from 78% at site 2 to 97% at research site 9. The answers to these questions were significantly correlated with one another (*Table 11*), but levels of correlation varied from very low (0.102) to, at best, moderate (0.487). It is likely that the lowest correlation proved statistically significant ( $p = 0.021$ ) only because this is a fairly large data set.

Responses to the seven-item error wisdom scale were normally distributed and varied significantly between hospitals ( $F = 5.593$ ;  $df 7, 509$ ;  $p < 0.0001$ ; *Table 12*). In *Table 12* higher mean scores correspond to higher levels of perceived error wisdom, which is considered to be a facet of positive organisational culture aligned with supporting patient safety.<sup>55,57</sup> Research site 6 was perceived to have the highest level of error wisdom (mean score 3.89 out of 5) and site 9 the lowest level (2.84).

The Pearson’s correlation between the organisation scale scores and error wisdom scores was moderate,  $r = 0.579$  ( $p < 0.0005$ ): scores on these two scales tend to increase and decrease together and the value of one accounts for 36% of the variation in values of the other ( $r^2 = 0.356$ ). *Tables 10* and *12* show that, while moving in step, scores for error wisdom were a little higher (mean 3.39 out of 5) than scores on the organisation scale (mean 2.97 out of 5) and a little less variable (SE of error wisdom, all sites, 0.028; SE organisation, all sites, 0.036).

TABLE 10 Organisation scale

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	2.838	0.129	34	2
2	2.842	0.199	19	8
3	3.423	0.083	82	1
4	3.054	0.108	37	1
5	3.245	0.112	34	1
6	3.389	0.094	27	1
7	2.833	0.192	17	0
8	3.381	0.205	7	0
9	2.394	0.128	36	1
10	2.989	0.105	30	0
11	2.486	0.204	24	1
12	2.519	0.185	27	0
13	3.032	0.174	21	0
14	3.027	0.167	25	0
15	2.842	0.130	40	0
16	2.806	0.095	55	0
All sites	2.968	0.036	515	16

TABLE 11 Spearman's rho correlations for questions concerning whistle-blowing and reporting errors, near-misses and incidents

	Believes there is a confidential system for whistle-blowing	Knows how to report errors, near-misses and incidents
<i>Knows how to whistle-blow</i>		
Correlation ( $\rho$ )	0.362	0.487
Two-tailed <i>p</i> -value	<0.0001	<0.0001
<i>n</i>	515	508
<i>Believes there is a confidential system for whistle-blowing</i>		
Correlation ( $\rho$ )		0.102
Two-tailed <i>p</i> -value		0.021
<i>n</i>		511

### Factors influencing the indicators of organisational climate

Multilevel modelling of scores on the organisation scale (individual responses nested within research sites nested within hospitals, as described in *Chapter 3, Multilevel modelling*) showed that one site-level variable (the number of people included in the staff survey) and two individual-level variables (whether a manager and the number of years working in the trust) mediated organisation scores (*Table 13*). The model included 510 cases and estimated the underlying level of organisation scores as 2.672 (SE 0.119). The estimated parameters in *Table 13* indicate that:

- Organisation scores were marginally higher in departments that distributed larger numbers of questionnaires, but only 0.0002 (estimated parameter  $\times$  SD) of organisation scores higher for every additional 62 (1 SD) people included in the staff survey so this statistically significant result has no practical significance.

TABLE 12 Error wisdom

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	3.078	0.088	35	1
2	3.481	0.148	22	5
3	3.561	0.057	82	1
4	3.327	0.097	38	0
5	3.473	0.091	32	3
6	3.888	0.094	28	0
7	3.179	0.143	16	1
8	3.796	0.187	7	0
9	2.837	0.097	35	2
10	3.631	0.096	29	1
11	3.091	0.149	25	0
12	3.058	0.223	27	0
13	3.429	0.149	21	0
14	3.456	0.103	25	0
15	3.307	0.080	40	0
16	3.555	0.060	55	0
All sites	3.386	0.028	517	14

- Respondents who managed other staff returned organisation scores that were a little higher (0.033 higher) than the scores from those who did not manage others.
- Compared with respondents in the modal category of 6–10 years' employment with their trust, respondents who had worked at the trust for <6 years or >15 years returned noticeably higher organisation scores [between 1.577 (1–2 years) and 0.552 (>15 years) higher].

The multilevel model partitioned the remaining unexplained variation in organisation scores follows:

- 11% at hospital level (estimated variance 0.067, SE 0.039,  $\chi^2$  2.959, 1 df,  $p=0.085$ )
- 89% with respondents within sites (estimated variance 0.562, SE 0.035,  $\chi^2$  251.0, 1 df,  $p<0.0001$ ).

There is significant residual variance in the organisation scores that cannot be accounted for by respondents' demographic and role-related factors. This variation may relate to unmeasured individual characteristics but we regard it as mainly identifying site-level differences in perceptions of organisational climate.

A similar analysis was performed for the error wisdom scale. The multilevel model used 475 cases and estimated the underlying level of error wisdom scores at 2.805 (SE 0.106). The estimated parameters in *Table 14* show that eight variables mediated error wisdom scores, as follows:

- Compared with staff in EDs, DU staff perceived higher levels of organisational error wisdom, 0.038 higher.
- As in previous analyses, there was a very small effect associated with the number of people included in the staff survey, which was too small to have practical significance, Error wisdom scores increased by 0.0003 for each additional 62 questionnaires distributed.
- At the individual level, age was a mediating factor: compared with the modal reference group aged 31–40 years, older respondents rated their trust's error wisdom more highly:

**TABLE 13** Parameter estimates for fixed factors in the best fitting multilevel model of organisation scores

Fixed factors	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
<i>Site level</i>		
Number of people sent survey <sup>a</sup>	0.002 (0.001)	5.380 ( $p=0.020$ )
<i>Respondent level</i>		
If a manager	0.208 (0.074)	7.901 ( $p=0.005$ )
Years worked for trust <sup>b</sup> :		
< 1 year	0.390 (0.131)	8.832 ( $p=0.003$ )
1–2 years	0.425 (0.117)	13.253 ( $p=0.0003$ )
3–5 years	0.273 (0.097)	7.937 ( $p=0.005$ )
11–15 years	0.113 (0.104)	0.841 ( $p=0.359$ )
> 15 years	0.174 (0.111)	4.640 ( $p=0.031$ )

a Centred on mean.

b Reference group 6–10 years.

scores were 0.020 higher for the 41- to 50-year age group and 0.017 higher from those aged 51–65 years.

- People who had a mentor in the previous 12 months returned higher error wisdom scores, 0.018 higher.
- Managers rated their trust's error wisdom more highly than non-managers, 0.017 higher.
- People who received updating or online training or who shadowed someone at some point during the previous year also rated their trust's error wisdom more highly; scores were increased by 0.017, 0.017 and 0.016, respectively.

The multilevel model partitioned residual variation as follows:

- 10% at hospital level (estimated variance 0.038, SE 0.022,  $\chi^2$  2.811, 1 df,  $p=0.094$ )
- 90% at the level of respondents nested within sites (estimated variance 0.348, SE 0.023,  $\chi^2$  233.6,  $p<0.0001$ ).

There remains significant residual variance in error wisdom scores, not accounted for by respondents' demographic and role-related factors. Although the influence of unmeasured factors cannot be ruled out, this will be taken to signal significant variation in error wisdom scores between research sites; reflecting underlying variation in organisational culture.

### Combining indicators of organisational climate

Responses to staff survey questions intended to elicit perceptions of organisational climate (described in *Chapter 3, Developing the organisational and safety climate questionnaire*) were combined as described in *Chapter 3, Combining indicators of organisational climate*. The mean and SD of the summary survey-based organisational climate scores were 0 and 0.752, respectively. In *Table 15* the research sites have been arranged so that the highest scoring sites appear at the top: these are the sites where staff perceptions of organisational climate were most positive. It was noticeable that there were more delivery units than emergency departments in the positively scoring sites, where perceptions of organisational climate were above average, but analysis of variance showed that the difference between scores for delivery units and emergency departments was not significant ( $F=2.115$ , df 1,14,  $p=0.168$ ). The scores in *Table 15* also show that staff in different clinical services within the same hospital can have different perceptions of

**TABLE 14** Estimated parameters fixed for factors in multilevel model of error wisdom scores

Fixed factors	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
<b>Site level</b>		
Service <sup>a</sup>	0.355 (0.073)	23.528 ( $p < 0.0001$ )
Number sent staff survey <sup>b</sup>	0.003 (0.001)	13.412 ( $p = 0.0003$ )
<b>Respondent level</b>		
Age (years) <sup>c</sup>		
16–30	0.053 (0.080)	0.0435 ( $p = 0.835$ )
41–50	0.185 (0.073)	6.507 ( $p = 0.011$ )
51–65	0.157 (0.079)	3.904 ( $p = 0.048$ )
Had mentor	0.167 (0.070)	5.682 ( $p = 0.017$ )
Is a manager	0.162 (0.060)	7.334 ( $p = 0.007$ )
Updating	0.162 (0.069)	5.436 ( $p = 0.020$ )
Online training	0.160 (0.061)	6.839 ( $p = 0.009$ )
Shadowed someone	0.155 (0.072)	4.680 ( $p = 0.031$ )

a Reference group ED so parameter relates to DU.

b Centred on mean.

c Reference group 31–40 years.

the organisational climate that they arguably share. The widest disparity occurred at hospital 4 and the smallest difference at hospital 1.

## Safety climate and team factors

Responses to the three-item overload scale were normally distributed and varied significantly between hospitals ( $F = 3.117$ ,  $df 7, 497$ ,  $p = 0.003$ ). Respondents from research site 3 reported the lowest levels of overload (mean 3.21 out of five, SE 0.112), while respondents from site 12 reported the highest levels (mean 2.40, SE 0.195; *Table 16*).

Responses to the five-point single item 'relationships at work are strained' were normally distributed and varied between hospitals ( $F = 4.496$ ,  $df 7, 519$ ,  $p = 0.0001$ ). The overall mean was 3.12 and respondents from site 6 reported least strained working relationships (mean 3.68, SE 0.146), whereas respondents from site 9 reported most strained working relationships (mean 2.32, SE 0.190).

Responses to the five-item scale evaluating immediate line management were normally distributed and varied significantly between hospitals ( $F = 8.613$ ,  $df 5, 510$ ,  $p < 0.0001$ ). Respondents from site 5 evaluated their line management most positively (mean 3.86 out of 5, SE 0.150), while those from site 9 evaluated it least positively (mean 2.51, SE 0.150; *Table 17*).

Ninety per cent of respondents (476) reported that they had to work closely with other team members to achieve the team's objectives. However, questionnaires distributed at research sites 1 and 2 contained a printing error in the boxes for responses to this question; when these were removed from the analysis the proportion rose to 95%. On the other hand, only 45% (237) reported that their team met regularly to discuss its effectiveness and how this could be improved.

**TABLE 15** Summary survey-based organisational climate (z-scores)

Hospital (service)	Site	Summary organisational climate
H3 (DU)	6	1.333
H4 (DU)	8	1.014
H2 (ED)	3	0.714
H7 (DU)	14	0.449
H5 (DU)	10	0.418
H2 (DU)	4	0.355
H7 (ED)	13	0.304
H8 (DU)	16	0.230
H3 (ED)	5	0.189
H8 (ED)	15	-0.031
H1 (ED)	1	-0.584
H4 (ED)	7	-0.623
H1 (DU)	2	-0.682
H6 (ED)	11	-0.787
H6 (DU)	12	-1.004
H5 (ED)	9	-1.293

**TABLE 16** Overload scores (reversed so high scores represent low levels of overload)

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	2.587	0.148	25	11
2	2.700	0.151	20	7
3	3.211	0.112	79	4
4	2.561	0.136	38	0
5	2.843	0.171	34	1
6	2.786	0.166	28	0
7	2.792	0.225	16	1
8	2.714	0.267	7	0
9	2.613	0.129	37	0
10	2.689	0.160	30	0
11	2.600	0.231	25	0
12	2.397	0.195	26	1
13	3.000	0.202	21	0
14	2.800	0.160	25	0
15	2.573	0.131	39	1
16	2.470	0.131	55	0
All sites	2.736	0.041	505	26

Individual responses to the six-item scale termed teamwork climate by Sexton and colleagues,<sup>25</sup> and their seven-item scale termed safety climate, were close to being normally distributed with some evidence of negative skew due to high levels of agreement with the constituent items (teamwork climate: mean 3.992, SE 0.038; safety climate: 3.793, SE 0.036). A range of common transformations was explored with a view to increasing normality,<sup>205</sup> but only ranking gave a slightly better fit to the normal distribution. This gain was not considered to outweigh the additional complexity of interpretation that this transformation would bring. Later analyses in *Chapter 8* will use average scores for each site rather than individual scores and, as expected from the central limit theorem (see, for example, Lane<sup>206</sup>), these are normally distributed.



**TABLE 17** Evaluations of immediate manager

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	3.000	0.144	34	2
2	3.009	0.139	22	5
3	3.700	0.096	82	1
4	3.137	0.154	38	0
5	3.865	0.150	34	1
6	3.864	0.120	28	0
7	3.047	0.251	17	0
8	3.543	0.301	7	0
9	2.514	0.150	37	0
10	3.476	0.100	29	1
11	2.844	0.280	23	2
12	2.948	0.229	27	0
13	3.829	0.192	21	0
14	3.336	0.177	25	0
15	3.410	0.152	39	1
16	3.280	0.105	55	0
All sites	3.326	0.042	518	13

Each site's average score for the teamwork climate scale developed in Texas, USA, is shown in *Table 18*. Sites 5 and 6, at the same hospital, returned the highest teamwork climate scores (mean 4.52 out of 5, SE 0.072, and mean 4.42, SE 0.113, respectively), whereas sites 11 and 12, at the same hospital, returned the lowest teamwork climate scores (mean 3.42, SE 0.210, and mean 3.41, SE 0.303).

Average scores for the safety climate scale developed in the USA are shown in *Table 19*. Site 6 returned the highest safety climate scores (mean 4.31 out of 5, SE 0.081), whereas research site 7 returned the lowest safety climate scores (mean 3.14, SE 0.243).

### Comparing indicators of safety climate

Responses to the two British scales, overload and line management, exhibited a low level of positive correlation [Pearson's  $r = 0.285$  ( $n = 497$ ),  $p < 0.0001$ ], but the score on one scale explains only 8% of the variation in scores on the other scale ( $r^2 = 0.081$ ). In contrast, the two scales developed in the USA were strongly correlated [ $r = 0.757$  ( $n = 485$ ),  $p < 0.0005$ ]: evaluations of teamwork climate and safety climate move in step with one another and the score on one explains 57% of the variation in the other ( $r^2 = 0.573$ ). Furthermore, *Tables 18* and *19* show that teamwork climate and safety climate were scored at similar levels (mean 3.99, SE 0.038, and mean 3.79, SE 0.036, respectively).

Correlations between scores on the British overload scale and the American scales were low (0.320 and 0.325,  $p < 0.0001$ ; *Table 20*). To improve readability, entries in the lower triangle of *Table 20* have been removed because they duplicate entries in the upper triangle. Hence, low levels of reported overload are to some degree associated with high scores for teamwork and safety climates, but overload scores only account for 10% of the variation in teamwork climate scores ( $r^2 = 0.102$ ) and for 11% of the variation in safety climate scores ( $r^2 = 0.106$ ). There were stronger correlations between the British line management scale and the American scales (0.570 and 0.600,  $p < 0.0005$ ; see *Table 20*). Positive evaluations of supportive line management are associated with higher scores for teamwork and safety climates. Line management scores account for 32% of the variation in teamwork climate scores ( $r^2 = 0.325$ ) and 36% of the variation in safety climate scores ( $r^2 = 0.036$ ).

**TABLE 18** Teamwork climate scores (Sexton *et al.*<sup>25</sup>)

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	3.801	0.161	31	5
2	3.936	0.159	26	1
3	4.264	0.070	74	9
4	3.995	0.107	35	3
5	4.417	0.113	26	9
6	4.524	0.072	28	0
7	3.464	0.296	14	3
8	4.048	0.259	7	0
9	3.438	0.146	35	2
10	4.244	0.122	30	0
11	3.417	0.210	24	1
12	3.407	0.303	25	2
13	4.095	0.196	21	0
14	4.276	0.122	25	0
15	4.064	0.116	39	1
16	3.929	0.096	55	0
All sites	3.992	0.038	495	36

**TABLE 19** Safety climate scores (Sexton *et al.*<sup>25</sup>)

Site	Mean	Standard error of mean	<i>n</i>	Missing owing to incomplete responses
1	3.597	0.128	34	2
2	3.600	0.165	25	2
3	4.022	0.078	73	10
4	3.927	0.112	37	1
5	4.077	0.119	28	7
6	4.311	0.081	28	0
7	3.143	0.243	15	2
8	3.898	0.311	7	0
9	3.437	0.135	35	2
10	4.010	0.127	30	0
11	3.373	0.199	23	2
12	3.49	0.262	26	1
13	4.109	0.175	21	0
14	3.957	0.146	25	0
15	3.707	0.122	39	1
16	3.662	0.086	55	0
All sites	3.793	0.036	501	30

The reverse-coded single item 'Relationships at work are strained' was moderately correlated with all four scales, in the range 0.428–0.502 ( $p < 0.0005$ ; see *Table 20*): perceiving strained relationships is somewhat linked to perceiving overload and lower evaluations of line management, teamwork and safety climates, accounting for between 18% and 25% of the variation in the scores on these scales (the range for  $r^2$  was 0.183–0.252).

**TABLE 20** Pearson's correlations between indicators of safety climate

		Line management	Sexton teamwork climate	Sexton safety climate	'Relationships at work are strained' (reverse coded)
Overload	$r(n)$	0.285 (497)	0.320 (472)	0.325 (478)	0.428 (499)
Line management	$r(n)$		0.570 (483)	0.600 (490)	0.498 (517)
Sexton teamwork climate	$r(n)$			0.757 (485)	0.502 (486)
Sexton safety climate	$r(n)$				0.427 (493)

$r(n)$ , the first value,  $r$ , shows the correlation, followed by the number of cases in parentheses.

The two-tailed significance for each correlation in the table was  $<0.0001$ , although apparent significance may have been boosted by the large data set.

### Factors influencing the indicators of safety climate

Multilevel modelling of the scales intended to capture aspects of safety climate showed that four respondent-level factors (professional and ethnic groups, years worked for trust and having a mentor at some time in the previous year) mediated scores on the overload scale (Table 21). The model used 472 cases and the estimated underlying level of overload scores at 2.812 (SE 0.172). The estimated parameters in Table 21 show that:

- Compared with the modal reference group ED nurse or DU midwife, respondents whose roles supported care rather than directly provided care (e.g. administrative and clerical staff, non-clinical managers, porters and housekeepers) reported lower levels of overload. The estimated difference in scores was 0.112.
- Compared with respondents in the reference group of 6–10 years working for the trust, which was the modal category and contained the median duration of employment, respondents who had worked for their trust for  $<3$  years reported lower levels of overload. The estimated difference in scores was 0.068 for those employed for 1–2 years and 0.070 for those employed for  $<1$  year.
- Compared with the reference category (and modal response) of White ethnicity (see questionnaire in Appendix 3 for finer-grained composition of ethnic grouping used by the NHS) respondents identifying with any of the categories within the Asian or Asian British ethnic group reported lower levels of overload. The estimated difference in scores was 0.065.
- Respondents who had a mentor reported lower levels of overload than those who did not. The estimated difference in scores was 0.061.

The multilevel model, which included these fixed factors, partitioned variance as follows:

- 3% at hospital level (estimated variance 0.020, SE 0.017,  $\chi^2$  1.394, 1 df,  $p=0.238$ )
- 97% with respondents within research sites (estimated variance 0.734, SE 0.048,  $\chi^2$  232.0, 1 df,  $p<0.0001$ ).

Although there is no discernible hospital-level variation in perceptions of overload, there is significant variation in responses to the overload scale from respondents nested within research sites, which remains after including the mediating effects of professional and ethnic groups, duration of employment, managerial responsibility and engagement with CPD. Perceptions of overload varied between the clinical departments participating in this study.

Multilevel modelling of line management revealed that one site-level variable (response rate) and six respondent-level variables (professional group, being a manager, gender, receiving on-the-job training, having a mentor and updating knowledge or skills scores) mediated scores on this scale.

**TABLE 21** Parameter estimates for fixed factors in multilevel model of overload scores

Fixed factors (respondent level)	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
Professional group <sup>a</sup>		
Doctors	0.064 (0.113)	0.317 ( $p=0.573$ )
Others – giving direct care	0.182 (0.135)	1.805 ( $p=0.179$ )
Others – not giving direct care	0.653 (0.134)	22.680 ( $p<0.0001$ )
Years worked for trust <sup>b</sup>		
< 1	0.407 (0.170)	5.727 ( $p=0.017$ )
1–2	0.395 (0.139)	8.117 ( $p=0.004$ )
3–5	0.100 (0.117)	0.731 ( $p=0.393$ )
11–15	0.239 (0.145)	2.704 ( $p=0.100$ )
> 15	-0.090 (0.121)	0.546 ( $p=0.460$ )
Ethnic group <sup>c</sup>		
Asian or Asian British	0.375 (0.162)	5.322 ( $p=0.021$ )
Black or Black British	0.186 (0.201)	0.859 ( $p=0.354$ )
Mixed background	0.072 (0.359)	0.040 ( $p=0.842$ )
Chinese and any other ethnic group	-0.249 (0.251)	0.986 ( $p=0.321$ )
Had a mentor	0.356 (0.098)	13.333 ( $p=0.0003$ )

a Reference group midwives (DU) or nurses (ED).

b Reference group 6–10.

c Reference group White.

The best-fitting model included 477 cases and the underlying level of line management scores was estimated at 2.496 (SE 0.191). The parameter estimates for mediating variables, listed in Table 22, show that:

- Research sites with above average response rates returned lower line management scores (0.289 lower).
- Compared with the modal reference group of DU midwives and ED nurses, doctors returned higher line management scores (0.081 higher).
- Respondents who had been supported to update skills or knowledge during the past year also returned higher line management scores (0.062 higher).
- Managers and women reported more favourable views of their direct managers, with line management scores 0.054 and 0.050 higher than non-managers and men, respectively.
- People who had received on-the-job training or had a mentor at some point during the past year were also more positive about their line managers, with scores, respectively, 0.048 and 0.046 higher.

The multilevel model partitioned variance as follows:

- 13% at hospital level (estimated parameter 0.108, SE 0.061,  $\chi^2=3.116$ , 1 df,  $p=0.078$ )
- 87% with respondents within sites (estimated parameter 0.725, SE 0.047,  $\chi^2=253.5$ , 1 df,  $p<0.0001$ ).

Most of the variation that remains after including the mediating effects of response rates, professional groups, managerial responsibility, gender and engagement with CPD lies at the level of respondents nested within research sites. Respondents from different research sites perceived the quality of their line management differently.

**TABLE 22** Parameter estimates for fixed factors in multilevel model of line management scores

Fixed factors	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
<b>Site level</b>		
Response rate <sup>a</sup>	-1.514 (0.710)	4.544 ( $p=0.033$ )
<b>Respondent level</b>		
Professional group <sup>b</sup>		
Doctors	0.424 (0.118)	12.797 ( $p=0.0003$ )
Others – giving direct care	-0.046 (0.144)	0.101 ( $p=0.751$ )
Others – not giving direct care	0.165 (0.144)	1.322 ( $p=0.250$ )
Updating	0.336 (0.103)	10.692 ( $p=0.001$ )
Is a manager	0.282 (0.088)	10.238 ( $p=0.001$ )
On-the-job training	0.249 (0.087)	8.268 ( $p=0.004$ )
Gender <sup>c</sup>	0.246 (0.114)	4.662 ( $p=0.031$ )
Had a mentor	0.241 (0.099)	5.907 ( $p=0.015$ )

a Centred on mean.

b Reference group midwives (DU) or nurses (ED).

c Reference group male so parameter relates to females.

Multilevel modelling of the teamwork climate scale developed by Sexton and colleagues<sup>25</sup> showed that one site-level variable (the number of people included in the staff survey) and seven respondent-level variables (professional group, gender, years employment with trust, three types of CPD and being a manager) mediated teamwork climate scores (Table 23). The model used 455 cases and the estimate of the underlying level of teamwork scores was 3.098 (SE 0.176). The parameter estimates show that:

- Compared with the reference group of ED nurses and DU midwives, doctors' teamwork climate scores were 0.077 higher.
- Women and managers had higher teamwork climate scores, 0.056 and 0.036 higher than men and non-managers, respectively.
- Longer than average employment with the trust was also associated with higher teamwork climate scores. Compared with the modal reference group who had worked for their trust for between 6 and 10 years, respondents who had been employed for 11–15 years returned scores 0.050 higher, and for those with more than 15 years' service the increase was 0.065.
- Updating knowledge or skills, having a mentor and doing online training each increased teamwork climate scores by between 0.038 and 0.039.
- The site-level variable of the number of people identified to receive the staff survey had a very small positive effect, adding 0.00005 to scores for every additional 62 (1 SD) people included. This small increase has no practical significance.

The best-fitting model partitioned residual variance as follows:

- 12% at hospital level (estimated parameter 0.072, SE 0.041,  $\chi^2 = 2.994$ ,  $p = 0.084$ )
- 88% with respondents nested within research sites (estimated parameter 0.525, SE 0.035,  $\chi^2 = 223.5$ ,  $p < 0.0001$ ).

Significant variation in the Sexton teamwork climate scores remains with respondents nested within the research sites. Although effects due to unmeasured individual-level variables cannot be ruled out, the result will be interpreted as identifying site-level variation in teamwork climate among the clinical departments that participated in this study.

**TABLE 23** Parameter estimates for fixed factors in the multilevel model of teamwork climate scores (Sexton *et al.*<sup>25</sup>)

Fixed factors	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
<b>Site level</b>		
Number of people sent staff survey <sup>a</sup>	0.003 (0.001)	11.037 ( $p=0.001$ )
<b>Respondent level</b>		
Professional group <sup>b</sup> :		
Doctors	0.438 (0.106)	17.096 ( $p<0.0001$ )
Others – giving direct care	-0.165 (0.125)	1.724 ( $p=0.189$ )
Others – not giving direct care	0.070 (0.149)	0.218 ( $p=0.641$ )
Gender <sup>c</sup>	0.316 (0.099)	10.233 ( $p=0.001$ )
Years worked for trust <sup>d</sup>		
<1		0.011 ( $p=0.917$ )
1–2	0.219 (0.122)	3.229 ( $p=0.072$ )
3–5	0.140 (0.102)	1.854 ( $p=0.173$ )
11–15	0.285 (0.128)	4.984 ( $p=0.026$ )
> 15	0.371 (0.105)	12.478 ( $p=0.0004$ )
Updating	0.222 (0.095)	5.403 ( $p=0.020$ )
Had a mentor	0.220 (0.087)	6.404 ( $p=0.011$ )
Online training	0.215 (0.079)	7.383 ( $p=0.007$ )
Is a manager	0.204 (0.081)	6.369 ( $p=0.012$ )

a Centred on mean.

b Reference group midwives (DU) or nurses (ED).

c Reference group male so parameter for females.

d Reference group 6–10.

Finally, we examined scores for Sexton and colleagues' safety climate scale.<sup>25</sup> *Table 24* shows that seven respondent-level variables mediated these scores (being a manager, updating skills or knowledge, age, professional group, receiving on-the-job training and having a mentor). The best-fitting model used 462 cases and the underlying estimate for Sexton safety climate was 2.879 (SD 0.153). The parameters in *Table 24* show that:

- Managers evaluated safety climate more positively than staff who did not manage others, scoring 0.049 higher.
- Respondents who had been supported to update their knowledge or skills during the past 12 months returned higher safety climate scores (0.048 higher).
- Staff who were older than the modal reference category of 31–40 years returned higher safety climate scores (0.027 higher for those aged 41–50 years and 0.048 higher for those aged 51–65 years).
- Women evaluated safety climate more highly than men, scoring 0.040 higher.
- Staff who had received on-the-job training or had a mentor at some point during the past year returned safety climate scores that were, respectively, 0.033 and 0.026 higher.

The model that contained the fixed factors shown in *Table 24* partitioned residual variation as follows:

- 6% at hospital level (estimated parameter 0.033, SE 0.022,  $\chi^2 = 2.300$ , 1 df,  $p = 0.129$ )
- 94% with respondents nested within research sites (estimated parameter 0.517, SE 0.034,  $\chi^2 = 227.1$ ,  $p < 0.0001$ ).

**TABLE 24** Parameters for best-fitting model of safety climate scores (Sexton *et al.*<sup>25</sup>)

Fixed factors	Parameter estimate (SE)	$\chi^2$ , 1 df (significance)
Is a manager	0.318 (0.076)	17.392 ( $p < 0.0001$ )
Updating	0.314 (0.090)	12.198 ( $p = 0.001$ )
Age (years) <sup>a</sup>	0.021 (0.098)	0.044 ( $p = 0.834$ )
16–30		
41–50	0.178 (0.089)	3.990 ( $p = 0.046$ )
51–65	0.311 (0.099)	9.956 ( $p = 0.002$ )
Professional group <sup>b</sup>	0.298 (0.098)	9.171 ( $p = 0.003$ )
Doctors		
Others – giving direct care	0.144 (0.124)	1.344 ( $p = 0.246$ )
Others – not giving direct care	0.019 (0.141)	0.018 ( $p = 0.893$ )
Gender <sup>c</sup>	0.264 (0.096)	7.530 ( $p = 0.006$ )
On-the-job training	0.218 (0.074)	8.772 ( $p = 0.003$ )
Had a mentor	0.171 (0.084)	4.143 ( $p = 0.042$ )

a Reference group 31–40.

b Reference group midwives (DU) or nurses (ED).

c Reference group male so parameter relates to female.

After adjusting for respondent-level mediating factors, the scale developed by Sexton and colleagues<sup>25</sup> shows significant site-level residual variation in safety climate scores.

### Combining indicators of safety climate

Summary survey-based safety climate scores were calculated as described in *Chapter 3, Combining indicators of safety climate*. The results are displayed in *Table 25*, which has been arranged so that the highest-scoring sites appear at the top: these are the sites where staff perceptions of the elements forming the summary safety climate score were most positive. The mean of the site-specific survey-based safety climate scores is zero (SD 0.835). Echoing the summary survey-based organisational climate scores (see *Table 15*), site 6 is at the top of the table, reflecting positive staff perceptions, whereas site 9 lies at the bottom.

### Comparing survey-based organisational and safety climate scores

This section addresses hypothesis 2a: there will be a strong correlation and good agreement between *questionnaire*-based evaluations of organisational and safety culture. We will see that the staff survey data were consistent with this hypothesis.

The summary survey-based organisational climate *z*-scores (see *Table 15*) and survey-based safety climate *z*-scores (*Table 25*) were strongly correlated ( $r = 0.845$ ;  $n = 16$ ). Research sites that returned high scores on the indicators of organisational climate also returned high scores on the indicators of safety climate and, similarly, sites returning low scores for organisational climate also returned low scores for safety climate (see *Figure 5*, in which each point has been annotated with the research site number). The scores on one evaluation of climate accounted for 71% of the variation in scores in the other evaluation ( $r^2 = 0.714$ ): it seems that staff found it difficult to view one aspect of culture positively (or negatively) without perceiving the other aspect of culture in a similar manner.

**TABLE 25** Summary survey-based safety climate (z-scores)

Hospital (service)	Site	Summary safety climate
H3 (DU)	6	1.276
H2 (ED)	3	1.177
H3 (ED)	5	1.147
H7 (ED)	13	0.839
H5 (DU)	10	0.550
H7 (DU)	14	0.542
H4 (DU)	8	0.099
H2 (DU)	4	-0.158
H8 (DU)	16	-0.306
H1 (DU)	2	-0.402
H1 (ED)	1	-0.656
H4 (ED)	7	-0.843
H8 (ED)	15	-0.892
H6 (ED)	11	-0.976
H6 (DU)	12	-1.042
H5 (ED)	9	-1.162

The SD of differences between organisational climate and safety climate scores was 0.449, and the Bland–Altman plot (*Figure 6*) is annotated with lines positioned at mean difference  $\pm$  2SD. These ‘limits of agreement’ are closely spaced, indicating good agreement between the two measures. Fourteen (88%) of the points on the graph lie between these lines, one point (research site 8) lies on the upper limit and one point (research site 5) lies just below the lower limit. There is no discernible pattern within the differences. Agreement between the standardised summary organisational climate and summary safety climate scores is good.

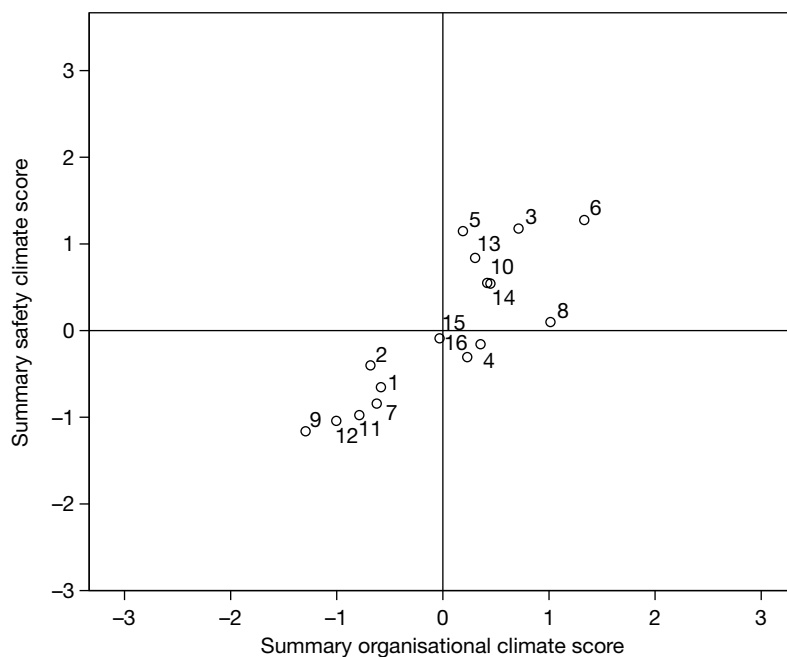
The difference between standardised summary organisational and safety climate scores at research site 5 was -0.958: respondents’ evaluations of safety climate were relatively high when compared with their evaluations of organisational climate. Conversely, the difference between standardised summary organisational and safety climate scores at research site 8 was 0.915: respondents’ evaluations of safety climate were relatively low when compared with their evaluations of organisational climate.

## Summary

A questionnaire survey using prevalidated scales and questions was distributed to staff at 16 research sites, two clinical departments in each of eight hospitals: the ED and the DU. A total of 531 questionnaires were returned, representing a response rate of 27.6%, although response rates differed greatly between research sites (range 9–47%). The demographic profile of respondents broadly reflected the wider NHS workforce with respect to gender and ethnicity. Most (89%) staff survey participants held clinical roles and 37% managed other staff.

Responses to prevalidated scales within the staff survey were investigated using multilevel modelling. A range of individual-level factors were found to mediate scores on one or more scales, including being a manager, gender, age, years of employment with the trust, professional group, ethnicity and participation in certain types of CPD. At site level, the type of service (DU or ED) mediated error wisdom scores (DUs perceived higher organisational error wisdom) and response rates mediated line management scores (departments with above average response rates

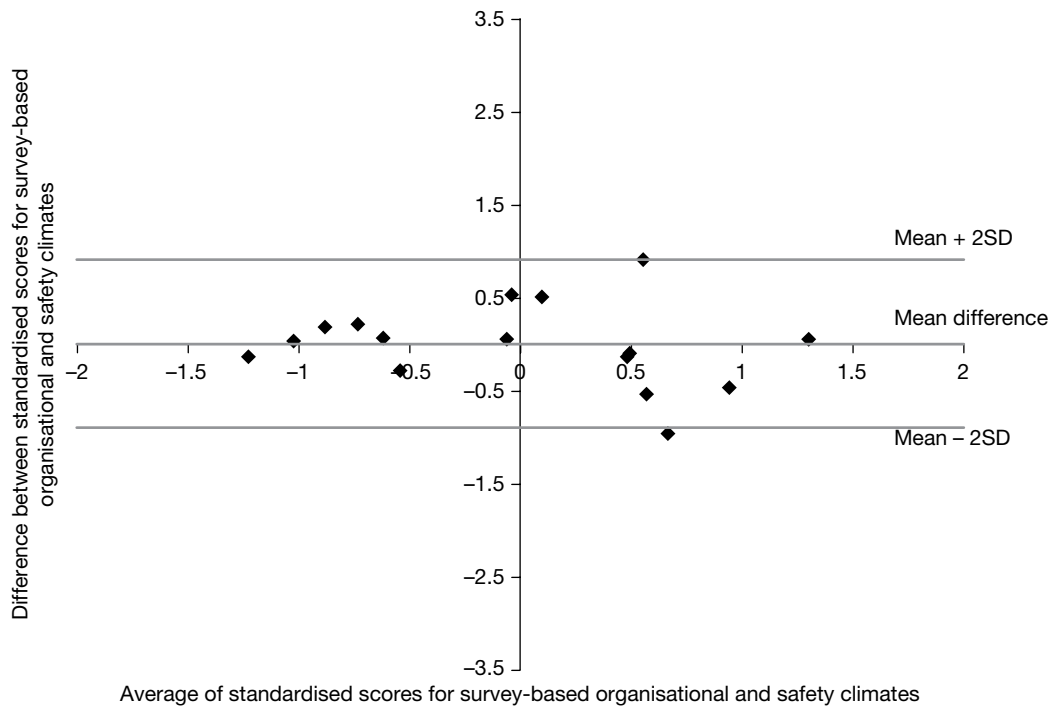




**FIGURE 5** Scatterplot of standardised summary survey-based organisational and safety climate scores. Data points are annotated with research site numbers.

returned lower line management scores). There was no statistically significant residual variation at hospital level, but there was considerable residual variation at the level of respondent nested within research sites after allowing for the effects of individual-level characteristics. This indicates that scores on the scales selected to elicit perceptions of aspects of organisational climate and safety climate predominantly varied at the level of the clinical department, not at hospital level.

Summary survey-based measures of organisational climate and safety climate were calculated for each site. The correlation between these was high (0.845) and agreement was reasonably good (see *Figure 6*).



**FIGURE 6** Bland-Altman plot for standardised summary organisational and safety climate scores.

## Chapter 6

# Strand B: observation-based holistic assessment of culture

This chapter fulfils objective 3: to generate quantified holistic evaluations of organisational and safety culture for each site using observation. It begins by summarising the observation visits and then presents scores allocated for safety-related aspects of culture. It concludes by testing hypothesis 2b, which compares observation-based assessments of organisational and safety cultures.

### Data collection summary

As described in *Chapter 3, Data collection*, non-participant observations were made at each research site, supported by brief interviews with key informants. Six 1-day visits were undertaken at each hospital (two sites) by each research fellow, with service user and service provider observers sometimes accompanying them. Data were collected at all periods of the day excluding midnight to 6 AM, and on all days except Sunday. An average of 31 hours of observation was carried out at each site. The observation record (*Table 26*) shows that a service user coresearcher observed alongside one of the research fellows at three research sites. A service provider coresearcher (midwife or ED nurse) observed alongside one of the research fellows at eight research sites.

Data collection at sites which had also participated in our earlier MOSES study (8, 12 and 14; see *Chapter 4, Recruiting research sites*) occurred in the second half of this study's data collection to ensure a break of around 2 years since the previous study's data collection. EJB and NM collected data at these research sites during the previous study but left this study after site 2. SA and MR collected data at sites 8, 12 and 14: they did not have access to the data from the previous study. After gaining research governance approval, EJB joined SA and MR during some observation visits to site 8. This updated her first-hand experience of using this study's data collection prompt list and provisional scoring scheme, helping her to contribute to the ongoing development of the scoring scheme (see *Chapter 3, Analysis*, and *Chapter 9, Tensions and challenges in quantifying observational data*). She was recognised and welcomed in this DU but, as in other sites and in the MOSES study, staff soon disregarded her presence (owing to the pressing nature of the work in the unit).

### Scoring

The observation data and notes of associated interviews and meetings were analysed as described in *Chapter 3, Analysis*. The resultant scoring is shown in *Table 27*. Two research fellows (SA and MR) independently scored each factor for each site and reached the same decision in most instances (74/128, 58%). Scores that required discussion and reconsideration of field notes before agreement (54, 42%) are annotated by an asterisk. SA and MR did not collect data at research sites 1 and 2 but worked from detailed field notes compiled separately by EJB and NM. Understanding was enhanced by detailed discussion with EJB, who subsequently reviewed all scores for research sites 1 and 2. Assistance from EJB was required to score the two entries marked #, as SA and MR were unable to evaluate these confidently.

TABLE 26 Observer presence by site

Site	Research fellows	Service user coresearcher	Service provider coresearchers
1	✓		✓
2	✓		
3	✓		
4	✓		
5	✓		
6	✓	✓	✓
7	✓		
8	✓		
9	✓		
10	✓		
11	✓		✓
12	✓		✓
13	✓		✓
14	✓		✓
15	✓	✓	✓
16	✓	✓	✓

TABLE 27 Scores summarising observation-based evaluation of culture

	Research site															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<b>1. Organisation and work environment factors</b>																
Staffing	2	2	2	1 <sup>a</sup>	2	2	2 <sup>a</sup>	1	1 <sup>a</sup>	0	2	1	1 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	2 <sup>a</sup>
Premises/equipment	2 <sup>ab</sup>	1	1 <sup>a</sup>	2 <sup>a</sup>	1 <sup>a</sup>	3 <sup>a</sup>	1 <sup>a</sup>	1 <sup>a</sup>	1	1	2	2	1	2 <sup>a</sup>	1 <sup>a</sup>	2
Admin, managerial and other support	2 <sup>ab</sup>	1 <sup>a</sup>	3	2	2 <sup>a</sup>	3 <sup>a</sup>	1	2 <sup>a</sup>	2 <sup>a</sup>	1	2 <sup>a</sup>	2	3 <sup>a</sup>	2	2 <sup>a</sup>	1 <sup>a</sup>
<i>Subtotal (out of 9)</i>	<i>6</i>	<i>4</i>	<i>6</i>	<i>5</i>	<i>5</i>	<i>8</i>	<i>4</i>	<i>4</i>	<i>4</i>	<i>2</i>	<i>6</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>5</i>
<b>2. Team factors</b>																
Informal training and supervision	3	3 <sup>a</sup>	2	3 <sup>a</sup>	3	3 <sup>a</sup>	1	1	1 <sup>a</sup>	3 <sup>a</sup>	2	3	2 <sup>a</sup>	3 <sup>a</sup>	3	3 <sup>a</sup>
Leadership and responsibility	3 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	2	3	2 <sup>a</sup>	1	1	3 <sup>a</sup>	3	2 <sup>a</sup>	3 <sup>a</sup>	1 <sup>a</sup>	3	3	2
Respect/warmth/collegiality	2 <sup>a</sup>	2	3	3	3	3	1 <sup>a</sup>	2	3	3	2	3	3	3 <sup>a</sup>	3	2
Information exchange within the team	3 <sup>a</sup>	2	3 <sup>a</sup>	3	3	3	1 <sup>a</sup>	0 <sup>a</sup>	3	3	3	3	1 <sup>a</sup>	3 <sup>a</sup>	3	2 <sup>a</sup>
Mutual support	3	2	3	3	3	3	1	1 <sup>a</sup>	3	3	3	3	3	3	3	2 <sup>a</sup>
<i>Subtotal (out of 15)</i>	<i>14</i>	<i>11</i>	<i>14</i>	<i>14</i>	<i>15</i>	<i>14</i>	<i>5</i>	<i>5</i>	<i>13</i>	<i>15</i>	<i>12</i>	<i>15</i>	<i>10</i>	<i>15</i>	<i>15</i>	<i>11</i>
<b>Grand total (out of 24)</b>	<b>20</b>	<b>15</b>	<b>20</b>	<b>19</b>	<b>20</b>	<b>22</b>	<b>9</b>	<b>9</b>	<b>17</b>	<b>17</b>	<b>18</b>	<b>20</b>	<b>15</b>	<b>20</b>	<b>20</b>	<b>16</b>

0, positive features consistently absent; 1, positive features frequently absent; 2, positive features frequently present; 3, positive features consistently present.

a Agreed after short discussion. The remainder were agreed immediately.

b SA and MR agreed 'don't know'. EJB rated 2.

## Observation-based evaluations of organisational, teamwork and safety cultures

The subtotals and grand totals in *Table 27* were converted to the mean scores, displayed in *Table 28*. The mean scores were inspected for normality. The one-sample Kolmogorov–Smirnov test indicated that each set of scores could be considered to have been drawn from a normal population (organisation and work environment factors,  $p = 0.362$ ; team factors,  $p = 0.288$ ; overall holistic evaluation of culture,  $p = 0.660$ ). Complementary inspection of summary statistics, histograms and P–P plots indicated that the distribution of team factors scores may be slightly negatively skewed, meaning that more high scores than low scores were awarded. For example, the mean score was 2.39 (SD 0.589), whereas the median was 2.67, and the method of scoring constrains the lower and upper bounds on the scores to be zero and 3. In view of the nature of the analyses that will follow, it was not considered necessary to transform the team factors scores to obtain a more symmetrical distribution.

As noted in *Chapter 3, Analysis*, the holistic observation-based safety culture score contained three organisation and work environment factors alongside five team factors. Consequently, the asymmetry was reduced in team factors scores but was still discernible in the holistic safety culture scores (mean 2.14, SD 0.463, median 2.28); however, this degree of asymmetry was not a concern. A greater concern was the number of ties within the three sets of scores, indicating that the method of scoring did not provide fine-grained differentiation between research sites.

The observation-based culture mean scores in *Table 28* were constrained by the scoring to lie between zero and 3. In fact, the observed range was 0.667–2.667 for organisation and work environment factors, which indicated organisational culture, and 1.000–2.833 for teamwork factors, which indicated important facets of teamwork and safety culture. In *Table 28*, observation-based organisational culture scores are lower than observation-based teamwork culture scores for 14 research sites (88%). Researchers' perceptions of teamwork culture were more positive than their perceptions of organisational culture. The exceptions to this were research sites 7 and 8, both at hospital H4, where researchers awarded low scores to the safety-related teamwork processes they witnessed.

The main contributors to lower observation-based evaluations of organisational culture were shortages of staff and equipment, and lack of space in clinical departments in relation to the number of patients. The observed clinical areas were frequently short-staffed, as evidenced in some places by the persistence of a very hectic pace of work, and in others (all of them DUs) midwives-in-charge were observed making repeated phone calls to midwives at home to see if they could fill gaps in the next shift. Generally, poor scores for premises and equipment derive from evidence such as the frequency of staff asking each other for help in locating routine equipment (e.g. thermometers, electrocardiography machines), or lack of cubicles, rooms or space for trolleys.

## Comparing observation-based organisational and teamwork culture scores

This section addresses hypothesis 2b (see *Chapter 1, Hypotheses*): there will be a strong correlation and good agreement between *holistic* evaluations of organisational and safety culture. We will see that the observation-based scores were poorly correlated but exhibited reasonably good agreement, although they did not agree as closely as survey-based measurements of organisational and safety climates.

**TABLE 28** Observation-based summary scores for organisational, teamwork and safety cultures

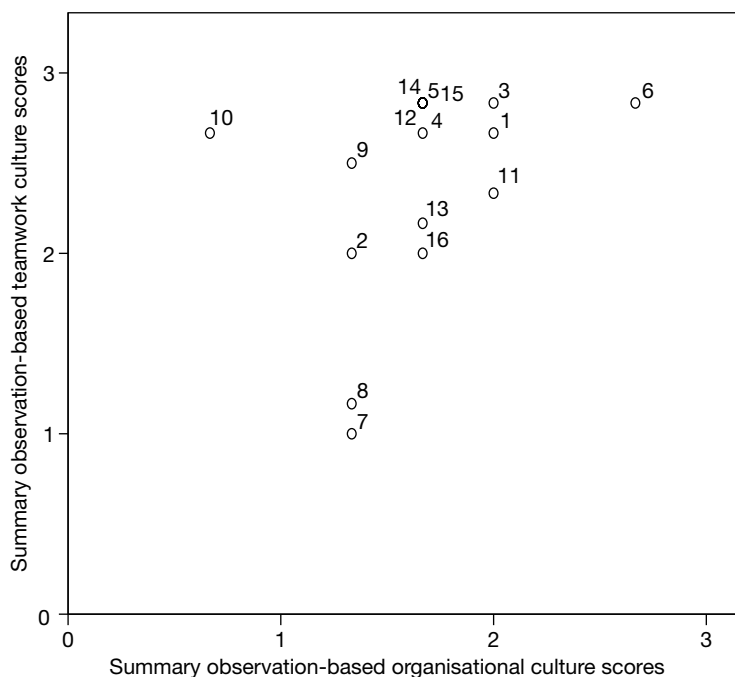
Hospital	Research site	Organisational culture score	Teamwork culture score	Holistic safety culture score
H1	1	2.000	2.667	2.444
	2	1.333	2.000	1.778
H2	3	2.000	2.833	2.556
	4	1.667	2.667	2.333
H3	5	1.667	2.833	2.444
	6	2.667	2.833	2.778
H4	7	1.333	1.000	1.111
	8	1.333	1.167	1.222
H5	9	1.333	2.500	2.111
	10	0.667	2.667	2.000
H6	11	2.000	2.333	2.222
	12	1.667	2.833	2.444
H7	13	1.667	2.167	2.000
	14	1.667	2.833	2.444
H8	15	1.667	2.833	2.444
	16	1.667	2.000	1.889
	All sites	Mean 1.646 (SD 0.429)	Mean 2.385 (SD 0.589)	Mean 2.139 (SD 0.463)

As noted in *Chapter 3, Analyses*, the summary observation-based assessments of organisational and holistic safety culture are not distinct as both contain scores for the organisation and work environment factors. Thus, hypothesis 2b cannot be tested in the form that it is written. It is possible to compare the summary observation-based organisational and teamwork culture scores, which are displayed in the scatterplot in *Figure 7*. The correlation between these scores was low and not statistically significant ( $r=0.356$ ,  $p=0.176$ ,  $n=16$ ). Moreover, when the two outliers in *Figure 7* (sites 7 and 8) were removed from the calculation, the correlation fell ( $r=0.230$ ,  $p=0.430$ ,  $n=14$ ), indicating that poor teamwork culture scores at these sites were having a disproportionate effect on the overall correlation: inflating the apparent level of correlation between teamwork culture and organisational culture. Based on their observations of teams at work, the working environment and supplementary information from key informants, the researchers mostly had different perceptions of organisational culture and teamwork culture in a given location, but these differences were not consistent across research sites.

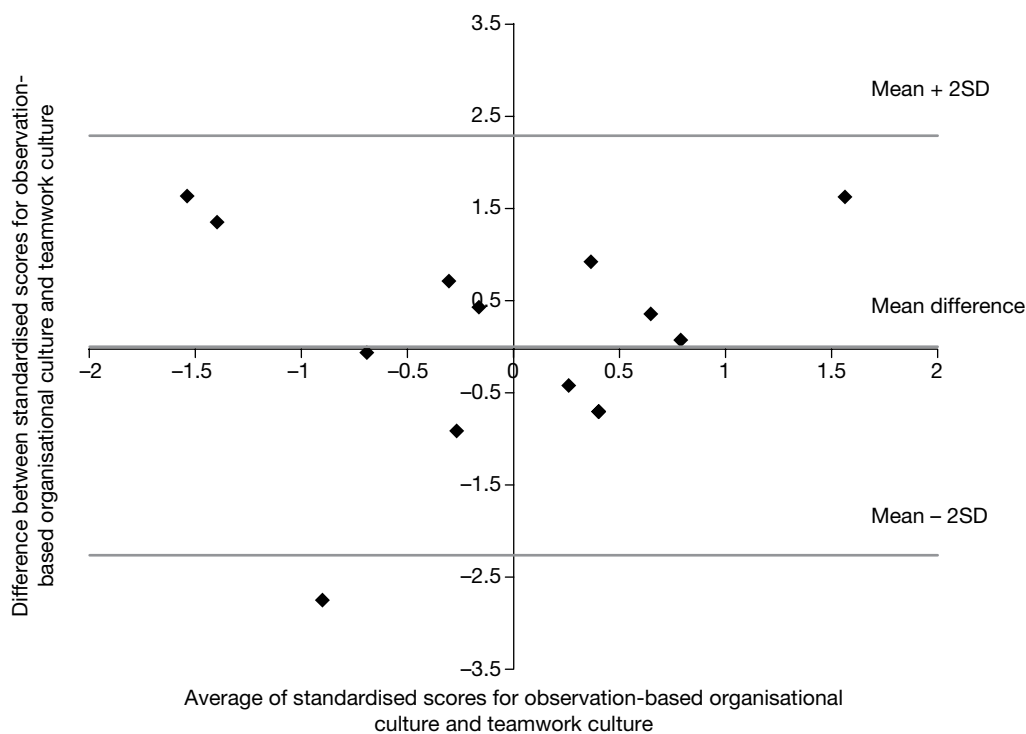
The SD of differences in *Figure 8* is 1.135 (compared with 0.449 for survey-based measurements). There is no discernible pattern and one outlier (research site 10, where researchers rated organisation and work environment factors as deficient but teamwork factors as above average). Only 13 distinct points are visible because tied scores caused the points for research sites 5, 12, 14 and 15 to coincide.

## Summary

Quantified, observation-based evaluations of safety-related aspects of organisational culture and of teamwork culture were made. The level of correlation between these two measures was not statistically significant, but the Bland–Altman plot in *Figure 8* indicated reasonably good agreement between the observation-based measures, although there was one observation which was outwith the limits of agreement. A number of tied scores were obtained, suggesting that



**FIGURE 7** Scatter diagram of standardised observation-based summary scores for organisational culture and teamwork culture. Data points are annotated with research site numbers.



**FIGURE 8** Bland-Altman plot for observation-based organisational culture and teamwork culture.

the scoring system did not provide sufficient differentiation between research sites. The eight measures contributing to either the observation-based organisational culture scores or the observation-based teamwork culture scores were averaged to form a holistic observation-based safety culture score. Consequently, in contrast to the survey-based data collection (strand A, *Chapter 5*), the observation-based strand B data collection did not yield independent evaluations of organisational culture and safety culture.



## Chapter 7

# Strand C: retrospective audit of evidence-based markers of the quality of care

This chapter fulfils objective 4: to obtain criterion-based measurements for the quality of care at each research site. It begins with summaries of the audit findings for each condition, and then reports a summary audit score for each research site.

### Condition-specific audit results

The audited markers were drawn from national guidelines with strong research evidence. They were described in *Box 1* (see *Chapter 2*). For convenience, the appropriate markers will be reproduced in each of the subsections below.

#### Acute coronary syndrome

The audit standards were:

1. Electrocardiography performed immediately.
2. Aspirin should be given immediately if not already given by ambulance service.
3. If patient in pain, times from arrival to the administration of pain relief:
  - in cases of severe pain, 50% in 20 minutes, 75% in 30 minutes, 98% in 60 minutes
  - in cases of moderate pain, 75% in 30 minutes, 98% in 60 minutes.

*Table 29* shows the percentage of audited cases for which the clinical records contained evidence of meeting the standards in these audit criteria (ACS1 and ACS2) and degree to which the targets in ACS3 were met, expressed as a percentage of fully meeting the target. *Table 29* also shows the number of audited cases to which this standard applied. For ACS1, 'immediately' was interpreted as within 10 minutes of arrival. Compliance with the ACS1 standard ranged from 14% to 48%: most patients do not have an ECG in the first few minutes after arrival. The ACS2 marker proved problematic. It did not apply to a large number of audited cases because aspirin had already been given by the ambulance service or the patient was taking aspirin daily. Audited notes were also checked for recorded allergies and any other noted contraindications. Irrespective of this, there was very low compliance with this standard for the cases to which it applied, ranging from zero to 27%. ACS3 required an assessment of pain for all patients and, when moderate or severe pain was detected, early administration of pain relief. Compliance with the ACS3 standard ranged from 7% to 62%: many patients experienced delays before receiving pain relief.

#### Acute severe asthma

The audit standards were:

1. Oxygen saturation should be measured and recorded on arrival in 98% of cases.
2. Salbutamol or terbutaline should be given within 10 minutes of arrival.
3. Intravenous hydrocortisone or oral prednisolone should be given within 30 minutes of arrival in 90% of cases.
4. Measurement of oxygen saturation should be repeated within 60 minutes in 75% of cases.

**TABLE 29** Audit results for acute coronary syndrome (%)

Research site	ACS1	ACS2	ACS3	ACS mean
1	31.1	3.1	33.7	22.7
3	20.8	0.0	61.7	27.5
5	32.0	5.0	26.8	21.3
7	46.0	0.0	16.4	20.8
9	47.8	3.3	19.6	23.6
11	14.3	3.0	16.0	11.1
13	31.9	26.7	58.9	39.2
15	46.0	0.0	6.8	17.6
All ED sites	33.7	4.3	29.6	22.5

There was a change in the guidelines for ASA after data collection at site 1.<sup>182,207</sup> In the earlier guidance, ipratropium was recommended both for moderate asthma, and for acute severe and life-threatening asthma; in the later guidance, for life-threatening asthma only. The new guideline groups these two types of asthma together, but our inclusion criteria were unable to distinguish between acute severe and life-threatening, and therefore had to omit this from our markers. In applying the new standards to sites 3–15, we therefore overlook the fact that life-threatening cases require higher standards than those we are applying: we treat them all as severe.

Table 30 shows the audit results for ASA. In ASA1, ‘on arrival’ was interpreted as within 10 minutes and compliance varied between 40% and 84%; this baseline observation was missing from many patient records. Compliance with audit standard ASA2 was much lower (range 6%–36%), with six EDs scoring no more than 20%. The degree to which the 90% target in audit standard ASA3 was met varied between 14% and 43%. There was wide variation in the degree to which research sites met the target in audit standard ASA4: two EDs scored <20%, whereas the others scored in the range 43–69%.

### Fractured neck of femur

The audit standards were:

1. Pain score recorded on arrival.
2. Time from arrival to receive analgesia:
  - i. if severe pain, 50% in 20 minutes, 75% in 30 minutes, 98% in 60 minutes
  - ii. if moderate, 75% in 30 minutes, 98% in 60 minutes.
3. Radiography performed within 60 minutes of arrival in 75% of cases.
4. If in pain, evidence of re-evaluation of pain (90% of those with severe pain re-evaluated within 30 minutes; 75% of those in moderate pain within 60 minutes).
5. Admitted within 4 hours of arrival.

In audit standard FNoF1, ‘on arrival’ was interpreted as within 10 minutes, and Table 31 shows that, although five EDs scored in the range 80–92%, levels of compliance fell as low as 11% elsewhere. The degree to which the targets for administering analgesia without unnecessary delays (FNoF2) were met mainly varied between 10% and 51%, with research site 9 scoring more highly at 76%. Compliance with the target for timely radiography (FNoF3) varied widely, between 11% and 97%, with half of the EDs scoring over 72%. Research site 9 had the highest level of compliance with FNoF3 (97%) but the lowest level of compliance with FNoF1 (11%). Levels of recorded compliance with FNoF4, the timely re-evaluation of pain, were extremely low:

**TABLE 30** Audit results for acute severe asthma (%)

Research site	ASA1	ASA2	ASA3	ASA4	ASA mean
1	42.5	13.2	26.5	17.1	24.8
3	76.0	16.3	19.6	45.6	39.4
5	67.4	38.3	42.5	57.8	51.5
7	71.4	14.9	25.6	53.3	41.3
9	68.7	19.6	20.0	68.9	44.3
11	63.3	5.9	14.0	13.3	24.1
13	83.7	36.2	39.1	64.4	55.9
15	39.6	14.3	17.1	43.1	28.5
All ED sites	64.9	19.9	24.6	46.3	38.9

zero for five EDs and elsewhere no more than 4% in cases to which this standard applied. In half of the EDs, admission within 4 hours (FNoF5) was achieved in over 85% of cases. Elsewhere the level of compliance dropped as low as 13%. It was noticeable that research site 1 had a high level of compliance with FNoF1 and an exceptionally low level of compliance with FNoF5, whereas the reverse was true for research site 13. These within-condition variations are smoothed out by the calculation of condition-specific means and further aggregation in *Summarising the site-specific audit results*. This will be discussed in *Chapter 9*.

### Normal delivery

The audit standards were:

1. Initial assessment should include temperature, pulse, blood pressure, urinalysis, length, strength and frequency of contractions, fundal height, lie, presentation, position and station, show, liquor, blood, pain and fetal heart rate.
2. Assessment during first stage should include fetal heart rate every 15 minutes, frequency of contractions every 30 minutes, pulse hourly, temperature and blood pressure 4-hourly, vaginal examination offered 4-hourly, frequency of emptying bladder.
3. Assessment during second stage should include fetal heart rate every 5 minutes, frequency of contractions every 30 minutes, pulse and blood pressure hourly, vaginal examination offered hourly, frequency of emptying bladder, woman's emotional/psychological needs.

Delivery unit midwives rarely recorded all the observations in the multifaceted audit standards for initial assessment and assessments during the first and second stages of labour. We will discuss this in *Chapter 9*. *Table 32* shows that compliance with audit standard ND1 was zero at all sites. The most frequently missing observations were blood and show. In half of the DUs, compliance with audit standard ND2 was between 21% and 26%, but elsewhere it was < 12%. Compliance with ND3 mostly varied between 17% and 33%, with research site 16 scoring only 2%. The applicability of some markers for standards ND2 and ND3 depended on the length of stages 1 and 2 of labour, respectively; where standards were inapplicable, cases were excluded from the denominator.

### Emergency caesarean section

The audit standards were:

1. Evidence of consultant obstetrician involvement in the decision to perform an ECS should be documented.

**TABLE 31** Audit results for FNoF (%)

Research site	FNoF1	FNoF2	FNoF3	FNoF4	FNoF5	FNoF mean
1	90.3	29.2	19.4	0.0	12.9	30.4
3	80.0	29.2	26.0	2.7	40.0	35.6
5	92.2	38.7	96.7	3.7	92.2	64.7
7	83.7	46.2	40.0	3.0	76.0	49.8
9	10.9	76.1	97.0	0.0	60.0	48.8
11	83.3	10.2	87.1	0.0	97.9	55.7
13	23.4	30.2	72.5	0.0	95.9	44.4
15	56.0	50.7	10.7	0.0	86.0	40.7
All ED sites	60.6	35.2	46.3	1.2	72.7	43.2

**TABLE 32** Audit results for ND (%)

Research site	ND1	ND2	ND3	ND mean
2	0.0	8.9	16.7	8.5
4	0.0	6.1	20.4	8.8
6	0.0	11.4	20.8	10.7
8	0.0	20.5	26.1	15.5
10	0.0	25.6	16.7	14.1
12	0.0	21.2	33.3	18.2
14	0.0	24.4	16.7	13.7
16	0.0	0.0	2.0	0.7
All DU sites	0.0	14.6	19.1	16.9

2. Women undergoing an ECS should be offered prophylactic antibiotics at the time of ECS.
3. All women undergoing ECS must receive thromboprophylaxis for VTE.
4. Women undergoing ECS should be offered regional anaesthesia (spinal or epidural).

Table 33 shows that in six DUs audit standard ECS1 was met in the majority of cases (range 55–84%); elsewhere levels dropped as low as 26%. Half of the DUs had high levels of compliance with audit standard ECS2 (81–98%) and the other half had moderate levels (48–55%). Compliance with audit standard ECS3 was generally high, between 92% and 100% at seven research sites, but was only 45% at research site 4. Discounting the very small number of cases for which regional anaesthesia was clinically inappropriate, compliance with audit standard ECS4 was high (range 92–100%).

### **Meconium-stained liquor (grade 2 or 3)**

The audit standards were:

1. Continuous electronic fetal monitoring (EFM) should be advised.
2. Health-care professionals trained in advanced neonatal life support should be available for the birth.
3. Baby assessment should include at 1 and 2 hours and then 2-hourly for 12 hours:
  - general well-being
  - chest movements and nasal flare
  - skin colour including perfusion

**TABLE 33** Audit results for emergency caesarean section (%)

Research site	ECS1	ECS2	ECS3	ECS4	ECS mean
2	55.1	54.6	96.0	100.0	76.4
4	62.0	52.0	44.9	98.0	64.2
6	26.0	53.3	96.0	96.0	67.8
8	60.0	48.0	92.0	100.0	75.0
10	64.0	98.0	96.0	96.0	88.5
12	44.9	98.0	98.0	95.9	84.2
14	83.7	80.9	98.0	98.0	90.1
16	60.0	98.0	100.0	92.0	87.5
All DU sites	56.9	74.0	90.2	97.0	79.6

- feeding
- muscle tone
- temperature
- heart rate and
- respiratory rate.

Table 34 shows that continuous electronic fetal monitoring (MSL1) was arranged in nearly all cases (range 81–100%); in many cases, this was already in place when meconium was detected. Compliance with audit standard MSL2 varied much more widely: between 69% and 98% in five DUs and dropping as low as 6% elsewhere. Similar to the results for the multifaceted ND audit standards, very few babies received the multifaceted observations required by MSL3. The most commonly missing observations were feeding and muscle tone.

## Summarising site-specific audit results

The distribution of mean audit scores for each condition was examined as described in *Chapter 3, Strand C: criterion-based assessment of quality of care, Analysis*. In each case the distribution was normal. The condition-specific means were scaled to the standard normal distribution (see *Chapter 3, Combining indicators of organisational climate*) and aggregated by finding the mean of the z-scores for the three conditions audited at each site. This formed the summary audit scores shown in *Table 35*, which has been arranged so that the research sites scoring most highly are at the top of the table. As expected, the mean of summary Audit scores was zero. Research sites with positive summary audit scores in *Table 35* were above average for the sample of hospitals in this study, whereas those with negative summary audit scores were below average. The SD of summary audit scores was 0.675.

## Summary

This method in general proved to be able to distinguish between sites. Although some markers were met in few or no cases, and others in nearly all cases, such variations disappeared in the averaging process.

**TABLE 34** Audit results for MSL (%)

Research site	MSL1	MSL2	MSL3	MSL mean
2	81.1	72.3	0.0	51.1
4	88.0	47.7	0.0	45.2
6	95.8	33.3	0.0	43.0
8	88.5	68.8	0.0	52.4
10	97.9	81.3	0.0	59.7
12	100.0	97.7	2.2	66.6
14	100.0	6.3	6.3	37.5
16	95.9	76.0	0.0	57.3
All DU sites	93.8	65.1	0.6	53.2

**TABLE 35** Summary audit scores

Hospital (service)	Research site	Mean of standardised audit scores
H6 (DU)	12	1.117
H7 (ED)	13	1.087
H3 (ED)	5	0.843
H5 (DU)	10	0.769
H5 (ED)	9	0.258
H4 (DU)	8	0.145
H4 (ED)	7	0.089
H7 (DU)	14	0.025
H2 (ED)	3	-0.118
H8 (DU)	16	-0.173
H1 (DU)	2	-0.282
H6 (ED)	11	-0.607
H8 (ED)	15	-0.673
H3 (DU)	6	-0.720
H1 (ED)	1	-0.880
H2 (DU)	4	-0.881

## Chapter 8

# Comparison results from strands A–C

This chapter tests four of the hypotheses listed in *Chapter 1, Hypotheses*, which compare the summary scores from strands A–C of data collection to examine the degree to which they correlate and agree.

### Comparing survey-based assessments with observation-based assessments

#### *H1a: organisational climate and culture assessments*

Hypothesis 1a (*Chapter 1, Hypotheses*) was ‘There will be a strong correlation and good agreement between questionnaire-based and holistic evaluations of *organisational culture*’. This was investigated using the summary survey-based organisational climate *z*-scores (*Table 15* and reproduced in *Table 36*) and the summary observation-based organisational culture scores from *Table 28* after they had been scaled to the standard normal distribution (see *Table 36*).

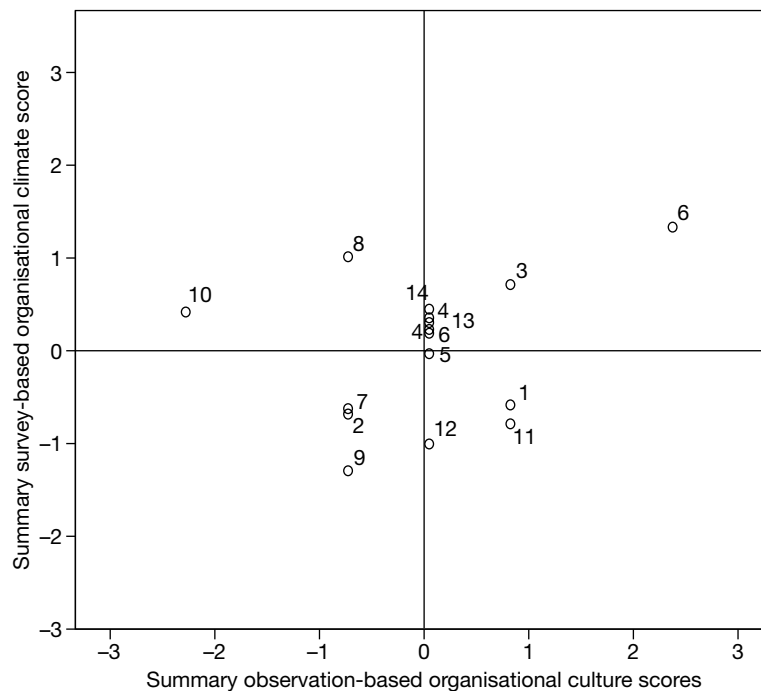
Visual comparison of the organisational climate and culture scores can be made by viewing the scatterplot in *Figure 9*. As noted in *Chapter 6, Scoring*, there is some concern about the number of ties in the observation-based scores. These show up in *Figure 9* as three vertical stacks of data points: sites for which the observation-based measurement resulted in the same total score for organisation and work environment factors. Visual inspection of *Table 36* and *Figure 9* also highlights that seven research sites had an above average (positive) score on one measurement and a below average (negative) score on the other measurement, while nine sites had two positive or two negative scores, signalling a likely lack of correlation between these measures. As expected from the visual inspections, the correlation between the organisational climate and culture scores in *Table 36* was low ( $r=0.252$ ) and not statistically significant ( $p=0.346$ ). In addition, the correlations for the prevalidated scales contributing to the summary organisational climate score (*Chapter 3, Combining indicators of organisational culture*) both had low correlations with the observation-based organisational culture score (organisation,  $r=0.236$ ; error wisdom,  $r=0.108$ ).

The SD of differences plotted in *Figure 10* is 1.089. Several differences cluster close to zero, but there is not a clear pattern to the distribution of differences. There is one point outside the ‘limits of agreement’ (mean difference  $\pm 2SD$ ). This relates to research site 10, where researchers assessed organisation and work factors as deficient but staff perceived the organisational climate more positively. The two scales contributing to the survey-based organisational climate score were also compared with observation-based organisation culture scores using Bland–Altman plots and neither offered notably better agreement.

Those aspects of organisational culture that could be observed during this study (the adequacy of staffing and premises, availability of sufficient functioning equipment, the availability and quality of support from administrative staff, senior managers and other supporting teams) did not correlate with staff perceptions of organisational climate, where the survey instruments focused on perceptions of senior management, organisational communication, patient care, whistle-blowing and error-reporting procedures and organisational error wisdom. These matters were touched upon in the brief key informant interviews that contextualised observations but they could not be observed directly. The level of agreement between the survey-based and

**TABLE 36** Summary survey-based organisational climate and observation-based organisational culture z-scores

Hospital	Site	Survey-based z-scores	Observation-based z-scores
H1	1	-0.584	0.824
	2	-0.682	-0.727
H2	3	0.714	0.824
	4	0.355	0.048
H3	5	0.189	0.048
	6	1.333	2.375
H4	7	-0.623	-0.727
	8	1.014	-0.727
H4	9	-1.293	-0.727
	10	0.418	-2.278
H6	11	-0.787	0.824
	12	-1.004	0.048
H7	13	0.304	0.048
	14	0.449	0.048
H8	15	-0.031	0.048
	16	0.230	0.048

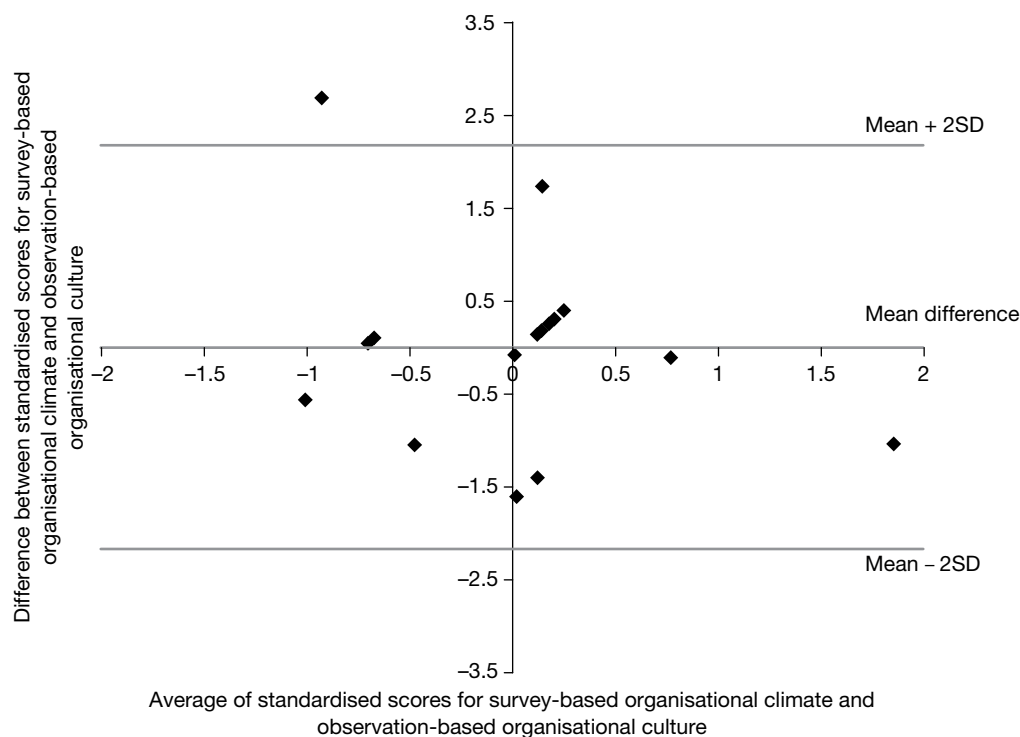
**FIGURE 9** Scatterplot of standardised summary organisational climate and culture scores. Data points have been annotated with the research site numbers.

observation-based measurements was reasonably good; only one point lay outside the limits of agreement (mean difference  $\pm 2SD$ ). Some clustering of points is visible, created by tied scores (lack of discrimination) in the observation-based measure.

### **H1b: Safety climate and culture assessments**

Hypothesis 1b proposed that ‘There will be a strong correlation and good agreement between questionnaire-based and holistic evaluations of *safety* culture’. This was investigated using the





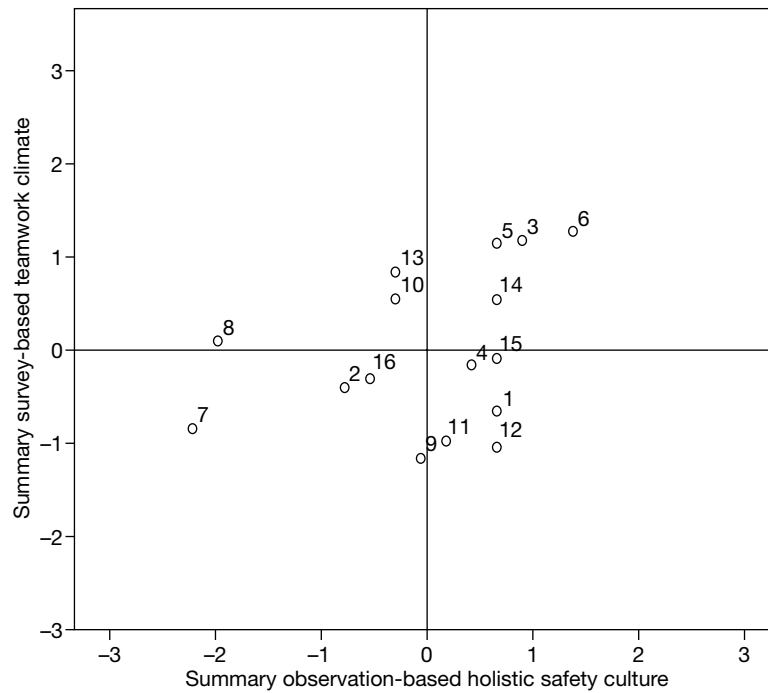
**FIGURE 10** Bland-Altman plot for survey-based organisational climate and observation-based organisational culture scores.

summary survey-based safety climate z-scores (Table 25 and reproduced in Table 37) and both the observation-based teamwork and holistic safety culture scores (Table 28 and, after scaling to standard normal distribution, z-scores in Table 37). Visual comparisons can be made by inspecting Figures 11 and 12. As in *H1a: organisational climate and culture assessments*, visual inspections of the scatterplots and Table 37 signalled that correlations would be low.

The correlation between the summary survey-based safety climate and observation-based holistic safety culture z-scores was low ( $r=0.345$ ) and not statistically significant ( $p=0.191$ ,  $n=16$ ; see Figure 11). Similarly, the correlation between the summary survey-based safety climate and observation-based teamwork culture z-scores was low ( $r=0.316$ ) and not statistically significant ( $p=0.234$ ,  $n=16$ ; see Figure 12). Moving to agreement, the SD of differences in Figure 13 was 1.059. There was no discernible pattern, and only one point lay close to the upper limit of agreement (research site 8, where researchers evaluated the safety culture deficient but staff returned average safety climate scores). Thus, there was good agreement between the contrasting measurements of safety climate and holistic safety culture. Ordinal evaluations of those aspects of safety culture that could be observed during this study (safety-related organisation and work environment factors; aspects of team members' interactions) agreed but did not correlate with staff perceptions of safety climate, where the survey questions focused on perceptions of overload, line management, teamwork, interactions between team members, the safety of care and management, organisational communication, patient safety concerns and responses to errors. There was some overlap between the components of these measurements. Similarly, the SD of differences in Figure 14 was 1.082, there was no discernible pattern, and again the point relating to research site 8 was the only one close to a limit of agreement (because observers judged safety-related teamwork factors to be poor but staff returned average safety climate scores). Overall, agreement was good.

**TABLE 37** Summary survey-based safety climate and observation-based teamwork and safety cultures (z-scores)

Hospital	Site	Survey-based summary safety climate	Observation-based teamwork culture	Observation-based holistic safety culture
H1	1	-0.656	0.477	0.659
	2	-0.402	-0.654	-0.779
H2	3	1.177	0.760	0.899
	4	-0.158	0.477	0.420
H3	5	1.147	0.760	0.659
	6	1.276	0.760	1.378
H4	7	-0.843	-2.352	-2.218
	8	0.099	-2.069	-1.978
H4	9	-1.162	0.194	-0.060
	10	0.550	0.477	-0.300
H6	11	-0.976	-0.088	0.180
	12	-1.042	0.760	0.659
H7	13	0.839	-0.371	-0.300
	14	0.542	0.760	0.659
H8	15	-0.892	0.760	0.659
	16	-0.306	-0.654	-0.539

**FIGURE 11** Scatterplot of standardised survey-based safety climate and observation-based holistic safety culture scores. Data points are annotated with research site numbers.

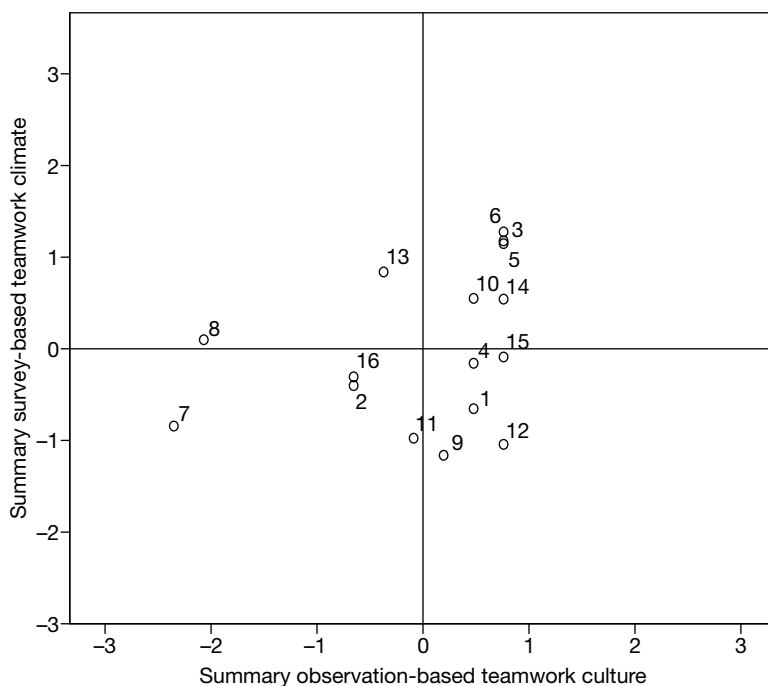


FIGURE 12 Scatterplot of standardised survey-based safety climate and observation-based teamwork culture scores.

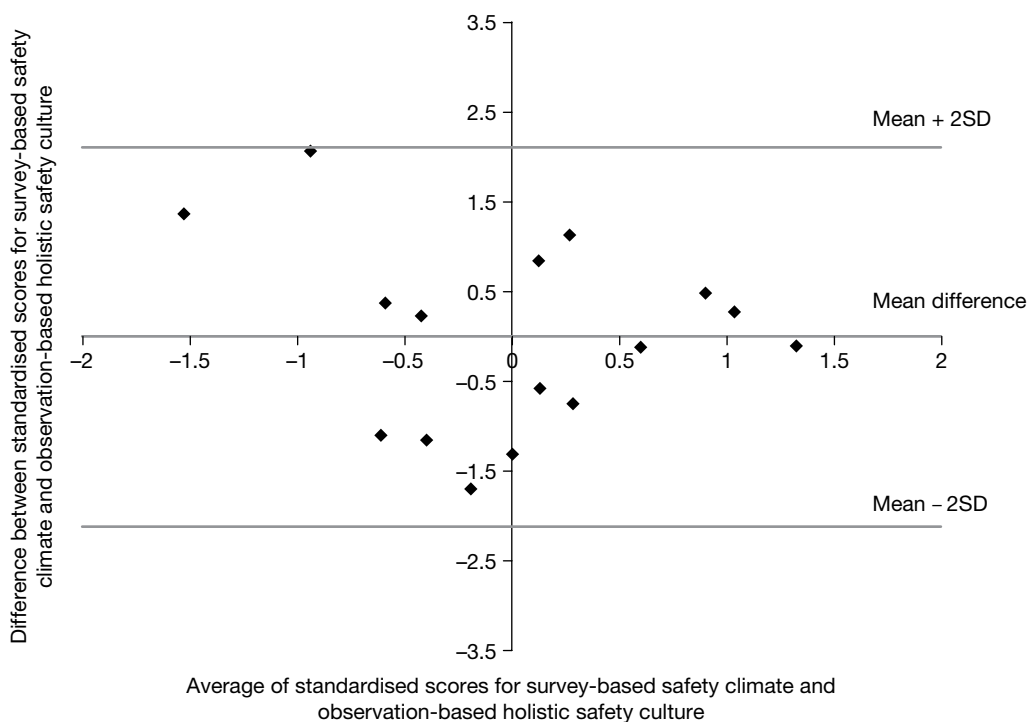
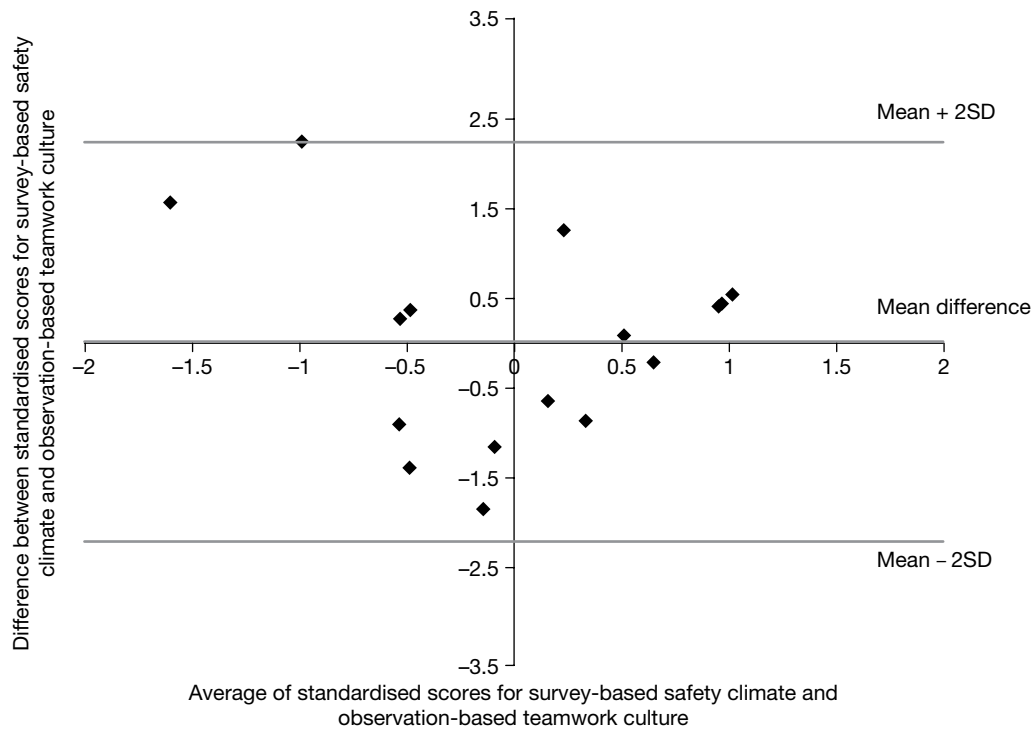


FIGURE 13 Bland-Altman plot for survey-based safety climate and observation-based holistic safety climate.



**FIGURE 14** Bland–Altman plot for survey-based safety climate and observation-based teamwork culture.

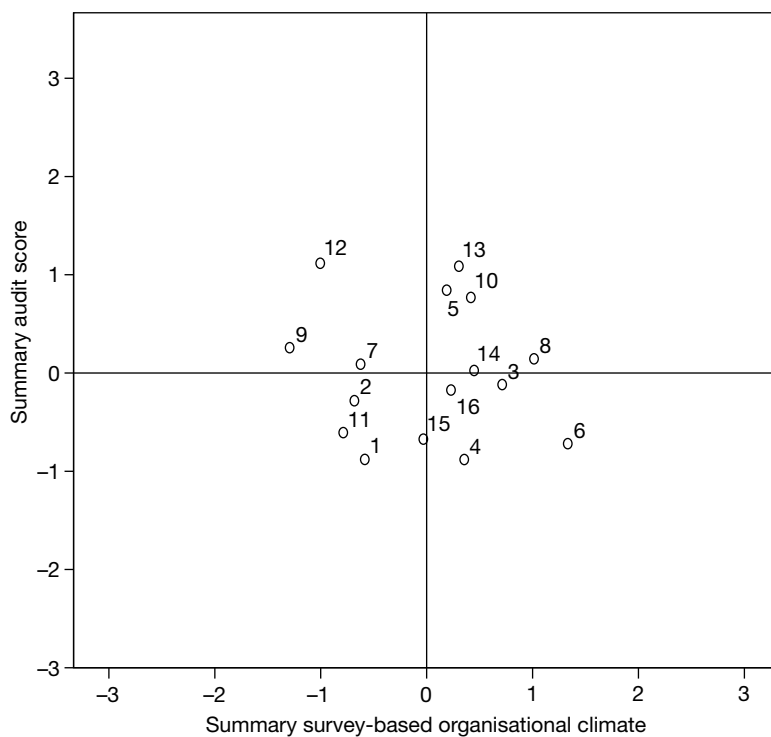
The prevalidated scales forming parts of the summary survey-based safety climate measure [overload, line management, teamwork climate (Sexton *et al.*<sup>25</sup>) and safety climate (Sexton *et al.*<sup>25</sup>)] were also compared with observation-based assessments of culture but none of these offered improved agreement.

## Comparing climate and culture assessments with the quality of care

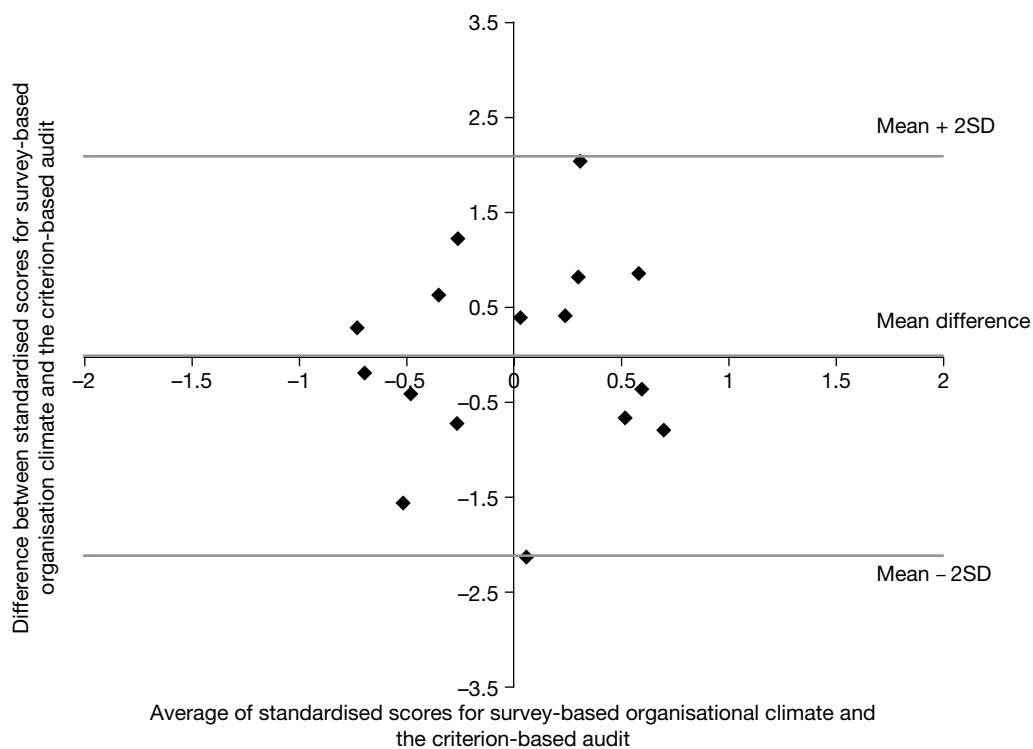
### *H3a(i): comparing survey-based organisational climate with audit assessments of the quality of care*

The first part of hypothesis 3a (*Chapter 1, Hypotheses*) proposed that ‘There will be a moderately strong correlation and reasonably good agreement between criterion-based measurements of the quality of care and questionnaire-based evaluations of organisational culture’. This was investigated using the summary survey-based organisational climate *z*-scores (see *Table 15*) and the summary audit *z*-scores (see *Table 35*). The correlation between these measures was negligible ( $r = -0.096$ ), which is evident from visual inspection of *Figure 15*. Moving to agreement, the SD of differences in *Figure 16* is 1.058; there is no discernible pattern but two points lie on the limits of agreement. There is reasonably good agreement between these measures.

The prevalidated scales contributing to the summary survey-based organisational climate scores ‘organisation’ and ‘error wisdom’ were also compared separately with the criterion-based audit scores, neither displayed better agreement.



**FIGURE 15** Scatterplot of standardised summary survey-based organisational climate and audit scores. Data points are annotated with research site numbers.



**FIGURE 16** Bland-Altman plot for survey-based organisational climate and criterion-based audit.

### **H3a(ii): comparing observation-based organisational culture with audit assessments of the quality of care**

According to the second part of hypothesis 3a, ‘There will be a moderately strong correlation and reasonably good agreement between criterion-based measurements of the quality of care and holistic evaluations of organisational culture.’ This was investigated using the observation-based organisational culture (see *Table 28*) and summary audit *z*-scores (see *Table 35*). There was a moderate negative correlation between these measures ( $r = -0.481$ ,  $p = 0.059$ ,  $n = 16$ ), although visual inspection of *Figure 17* reveals that there can be considerable variation in audit scores among research sites with the same observation-based organisational culture scores and that the results from research sites 6 and 10 may be having a disproportionate effect on the correlation coefficient. Removing research sites 6 and 10 from the analysis reduces the correlation to  $r = -0.347$  ( $p = 0.225$ ,  $n = 14$ ). Moving to agreement, the SD of differences in *Figure 18* is 1.451, tied observation scores are linked to a pattern of three diagonal lines within the points within the limits of agreement (mean difference  $\pm 2SD$ ), and the points relating to research sites 6 and 10 are once again anomalous, lying, respectively, above and below the limits of agreement. At research site 6 researchers perceived a positive organisational culture but audit scores were below average. Conversely, researchers awarded site 10 the lowest score for organisational climate but audit scores were above average. Agreement between these measures is poor.

### **H3b(i): comparing survey-based safety climate with audit assessments of the quality of care**

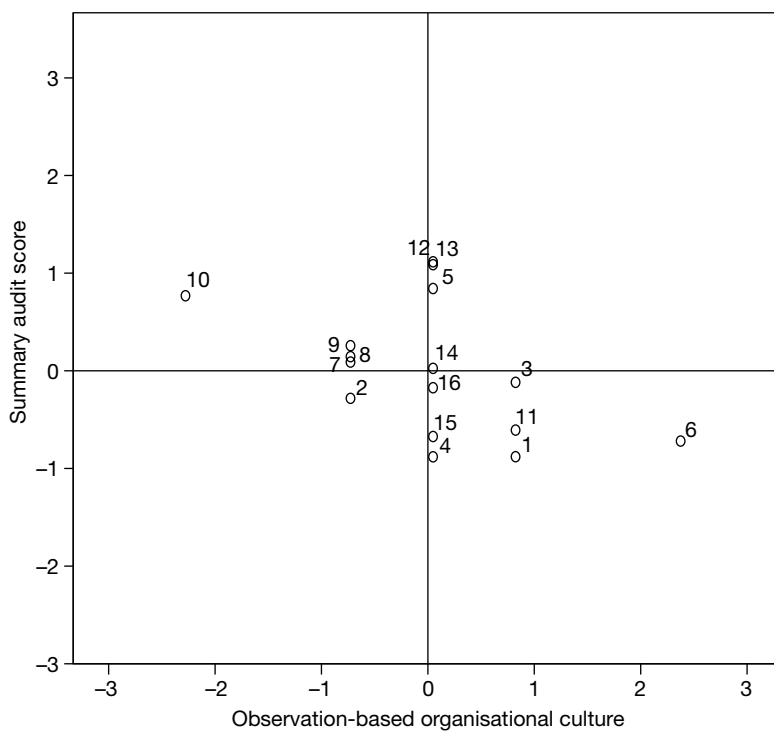
According to the first part of hypothesis 3b, ‘There will be strong correlations and good agreement between criterion-based measurements of the quality of care and questionnaire-based evaluations of safety culture.’ This was investigated using the summary survey-based safety climate (see *Table 25*) and summary audit *z*-scores (see *Table 35*), which are displayed in *Figure 19*. The correlation between these sets of scores was very low: 0.150. Moving to agreement, the SD of differences in *Figure 20* is 0.992, there is no discernible pattern to the points, and one point lies outside the limits of agreements and one point lies on the upper limit of agreement. Agreement between these measures is reasonably good.

The four prevalidated scales (overload and line management,<sup>175</sup> and teamwork and safety climates<sup>25</sup>) were also investigated and none provided better agreement with audit scores. Correlations also remained low, maximum 0.209.

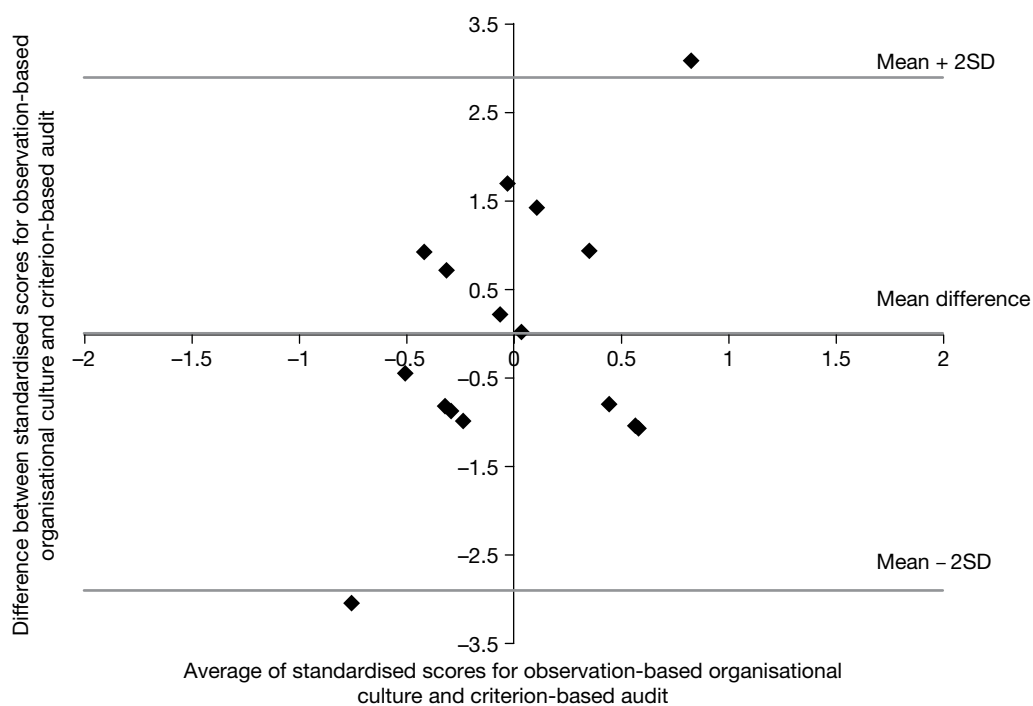
### **H3b(ii): comparing observation-based safety culture with audit assessments of the quality of care**

The second part of hypothesis 3b proposed that ‘There will be strong correlations and good agreement between criterion-based measurements of the quality of care and holistic evaluations of safety culture.’ This was investigated using the observation-based teamwork and holistic safety culture *z*-scores (see *Table 37*) and the summary audit *z*-scores (see *Table 35*). The comparison between observation-based holistic safety climate scores and audit scores found low negative correlation ( $r = -0.201$ ; *Figure 21*), and research sites 7 and 8, at the same hospital, appear to be outliers: researchers awarded very low scores for safety culture but audit results were around average. Moving to agreement, the SD of differences in *Figure 22* is 1.314 and the points relating to research sites 7 and 8 lie towards the bottom-left of the graph. There is no discernible pattern to the remaining differences and all points lie within the limits of agreement. Overall, the agreement between these two measures is reasonably good.

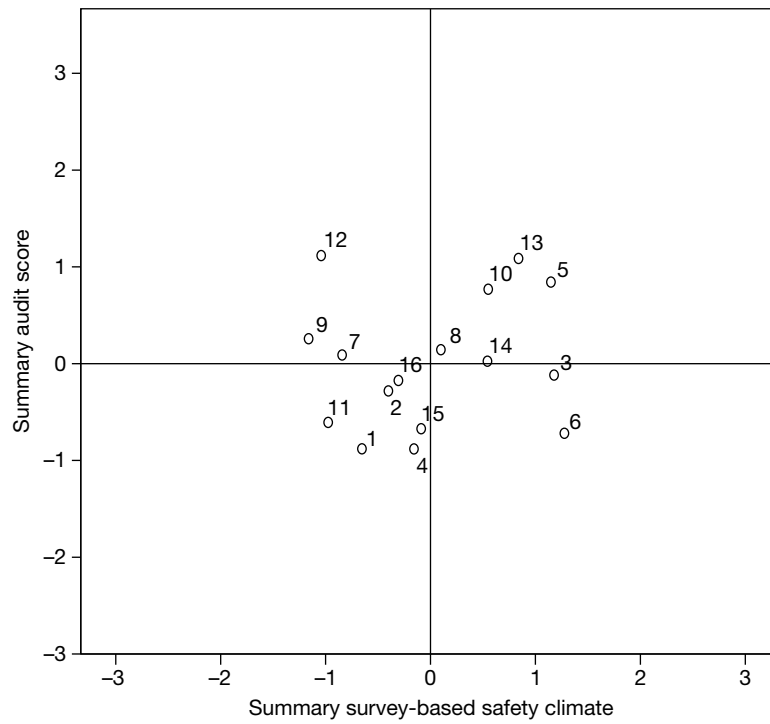
The correlation between observation-based teamwork culture scores and summary audit scores was negligible ( $r = -0.062$ ; *Figure 23*), and research sites 7 and 8 remain outliers, with the researchers perceiving poor teamwork culture, whereas audit results were around average. Moving to agreement, the SD of differences in *Figure 24* is 1.241. The point relating to research



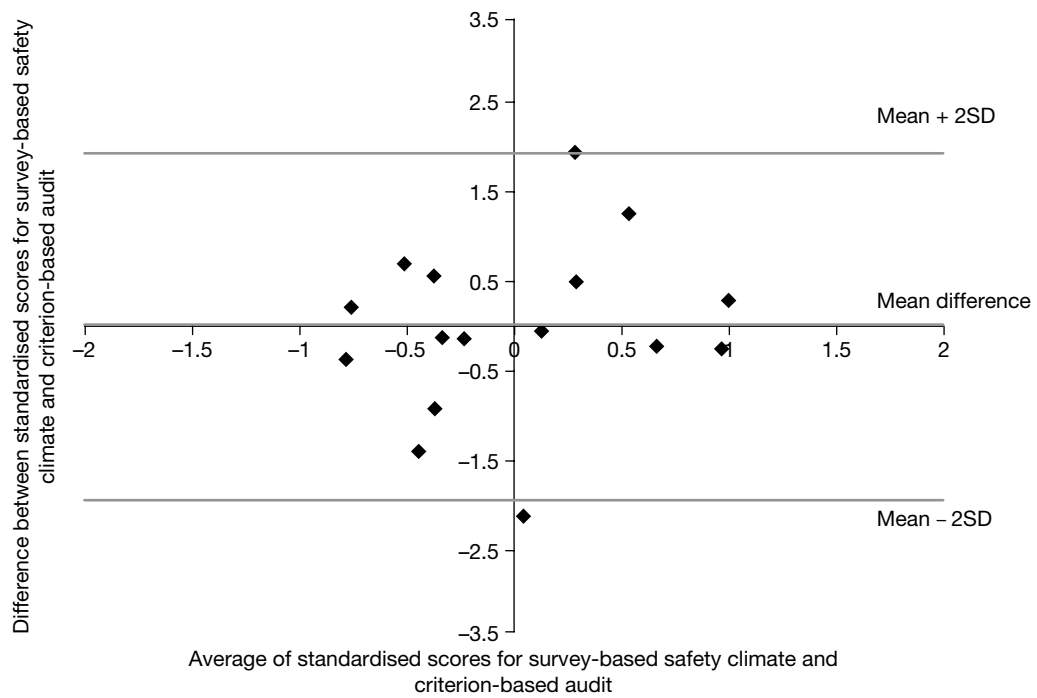
**FIGURE 17** Scatterplot of standardised summary observation-based organisational culture and audit scores. Data points are annotated with research site numbers.



**FIGURE 18** Bland-Altman plot for observation-based organisational culture and criterion-based audit.

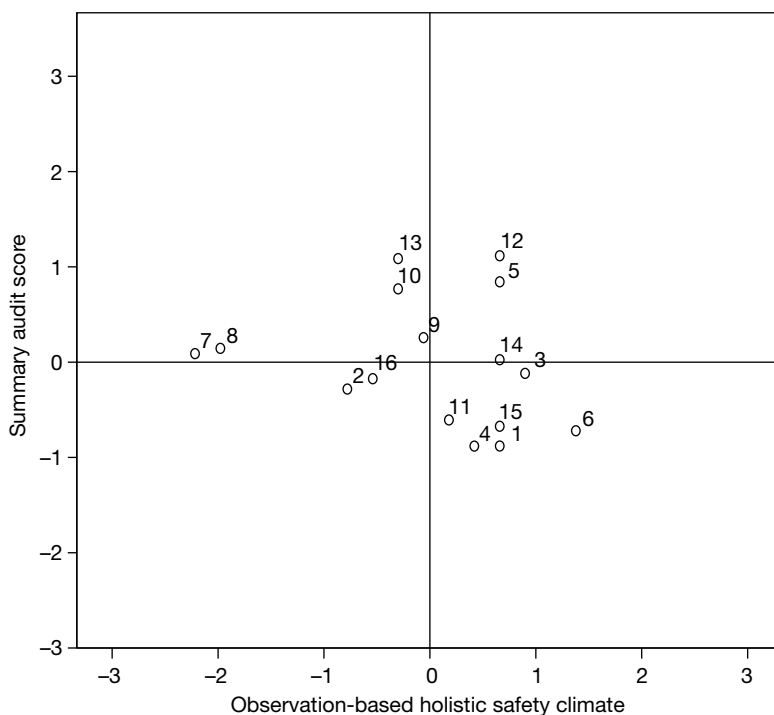


**FIGURE 19** Scatterplot of standardised summary survey-based safety climate and audit scores. Data points are annotated with research site numbers.

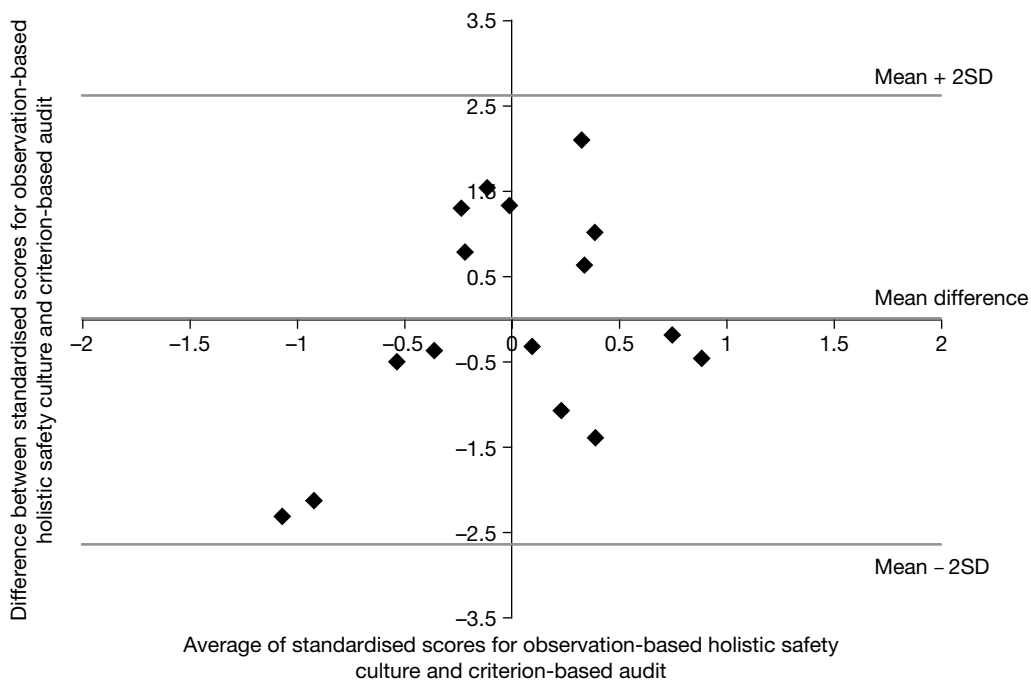


**FIGURE 20** Bland–Altman plot for survey-based safety climate and criterion-based audit.

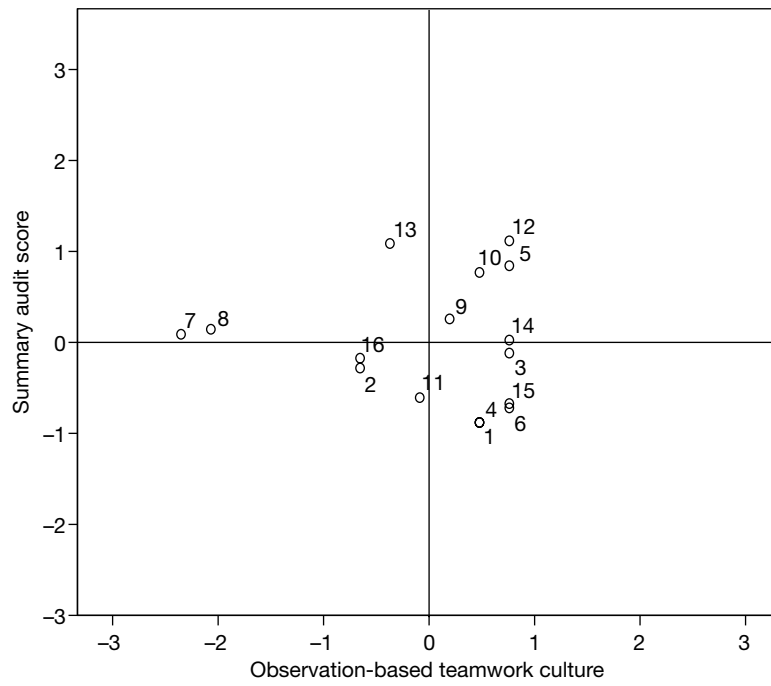




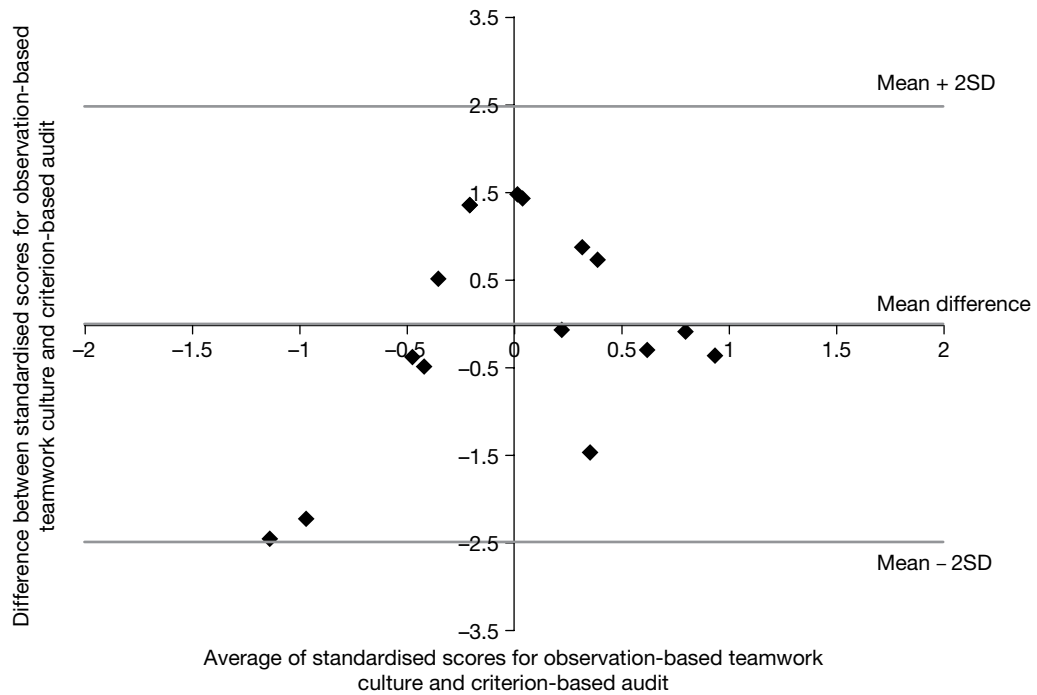
**FIGURE 21** Scatterplot of standardised summary observation-based holistic safety culture and audit scores. Data points are annotated with research site numbers.



**FIGURE 22** Bland-Altman plot for observation-based holistic safety culture and criterion-based audit.



**FIGURE 23** Scatterplot of standardised summary observation-based teamwork culture and audit scores. Data points are annotated with research site numbers.



**FIGURE 24** Bland–Altman plot for observation-based teamwork culture and criterion-based audit.

site 7 now lies very close to the lower limit of agreement although all points are between the lines. There is no pattern to give rise to concern. The agreement between these measurements is reasonably good.

## Summarising the comparisons

The comparisons made in this chapter and *Chapter 5, Hypothesis 2a: comparing survey-based organisational and safety climate scores*, and *Chapter 6, Hypothesis 2b: comparing observation-based organisational and teamwork culture scores*, are summarised in *Table 38*. Correlations were generally low, indicating that the pairs of measurements did not have a linear relationship. The only comparison producing a strong correlation was between survey-based organisational and safety climates, both representing staff perceptions. The level of agreement was good for two comparisons: between the summary survey-based safety climate measure and observation-based assessments of holistic safety culture and teamwork culture (the latter is a subset of the former). The level of agreement was reasonably good for seven comparisons: between survey-based organisational and safety climates (both representing staff perceptions) and between observation-based organisational and teamwork cultures (both predominantly representing researchers' perceptions); between survey-based organisational climate (staff) and observation-based organisational culture (researchers); and between criterion-based audit scores and (a) survey-based organisational climate (staff), (b) survey-based safety climate (staff), (c) observation-based holistic safety culture (researchers) and (d) observation-based teamwork culture (researchers) (d is a subset of c). Agreement was poor for one comparison, between criterion-based audit and observation-based organisational culture (predominantly researchers' perceptions).

But how close does agreement need to be to be useful for practical application? This is a moot point. However, smaller SDs (narrower limits of agreement) would be considered better than larger SDs and a high proportion of small differences would indicate that the measurements are in close agreement at most research sites. As an exercise, *Table 39* reproduces the SD of differences for each comparison and shows the proportion of differences that lie in a narrow range of  $-0.5$  to  $0.5$ . This table suggests that the comparison for which measurements most closely agree is between survey-based assessments of organisational and safety climates, representing staff perceptions. The measurements producing the next closest comparisons were (a) survey-based safety climate and criterion-based audit and (b) survey-based organisational climate and observation-based organisational culture. As expected from earlier tables and graphs, the comparison between observation-based organisational culture and criterion-based audit yielded the poorest level of agreement.

In summary, the survey-based assessment of safety climate developed for this study provided the closest agreement with the criterion-based audit of markers of the quality of care, also developed for this study. The summary survey-based assessment of safety climate comprised prevalidated scales from the NHS annual staff survey,<sup>175</sup> which we termed 'overload' and 'line management', teamwork and safety climate scales developed by Sexton and colleagues<sup>25</sup> and three individual items<sup>175</sup> concerning strained relationships and teamwork (see *Chapter 3, Combining indicators of safety climate*). The basket of well-evidenced and clinically acceptable markers of the quality of care focused on three conditions commonly encountered in research sites (separate markers for DUs and EDs; see *Chapter 3, Audit data collection*).

TABLE 38 Summary of comparisons

Comparison	Correlation	Agreement
Survey-based organisational and safety climates	Strong ( $r=0.845$ )	Good, SD of differences small at 0.449, no discernible pattern, one point (6%) outside 'limits of agreement', i.e. mean difference $\pm 2SD$ , one point at upper limit
Observation-based organisational and teamwork cultures	Low ( $r=0.356$ )	Reasonably good, SD differences 1.135, no discernible pattern, one outlier, tied scores evident
Survey-based organisational climate and observation-based organisational culture	Low ( $r=0.252$ )	Reasonably good, SD differences 1.089, one point outside limits of agreement, some clustering of points
Survey-based safety climate and observation-based holistic safety culture	Low ( $r=0.345$ )	Good, SD differences 1.059, no discernible pattern, one point on upper limit of agreement
Survey-based safety climate and observation-based teamwork culture	Low ( $r=0.316$ )	Good, SD differences 1.082, no discernible pattern, one point on upper limit of agreement
Survey-based organisational climate and criterion-based audit	Negligible ( $r=-0.096$ )	Reasonably good, SD differences 1.058, no discernible pattern, two points on limits of agreement
Observation-based organisational culture and criterion-based audit	Moderate negative ( $r=-0.481$ ) but concern that two sites had a disproportionate effect	Poor, SD differences 1.451, discernible pattern within limits of agreement and two points lie outside the limits
Survey-based safety climate and criterion-based audit	Low ( $r=0.150$ )	Reasonably good, SD of differences 0.992, no discernible pattern, one point on upper limit of agreement and one outside the limits of agreement
Observation-based holistic safety culture and criterion-based audit	Low, negative ( $r=-0.201$ ); concern about two outliers	Reasonably good, SD differences 1.314, two outliers visible but otherwise no discernible pattern, all points within limits of agreement
Observation-based teamwork culture and criterion-based audit	Negligible ( $r=-0.062$ ); two outliers visible	Reasonably good, SD differences 1.241, two outliers visible but otherwise no discernible pattern, all points within limits of agreement

TABLE 39 Examining the closeness of agreement

Comparison	SD of differences	Differences within range $\pm 0.5$
Survey-based organisational and safety climates	0.449	14 (88%)
Observation-based organisational and teamwork cultures	1.135	5 <sup>a</sup> (31%)
Survey-based organisational climate and observation-based organisational culture	1.089	9 (56%)
Survey-based safety climate and observation-based holistic safety culture	1.059	6 (38%)
Survey-based safety climate and observation-based teamwork culture	1.082	7 (44%)
Survey-based organisational climate and criterion-based audit	1.058	6 (38%)
Observation-based organisational culture and criterion-based audit	1.451	3 (19%)
Survey-based safety climate and criterion-based audit	0.992	9 (56%)
Observation-based holistic safety culture and criterion-based audit	1.314	5 (31%)
Observation-based teamwork culture and criterion-based audit	1.241	7 (44%)

a Coincident points lie at  $-0.7$ .

## Part III

# Discussion, conclusions and recommendations



## Chapter 9

### Discussion

This chapter will revisit the study's aims, objectives and hypotheses, evaluating what was learnt from the three strands of data collection and associated comparisons. We then consider the extent to which the empirical results might be considered predictable. A range of challenges that were encountered during this study, and have implications for future research studies, are then discussed. Finally, the strengths and weaknesses of the study are discussed.

High-hazard industries such as health care have focused on predictive measures of safety. One particular focus has been the evaluation of 'safety climate', a term that generally refers to the measurable components of 'safety culture' such as management behaviours, safety systems, and employee perceptions of safety (see *Chapter 2*, particularly *Organisational and safety cultures – concepts and definitions*). This has prompted the development of many checklists and questionnaires aimed at capturing facets of safety climate.<sup>21,140</sup> However, there is a paucity of research that explores the relationship between organisational culture and patient safety climate scores and process measures theorised to be associated with improved patient outcomes. It is thus premature to anticipate that measures of patient safety climate reliably indicate patient safety outcomes.

Of the survey instruments listed in *Chapter 2, Structured questionnaire research instruments*, the SAQ<sup>24</sup> has been used most extensively to explore the relationship between safety climate scores and patient outcomes. During the SAQ's development and benchmarking,<sup>24,25</sup> favourable scores were associated with shorter lengths of stay, fewer medication errors, lower ventilator-associated pneumonia rates, and lower bloodstream infection rates. Sexton<sup>208</sup> also found that favourable scores were associated with lower risk-adjusted patient mortality rates. However, there is a need for more studies to investigate how organisational and safety climates relate to the quality of care, and this study contributes to that gap, using two scales from the SAQ, teamwork climate and safety climate,<sup>25</sup> and extending exploration to include questions drawn from the annual NHS Staff Survey<sup>175</sup> (see *Chapter 3, Developing the organisational and safety climate questionnaire*).

Health-care staff surveys have been found to generate poor response rates, which are thought to be falling,<sup>29,30</sup> possibly because of workload pressures and the rising number of forms, checklists and surveys to be completed, giving rise to concerns about non-response bias, although bias is not inevitable when there are low response rates.<sup>209</sup> An additional concern in an era when the concepts embedded in safety climate measures have been the subject of high-profile quality improvement campaigns is the risk of social desirability bias. Indeed, the contemporary focus on ranking the performance of units may accelerate any trend towards socially desirable responses and, thus, undermine the usefulness of established safety climate measures. These concerns have resulted in increased interest in observational measures.

Although there has been growing interest in the potential of ethnographic studies relating to patient safety, ethnographic observations do not seek to measure in any quantitative sense and are not well suited to quantitative comparisons with, for example, measures of clinical outcomes. We support continued investment in ethnographic studies focused on patient safety, the benefits of which were well argued by Dixon-Woods,<sup>210</sup> but this study was commissioned to focus on the development of quantifiable observations. However, we wished to stop short of the checklist approach which has flourished in recent years,<sup>211</sup> but which has given rise to

concerns about the consequences of naively seeing checklists as a panacea.<sup>212</sup> At the opposite end of a spectrum to ethnographies, observation through inspection visits lasting a few days, as conducted by quasigovernmental bodies, is not an appropriate model for research. Such visits would be unacceptably disruptive to clinical departments and the idea of inspection would be counterproductive. Furthermore, the veracity of assessments made from brief inspection visits depends on the quality of preparatory work by members of the inspected department: researchers cannot ask members of busy clinical departments to undertake extensive and complex work on their behalf unless backfill for staff can be identified and funded. Consequently, the approach taken in this study was an intermediate position: semistructured observation and strategic immersion in the ebb and flow of work in the researched clinical departments. This produced quantifiable assessments of organisational and safety culture which could be compared with more established measures of organisational and safety climate. This is the first study to attempt this methodological innovation and the associated comparisons.

This study operationalised culture and the quality of care as follows:

- organisational climate and safety climate using a postal questionnaire survey of staff in purposively selected clinical departments
- organisation and safety culture using semistructured observation and strategic immersion in the clinical departments for approximately 12 half-days
- the quality of care using retrospective, criterion-based audits of clinical notes using markers drawn from national clinical standards.

While prevalidated questions and scales were available for the postal questionnaire, no established method of obtaining quantifiable, time-limited yet holistic observation-based evaluations of culture existed. An observation framework and scoring system was developed during this study, which will need further refinement and testing in different contexts.

After obtaining separate assessments of climate, culture and the quality of care these contrasting patient safety indicators were compared in pairs. This is best illustrated by revisiting each of the multifaceted aims, objectives and hypotheses examined by this study.

## Revisiting the study's original aims, objectives and hypotheses

### Aims

The study's aims were to compare:

- questionnaire and holistic (observation-based) assessments of organisational and safety climate/culture and
- these assessments of organisational and safety climate/culture with criterion-based assessment of the quality of care.

These comparisons were made and the empirical results were reported in *Chapter 8*. Whether such comparisons are legitimate and useful are matters of epistemology and practicalities. The study's three strands are epistemologically and empirically different:

- A self-reported perceptions, recorded in a staff survey
- B non-participant observation made mainly by non-clinical health services researchers, supported by nurses, midwives and a service-user and
- C criterion-based, retrospective case note audit by non-clinical auditors.



Strand A staff perceptions may be viewed as qualitative data, which the survey participants quantified as they responded to the Likert scales in the questionnaire. Strand B observations were recorded in semistructured field notes and contained qualitative evaluations of behaviours, workplace practices and facets of the physical environment that were pertinent to safety (see data collection prompt list, *Appendix 5*). Researchers subsequently quantified these qualitative evaluations using the approach described in *Chapter 4, Quantifying semistructured observations* (which will be discussed in *Tensions and challenges in quantifying observational data*). Case notes contain a mixture of qualitative and quantitative records made by health-care staff describing observations, requests and decisions made during the care process, and actions taken. Criterion-based audit (strand C) records the presence, partial presence or absence of selected indicators of the quality and safety of care, thus synthesising qualitative and quantitative notes into quantitative data.

The three strands of this study powerfully illustrate the process of reducing qualitative evaluations to quantitative representations to facilitate quantitative analyses and comparisons, but it is acknowledged that the reductionism in this process discards or disguises the nuances that aid understanding of the numerical results. For this reason, in earlier work,<sup>132</sup> we have treated qualitative and quantitative observations of workplace practices as complementary to support a better-elaborated understanding than either approach delivered alone. In contrast, the purpose of this study was to 'triangulate' (commissioning brief, see *Appendix 1*) in the sense of examining the degree to which different data sets provide the same understanding of workplace practices.

In multimethod research<sup>213</sup> (sometimes termed mixed methods<sup>214</sup> research) integration of different types of data is most commonly thought of in terms of complementarity, and Brannen<sup>215</sup> (p. 13) notes that 'differences between data sets are likely to be as illuminating as their points of similarity'. Because this study was commissioned to test similarities, it tends to position differences as problems. However, we continue to view differences as illuminating and worthy of further research. The observation field notes from strand B provide illumination of the distribution of points in *Figures 9–24* which cannot be included in this report but which we hope to publish elsewhere.

## Objectives

Objective 1: To work *with* staff in the participating trusts and user stakeholders such that their organisational and professional knowledge is respected, the study is understood and supported within participating departments, and prompt feedback to participant departments allows local development in advance of the study reporting.

Examples of fulfilling this objective include asking clinical leads at each research site to define the categories of staff to be included in the staff survey, with minimal exclusions imposed by the research team (see *Chapter 3, Sampling, delivery and maximising returns*). Early feedback was offered to clinical teams at all (16) research sites and taken up by 10. Feedback included additional contextual data for clinical teams to support local quality improvement, but not included in this report as they were not necessary for the methodological comparisons which form the report's focus. According to local preferences, feedback was sometimes provided to a small group of senior staff and sometimes presented during a naturally occurring meeting of the wider clinical team.

Objective 2: To use questionnaires to obtain quantitative assessments of the organisational and safety climate at each site.

Questions from the NHS annual staff survey<sup>175</sup> and teamwork and safety climate scales<sup>25</sup> formed the staff survey questionnaire used in this study (see *Chapter 3, Developing the organisational and*

*safety climate questionnaire*). Prevalidated scales were first examined separately, then weighted averages were calculated for the responses to questions selected to elicit aspects of organisational climate and safety climate, as described in *Chapter 3, Combining indicators of organisational climate* and *Combining indicators of safety climate*. This process created summary survey-based assessments of organisational climate and safety climate. The questionnaire was shortened after slightly disappointing response rates at research sites 1 and 2, and consequently fewer questions contributed to the summary survey-based assessment of organisational climate than to the summary survey-based assessment of safety climate. The former comprised perceptions of senior management, organisational communication and error wisdom. The latter comprised perceptions of overload, line management, team work and safe care. The results of the staff survey were reported in *Chapter 5*.

The staff survey yielded variable response rates, although less variable than elsewhere.<sup>216</sup> The overall response rate was 27.6% and the range of site-specific response rates was 9–47% [interquartile range (IQR) 23–36%]. Extra measures (beyond salience, basic design, piloting, provision of addressed response envelopes, incentive in the form of a prize draw and a reminder after 2 weeks) were taken to improve response rates, including shortening the questionnaire and offering an online version of the survey in parallel with the postal version. Neither of these measures had a discernible effect on response rates, which continued to vary unpredictably. However, 531 completed questionnaires were received and analysed for this report, which includes comparisons to gauge whether the study respondents were broadly representative of the NHS workforce as a whole (see *Chapter 5, Demographics and role-related characteristics*, and objective 8 below).

Although the overall response rate was similar to other UK patient safety studies,<sup>100,178</sup> at each research site the study response rate was lower than trust-wide response rates for the longer NHS annual staff survey (see *Chapter 4, Samples and response rates*). However, the small, trust-wide samples of staff randomly selected to receive the NHS national staff survey are more strongly encouraged to complete and return it than would be regarded as ethical in research studies. Encouragements include multiple reminders and appeals from all levels of management, often including the trust chief executive. Similarly, some patient safety studies report considerably higher response rates than those obtained in this study.<sup>96,106,217</sup> In these studies, questionnaire distribution was managed by hospital staff who took the role of champions for the studies and were active in increasing response rates. This raises concerns about coercion. For example, Holden and colleagues<sup>218</sup> obtained response rates of 59.6% and 54.2% in a two-hospital study, but potential participants were handed an individually addressed survey pack during in-service training, staff meetings or shift handovers. The pack contained a small cash incentive and was followed up with three reminders over a 3- to 4-week period.

Site-specific response rates were included as possible explanatory variables in multilevel modelling of factors that may mediate scores on the prevalidated scales included in this study's survey. Response rate was not a significant factor. This echoes findings of another recent study,<sup>216</sup> which experienced even more variable response rates to the (hand-delivered) SAQ<sup>25</sup> from 63 surgical departments (range 5–100%, IQR 25–64%). Watts and colleagues<sup>216</sup> included response rates in a regression analysis and found that response rates were not significantly correlated with scores on any of the six SAQ domains (two of which, teamwork climate and safety climate, formed part of this study's summary survey-based safety climate measure).

In our survey, responses to prevalidated scales drawn from the annual NHS staff survey were normally distributed, but there was some evidence of a degree of negative skew (a high proportion of positive responses) within responses to the teamwork and safety climate scales.<sup>25</sup>

These teamwork and safety climate scales, developed in the USA and benchmarked more widely,<sup>96</sup> address matters such as the coordination of different professional teams' contributions to care, speaking up, questioning, resolving conflict, responding to errors and personal feelings about the safety of care. In recent years these matters have become basic tenets of quality improvement initiatives and a strong focus within initial professional education and CPD for health-care staff.<sup>219–221</sup> This could have sensitised respondents to 'correct' or socially desirable answers. The slight skew towards positive responses for the Sexton teamwork and safety climate scales could be limited evidence of improvements in these aspects of safety culture or limited evidence of social desirability bias.<sup>26</sup> Further research will be needed to disentangle these matters.

Objective 3: To generate quantified holistic evaluations of organisational and safety culture for each site using observation.

No established method of data collection existed for objective 3, so a suitable framework and scoring system was developed during the study. This was achieved using semistructured observations and strategic immersion in the ebb and flow of activity within the purposively selected clinical departments. Two experienced research fellows separately conducted non-participant observations at each research site during six half-day visits. In recognition of fluctuating workloads and staffing, the visits included 6 days (Monday–Saturday) and time sampling in the period 6 AM to midnight (see *Chapter 3, Data collection*, for further details of the non-participant observation). At eight research sites additional 1-day visits were made by service provider researchers. This provided a clinical gaze to complement the non-clinical researchers' understandings of what was being observed. This increased researchers' confidence that they were not overlooking or misunderstanding important features of the workplace activity. At sites where service provider researchers were unable to attend, research fellows discussed their observations in detail with service provider researchers. Selected issues were also discussed with the project's senior clinical advisors.

At three research sites, a service user observer made observations during one half-day visit. Her observations were valued by the research fellows to check that they had not 'gone native', in the anthropological sense, through gradually developing too close an identification with the clinical perspective. The service user observer also proved exceptionally good at focusing on safety-related non-technical skills<sup>12</sup> exhibited by staff during their work. A service user observer could only be present in three research sites, owing to difficulties in recruitment of interested individuals and subsequent difficulties in securing research governance approvals for their participation (see *Chapter 4, Recruiting service user and service provider coresearchers*). Plans to streamline research governance processes and speed recruitment to studies may help future studies.<sup>196,222</sup>

The value of ethnography in patient safety research is recognised:<sup>210</sup> our use of strategic immersion is less established. The strengths of the semistructured observations and strategic immersion used in this study include economy and unobtrusiveness. Compared with much more prolonged ethnography, strategic immersion allows less time (in the form of fewer visits) for staff to become accustomed to the presence of observers such that the latter become 'invisible' so, in theory, reactivity to researchers may be present in a higher proportion of the observation data set. However, reactivity to researchers subsides very quickly in busy clinical environments: well within the half-day observation periods used in this study. Making more or fewer visits would not be expected to change the extent of researcher reactivity because staffing for each shift would be different. Limitations of less intensive observation included fewer opportunities to have informal conversations with members of the clinical team, which may have added more texture and greater discrimination to the observations made.

Sequential observation at the eight participating hospitals allowed researchers to form, test and consolidate observation-based assessments of cultures within each research site with the minimum risk of contamination from observations at other sites.

Once observations were complete at all research sites, two research fellows separately read all observation materials and notes from associated orientation interviews and feedback meetings. This included personal data collection and data collection by the other research fellow, service-user and service provider coresearchers. Each research fellow ascribed scores for every research site, using a four-point scale across eight categories. The categories and scoring system were developed during this study (see *Chapter 3, Analysis*, and *Appendix 6*). After independent scoring, the research fellows met to compare scores, discuss discrepancies and agree final scores. Fifty-eight per cent of the research fellows' independent scores were in agreement. All scores were agreed after a short discussion. Five-point and six-point scoring systems were also trialled, but this study's research fellows found it impossible to make confident and consistent assessments with any greater level of discrimination than the four-point scale reported here. As a result of this and the process of adding scores across eight categories (three relating to organisation and work environment factors and five relating to team factors), the scores for several research sites tied. This was a particular problem for the organisation and work environment factors. In future research it may be possible to obtain greater discrimination by narrowing the scope of what is observed. That would represent a trade-off between greater discrimination among observed clinical departments but narrower understanding of facets of safety culture.

Researchers were cautious about ascribing scores for research sites 1 and 2, where observations were made by two project research fellows who then left the study. One of the former research fellows joined discussions, and minor doubts about scoring were soon resolved. For future research this highlights that, ideally, scoring should be conducted by those who observe. However, detailed input from the former research fellow was only essential for 2 of the 16 scores required for sites 1 and 2, demonstrating that experienced researchers, using a sufficiently structured scoring system, can score from observation records made by others. Where staff changes are inevitable, a variety of mechanisms may help increase confidence in scoring, for example ongoing contact between outgoing and incoming research staff, as in this study; review and annotation of observation records by departing staff at the time they leave the study (done in this study); a formal handover and discussion of observation records; and more structured observation records (something which would have been difficult in this particular study as this strand was mainly exploratory to investigate whether semistructured observation could yield trustworthy quantifiable evaluations of visible aspects of safety culture).

How much is sufficient observation to inform a secure assessment of organisational and safety culture? This is equivalent to asking the broader research question regarding data saturation in qualitative research. In qualitative studies diminishing returns from continued data collection prompt withdrawal from research environments, but there can never be a guarantee that the next observation visit would not have revealed something important and hitherto unobserved. Having noted this caveat we will reflect on the quantity of observation visits conducted during this study. We will see later that the observation vantage points used in this study did not achieve sufficient purchase on the matter of organisational culture. More or longer observations from the same vantage points would not have remedied this. Aided by earlier discussions with service provider coresearchers, the non-clinical health services researchers became confident about their evaluations of team-level factors relating to safety culture. A total of 12 half-day visits (six per researcher) was sufficient for confident scoring. Scoring after fewer visits, say four or five per researcher, would have been possible once researchers had gained confidence in observing these particular clinical environments. However, reduction in the number of visits reduces the sampling of days of the week and times of the day. We believe that it is important to collect data

in the week (Monday–Friday) and at weekends, during ‘office hours’ (9 AM to 5 PM) and at other times of the day and night, at the beginning, middle and ends of shifts (noting that different staff groups have different shift patterns).

Theoretically, there was a risk of a Hawthorne effect: staff providing care more safely than when unobserved. However, clinicians’ work is routinely observed by others, including patients and their visitors; official visitors to the hospital; staff from other hospital departments; junior doctors on rotation; agency and locum staff; and patients, visitors, students and colleagues. Our experience was that busy clinicians quickly appear unaware of or unperturbed by researchers conducting observations. We looked for but saw no evidence of staff concern about being observed, although at one site there was concern that we could overhear discussion of named patients. Monahan and Fisher<sup>223</sup> point out that, even if there is an undetected Hawthorne effect, this does not render observations worthless: if members of the clinical department ‘stage a performance’ then this reveals their perceptions of the performance that ought to occur. If the staged performance demonstrates bad practice then that provides quite robust evidence of low safety standards.

The tender specification (see *Appendix 1*) acknowledged the possibility that it may not prove possible to obtain distinct observation-based evaluations of organisational and safety culture and this proved to be the case (see *Chapter 9, Limitations of this study*). It was possible to assess three categories of organisation and work environment factors and five categories of safety-related teamwork factors (see *Appendix 6*). We were able to gather evidence of the adequacy of staffing, premises and equipment, the quality of support offered by administrators, senior managers and other related teams, and key aspects of teamworking: task management, decision-making, cooperation, coordination, leadership, communication, mutual performance monitoring, back-up behaviour, adaptability/flexibility and team/collective orientation.<sup>11,12</sup> We could observe the ‘active sharing and updating of knowledge, enabling risks to be collective and progressively monitored’;<sup>224</sup> the maintenance of situational awareness;<sup>34,74,75</sup> and other safety-related behaviours. The observations made during this study were consistently validated by staff when researchers reported back to departments, suggesting that the method developed is effective in capturing core team characteristics.

Objective 4: To obtain criterion-based measurements for the quality of care at each site.

This was achieved using a retrospective criterion-based audit of case notes, conducted by non-clinical auditors. Three common conditions, which had national guidelines for care, but had not been the subject of recent national audits or quality improvement campaigns, were selected. The conditions audited in delivery suites were normal labour and delivery, ECS and care following detection of grade 2 or grade 3 MSL. In EDs the audited conditions were ACS, ASA and FNoF. For each condition between three and five evidence-based audit standards were selected (see *Chapter 3, Strand C: criterion-based assessment of quality of care*). Some standards concerned treatment interventions, whereas others concerned recording important observations.

The study design planned sequential audits at the eight participating hospitals over a 2-year period, auditing consecutive cases and working backwards for no more than 6 months from the date observers first became noticeable in the clinical department. This plan addressed several points of general methodological interest. Sequential rather than parallel data collection at eight research sites was planned for efficiency; the study was feasible with two half-time research fellows. The selected time period simultaneously addressed concerns about a possible Hawthorne effect and the need to treat the survey, observation and audit data as contemporaneous in subsequent data analysis. However, sequential data collection requires vigilance with respect to the publication of revised care standards. It was important to conduct the case note review using

the relevant standards as they were at the time the clinical records were made. Using personal research networks and seeking advice from the study's senior clinical advisors, the study team had taken care to learn as much as possible about any likely revisions to standards before the final selection was made. Happily, only 1 of the 22 audit standards (ASA, see *Chapter 7, Acute severe asthma*) changed unexpectedly<sup>182,207</sup> and particular care was taken with auditing this condition. Vigilance with respect to changes in national care standards is an important feature of audit-based research.

The audit results exhibited substantial variation between standards. For example, there was a very high rate of compliance with standard ECS4, 'Women having an emergency caesarean section should be offered regional anaesthesia (spinal or epidural)' (range 92–100%), whereas compliance with standard ECS1, 'Documentary evidence of consultant obstetrician involvement in the decision to ECS', was lower and more variable (range 55–84%). There was also substantial variation in compliance within audit standards; for example, compliance with the standard ACS3 (below) varied between 7% and 61% at different research sites.

If patient in pain, times from arrival to the administration of pain relief:

- in cases of severe pain, 50% in 20 minutes, 75% in 30 minutes, 98% in 60 minutes
- in cases of moderate pain, 75% in 30 minutes, 98% in 60 minutes.

Both between-standard and within-standard variation provide challenges for audit-based research. Safety-related markers of the quality of care are not all equally important: poor scores for certain markers may have disproportionately serious consequences. Therefore, it is important to keep variations in performance within individual standards visible in the reporting of results, as we did in *Chapter 7*. Nevertheless, some situations require an aggregate view of the quality of care, which was the case in this study. This is achieved by auditing a 'basket' of care standards so that the aggregation of results smooths out unusual results for one or two standards to avoid results which are simply artefacts of the care standard or narrow range of standards chosen. The greater the number of standards, the greater the smoothing: Hutchinson and colleagues,<sup>156</sup> for example, used much larger numbers of standards than this study.

Power calculations for this study showed that nine audit standards were required for 80% power at 5% significance,<sup>174</sup> hence the selection of a minimum of three standards for each of three common conditions for each type of site. Where there was clinical concern that certain standards may not differentiate between departments because of near-universal compliance or very low levels of compliance, additional audit standards were identified to mitigate the problems. The standards selected for this study were selected from those identified as having strong underpinning evidence, each was considered clinically important and they were equally weighted in analyses (*Chapter 3, Strand C: criterion-based assessment of quality of care*).

Care standards requiring the recording of repeated observations generally showed low levels of compliance. In some situations, such as FNoF a single follow-up observation FNoF5, 'If in pain, evidence of re-evaluation of pain (90% of those with severe pain re-evaluated within 30 minutes; 75% of those in moderate pain within 60 minutes)', exhibited compliance levels in the range 0–4%. It is not known whether this represents poor care or poor recording of care. For care standards such as MSL4 (baby assessment should include at 1 and 2 hours and then 2-hourly for 12 hours: general wellbeing; chest movements and nasal flare; skin colour including perfusion; feeding; muscle tone; temperature; heart rate; and respiratory rate) compliance was poor, in this case in the range 0–6%. A similar situation was found with the care standards for repeated observations during normal labour and delivery. This raises questions about the national care standards themselves. Possible explanations include that the care standards may not have

widespread clinical support owing to considered professional judgement or lack of knowledge of the detail of the standard; staff may find the workload of making and/or recording multiple and repeated observations too high, signalling a need to prioritise those which are most important. Dixon-Woods and colleagues<sup>125</sup> discuss how rules in practice are so numerous that they quickly exceed individuals' ability to act on them. Staff then have to choose which rules to ignore, and may also ignore others without realising that they are doing so: partial compliance becomes the accepted 'normal-illegal' state.<sup>225</sup>

The recording of multiple and repeat observations was also influenced by site-specific record forms. The structure of these enhanced or limited the recording of pertinent clinical information, depending on the design of the form. It is possible that structured record forms also prompt the completion of included clinical actions or, by the same token, could direct attention away from other clinical actions.

Some audit-based research, including this study, seeks to differentiate research sites by audited performance against standards. Where the national care standards failed to differentiate between departments, because compliance was very high or very low overall, researchers can still discern variation in performance. In the cases where audit standards detect uniformly low levels of compliance, it may be useful in future to devise a modified way of measuring performance. For example, in the case of ND (all three standards), this could be done by selecting a smaller number of observations agreed to be priorities by an expert panel, although this introduces the risk of different studies selecting different subsets of observations and compromising the comparability of study findings. In the case of treatment or tests that should happen within a specified time (ACS2, ASA2 and ASA3), the more inclusive measure could include moderately but not very late performance of these tasks: again, consistency between studies in the definition of 'late' would facilitate comparison of results. When standards failed to discriminate between sites because they were met in the majority of cases at all sites, such as the offer of local anaesthesia for ECS, it is not possible to increase discriminatory power in the same way and a replacement audit standard would need to be identified.

Audit standards such as MSL4, above, highlight the close-coupling or interdependent nature of clinical departments and teams,<sup>139,144</sup> which we highlighted in *Chapter 4, Close-coupling of departments and services*. The MSL4 observations must begin in the DU but would normally be completed on a post-natal ward. Similarly, standard FNoF3, 'Radiography performed within 60 minutes of arrival in 75% of cases', does not entirely lie within the control of the emergency department team. Audit-based research needs to address the complexity of overlapping responsibilities or flexible boundaries. Challenges may be less if the unit of analysis is a care pathway, which might be viewed as a control system,<sup>226</sup> but sometimes team- or department (subsystem)-level evaluations are required and boundaries of responsibility and control must be considered. However, this is difficult, as Osman<sup>226</sup> (pp. 122–3) cautions: 'Reasoning about a network of decisions with respect to a single goal is imperative in order to monitor the behaviour of the system.' But dynamic information exchanges across social and technical networks, which form and link subsystems, generate multiple decisions that will vary in extent and predictability in their influence on the particular goal being investigated in the evaluation of a particular subsystem. We may not be able to fully disentangle the effects of contributions (and delayed or missed contributions) from closely coupled departments or teams when auditing the performance of a particular department or team in a wider system.

In this study non-clinical auditors were preferred because earlier research found this reduced hindsight bias,<sup>186</sup> and more recent research has found that non-clinical auditors are no less reliable than clinical.<sup>156,160</sup> This was by far the most challenging aspect of this study. The current climate in research governance places such severe restrictions on audits by non-clinical auditors

that future studies using this methodologically sound and low-risk research approach seem unlikely. Extremely busy staff within the care team do not have the capacity or inclination to provide anonymised data for research studies. It is a moot point whether the small number of trust-based central audit clerks are members of the care team as defined by the NIGB, although trusts participating in this study all viewed access to medical records by these staff as unproblematic. By the same token, most trusts participating in this study believed that, as researchers were in any case already screened with all the employment checks conducted for trust staff and, after clearance, awarded honorary contracts, they should be able to conduct audit-based research. Recent guidance<sup>196</sup> regarding increased access to patient data in health services research may ease some of the difficulties experienced during this study. However, the emphasis on restricting access to clinical records remains. Furthermore, the main focus of current policy<sup>223</sup> and initial developments is on increasing and speeding recruitment to clinical trials; improvements for audit-based research may take longer to achieve.

A separate trend that occurred over the duration of this study was improved arrangements for reimbursing trusts for research support costs, which has eased the participation of NHS staff in research studies, but, once again, the emphasis is on work with consented patients for clinical studies. The extraction of anonymous audit data is still poorly supported. When this study sought advice from the NIGB the advice indicated that trust provision of anonymised data was considered feasible and appropriate; a second approach would be for trusts to send a letter to each patient requesting individual consent before allowing researchers to audit records. This study had already established that members of the care team could not provide anonymised data. The individual consent route was not feasible for this study because electronic records are not yet sufficiently well developed to act as anything more than a guide to identify cases that meet the specified audit criteria. Paper and electronic case notes have to be retrieved and read to establish eligibility for inclusion in the audit before letters can be targeted appropriately; in some cases, the preliminary screening alone equates to the level of work required to audit the case notes. This study was completed following lengthy efforts by trusts to identify permanent or temporary clerical staff to extract audit data. In one trust where all such efforts failed, a health-care assistant was engaged to complete the task. Arguably, the use of carefully screened non-local auditors is more ethical than the use of local non-clinical auditors, since the non-local auditors are far less likely to know any of the patients or staff personally.

A second methodological difficulty and ethical concern with excessive restrictions on case note review is that data will be disproportionately lost from members of the most mobile populations, including travellers, refugees and people in temporary accommodation; people with poor literacy or poor understanding, whether due to unidentified or unmet translation needs, general reading difficulties or cognitive impairment; and people whose deteriorating health has precipitated a move from the home they had at the time of their admission. The cumulative effect of these losses is to bias results, and there is already concern that the experiences of minority and vulnerable groups are insufficiently captured by current NHS data collection.<sup>227</sup>

Objective 5: To compare levels of agreement between the questionnaire and holistic measurements of climate/culture.

This objective was linked to two hypotheses:

H1a: There will be a strong correlation and good agreement between questionnaire-based and holistic evaluations of *organisational* culture;

H1b: There will be a strong correlation and good agreement between questionnaire-based and holistic evaluations of *safety* culture.



These comparisons were reported in *Chapter 8*, and *Table 40* reproduces the relevant rows from *Table 38* with additional information on the closeness of agreement, extracted from *Table 39*. For the three comparisons between survey-based assessments of climate and observation-based assessments of culture low levels of correlation were found, indicating that there is no linear relationship between these measures.

Agreement addresses the closeness of two measurements made on the same scale. In this study, all summary scores were scaled to the standard normal distribution before comparisons were made. *Table 40* shows reasonably good agreement between survey-based organisational climate and observation-based organisational culture scores. The SD of differences between scores was 1.089 and the majority (56%) lay in the range  $\pm 0.5$  points. The Bland–Altman plot for this comparison, *Figure 10*, drew attention to some clustering of points. This was caused by lack of discrimination in the observation-based organisational culture measurement. Although these two measurements were mostly very similar, the survey-based organisational climate measure is preferred. This is because of the disadvantage of poor-to-moderate survey response rates being considered less serious than the limited facets of organisational culture that could be observed in this study (see objective 3 above) and concerns about lack of discrimination in the scoring of the observation-based organisational culture measure.

The prevalidated scales from the NHS annual staff survey, which contributed to this study's summary organisational climate score, were also compared with observation-based organisational culture scores; neither exhibited better correlation or agreement than our summary measure.

*Table 40* shows that good agreement was found between scores for this study's summary measure of safety climate and observation-based teamwork culture scores (five factors) and also the observation-based holistic safety culture scores (three organisation and work environment factors and the five teamwork factors from the previous measure). The SDs of differences were similar, 1.082 and 1.059 points, the distribution of points on the Bland–Altman plots (*Figures 13* and *14*) showed no discernible patterns, and around 40% of differences between measurements were small (within  $\pm 0.5$ ). This suggests that survey-based and observation-based assessments of safety climate/culture may be considered equivalent. Adequately trained and experienced non-clinical researchers, who made 12 visits to clinical departments at purposively sampled times to conduct semistructured observations lasting 4–6 hours, made assessments of safety culture in clinical departments at similar levels to the perceptions of people working in these clinical departments, as reported in responses to questionnaire items drawn from the NHS annual staff survey<sup>175</sup> and Sexton and colleagues<sup>25</sup> teamwork and safety climate scales. Thus, observation-based assessments of safety culture using the framework and scoring system developed during this study might be a useful alternative to staff surveys of safety climate, particularly as survey response rates appear to be falling and concerns about social desirability bias have grown.

An additional use for the observation-based assessment framework and scoring system developed during this study would be to complement staff surveys: the two measurements may be similar for most sites, but *Figure 14* shows one research site on the upper limit of agreement. Here, staff survey responses indicated a much poorer climate than observer evaluations of teamwork culture. This multimethod study was able to trace this result back through field notes and survey responses to discontent with management style, despite generally well-executed teamwork. Using the contrasting measures in this complementary way was not envisaged in the commissioning of this study, but is a strength of multimethod studies.

**TABLE 40** Summary of results from comparisons relating to H1a and H1b

Comparison	Correlation	Agreement
Survey-based organisational climate and observation-based organisational culture	Low ( $r=0.252$ )	Reasonably good, SD differences 1.089, one point outside limits of agreement, some clustering of points, nine (56%) differences in range $\pm 0.5$
Survey-based safety climate and observation-based teamwork culture	Low ( $r=0.316$ )	Good, SD differences 1.082, no discernible pattern, one point on upper limit of agreement, seven (44%) differences in range $\pm 0.5$
Survey-based safety climate and observation-based holistic safety culture	Low ( $r=0.345$ )	Good, SD differences 1.059, no discernible pattern, one point on upper limit of agreement, six (38%) differences in range $\pm 0.5$

The prevalidated scales that contributed to this study's summary survey-based safety climate measure were also separately compared with the observation-based assessments but the summary measure provided superior agreement.

Objective 6: To compare organisational culture with safety culture.

This objective was linked to the following hypotheses:

H2a: There will be a strong correlation and good agreement between *questionnaire*-based evaluations of organisational and safety culture;

H2b: There will be a strong correlation and good agreement between (quantified) *holistic* evaluations of organisational and safety culture.

Table 41 reproduces the relevant extracts from Tables 38 and 39. It shows that the survey-based measurements were strongly correlated ( $r=0.845$ ), meaning that a linear relationship exists between scores on these measures. Agreement between the scores on these measures was good, with 88% of differences lying in the range  $\pm 0.5$  points. Clinical departments in which staff rated safety climate favourably also returned favourable organisational climate scores. Further research would be needed to examine whether one favourable assessment is antecedent to the other. Zohar and Luria<sup>142</sup> examined this in the context of manufacturing plants. They found that organisation-level and workgroup-level climates were globally aligned (as was the case in this study). In the manufacturing context, multilevel modelling revealed that supervisory practices varied between workgroups, producing variations in safety behaviour, but organisational climate set limits on variations in supervisory behaviour and, therefore, on safety behaviour within workgroups. It is not known whether this finding would be replicated in health-care contexts.

The comparison between observation-based organisational and teamwork culture scores (rather than organisational and safety cultures scores) was made because it did not prove possible to develop an observation-based holistic safety culture measure that was distinct from the observation-based organisational measurement. The comparison between organisational and teamwork cultures found low correlation ( $r=0.356$ ) and a lower level of agreement than was found in the previous comparison between survey-based measures. The SD of differences for observation-based scores was 1.135 points (compare with 0.449 for survey-based scores), and only five (31%) differences were small (within  $\pm 0.5$  points). As previously discussed, in this study, observation-based assessment of organisational culture was felt to be limited. We are more inclined to state that the methods of observation and scoring did not allow sufficient purchase on organisational culture than to state that the survey-based assessments produced different results from the observation-based assessments. Time-limited non-participant observation in clinical settings cannot adequately assess organisational culture. Staff surveys remain an inexpensive form of assessment, notwithstanding concerns about falling response rates and non-respondent

**TABLE 41** Summary of key results from comparisons relating to H2a and H2b

Comparison	Correlation	Agreement
Survey-based organisational and safety climates	Strong ( $r=0.845$ )	Good, SD of differences 0.449, no discernible pattern, one point (6%) outside 'limits of agreement', i.e. mean difference $\pm 2$ SD, one point at upper limit, 14 (88%) differences in range $\pm 0.5$
Observation-based organisational and teamwork cultures	Low ( $r=0.356$ )	Reasonably good, SD differences 1.135, no discernible pattern, one outlier, tied scores evident, five (31%) differences in range $\pm 0.5$

bias. Continued investment in traditional ethnography would also support analytical purchase on organisational culture.

Objective 7: To compare culture measurements and criterion-based measurements of the quality of care.

This objective was linked to two hypotheses:

H3a: There will be a moderately strong correlation and reasonably good agreement between criterion-based measurements of the quality of care and, firstly, questionnaire-based and, secondly, holistic evaluations of *organisational* culture;

H3b: There will be strong correlations and good agreement between criterion-based measurements of the quality of care and, first, questionnaire-based and, secondly, holistic evaluations of *safety* culture.

Table 42 reproduces five rows from Table 38, with additional annotations reproduced from Table 39. These show the comparisons between assessments of climate or culture and assessment of the quality of care from the criterion-based audit. In each case correlations were low, less than  $\pm 0.2$ , or in one case moderate ( $r=-0.481$ ) because two outliers had a disproportionate effect. This study did not find a linear relationship between assessments of climate or culture and the quality of care.

Reasonably good agreement was found between the quality of care, represented by standardised criterion-based audit scores, and the following assessments of climate and culture:

- survey-based organisational climate (SD differences 1.058, with six (38%) differences in range  $\pm 0.5$ )
- survey-based safety climate (SD of differences 0.992, with nine (56%) differences in range  $\pm 0.5$ )
- observation-based holistic safety culture (SD differences 1.314, and five (31%) differences in range  $\pm 0.5$ )
- observation-based teamwork culture (SD differences 1.241, with seven (44%) differences in range  $\pm 0.5$ ).

It can be seen that the closest agreement was with the summary survey-based safety climate measure developed for this study, which had the lowest SD of differences (0.992), giving rise to the narrowest limits of agreement ( $0 \pm 1.98$ ), and also having the highest percentage (56%) of small differences ( $0 \pm 0.5$  points). Survey-based evaluation of safety climate provides the closest agreement with audit-based assessment of the quality of care. However, in studies in which an observation-based assessment is preferred, the observation-based teamwork culture scores provided closer agreement than the observation-based holistic safety culture scores (SD of

**TABLE 42** Summary of results of comparisons relating to H3a and H3b

Comparison	Correlation	Agreement
Survey-based organisational climate and criterion-based audit	Negligible ( $r = -0.096$ )	Reasonably good, SD differences 1.058, no discernible pattern, two points on limits of agreement, six (38%) differences in range $\pm 0.5$
Observation-based organisational and criterion-based audit	Moderate negative ( $r = -0.481$ ), but concern that two sites have a disproportionate effect	Poor, SD differences 1.451, discernible pattern within limits of agreement and two points lie outside the limits, three (19%) differences in range $\pm 0.5$
Survey-based safety climate and criterion-based audit	Low ( $r = 0.150$ )	Reasonably good, SD of differences 0.992, no discernible pattern, one point on upper limit of agreement and one outside the limits of agreement, nine (56%) differences in range $\pm 0.5$
Observation-based holistic safety culture and criterion-based audit	Low, negative ( $r = -0.201$ ), concern about two outliers	Reasonably good, SD differences 1.314, two outliers visible but otherwise no discernible pattern, all points within limits of agreement, five (31%) differences in range $\pm 0.5$
Observation-based teamwork culture and criterion-based audit	Negligible ( $r = -0.062$ ), two outliers visible	Reasonably good, SD differences 1.241, two outliers visible but otherwise no discernible pattern, all points within limits of agreement, seven (44%) differences in range $\pm 0.5$

differences 1.241 compared with 1.314, and 44% small differences compared with 31%). There is still much that is unknown about the link between clinical performance and measurements of safety climate or teamwork culture. These measurements may complement and illuminate some audit results, but audits of performance must continue.

This study found poor agreement between observation-based organisational culture scores and criterion-based audit scores. This was predominantly due to lack of discrimination in the observation-based measurement of organisational culture.

Objective 8: To collect data such that, where sufficient respondents exist within a category to protect anonymity, data can be explored by: stakeholder group (e.g. managers, midwives, nurses, doctors, allied health professionals, support staff) and level (e.g. management responsibility).

The demographic profile of respondents broadly reflected the wider NHS workforce with respect to gender and ethnicity. Most (89%) staff survey participants held clinical roles and 37% managed other staff.

Responses to prevalidated scales within the staff survey were investigated using multilevel modelling. A range of individual-level factors were found to mediate scores on one or more scales, including being a manager, gender, age, years of employment with the trust, professional group, ethnicity and participation in certain types of CPD (particularly having a mentor, but also online training, updating skills and knowledge, shadowing someone or receiving on-the-job training). This expands the range of factors investigated in previous studies.<sup>217,228</sup> The details of the mediating effects can be found in *Chapter 5, Factors influencing the indicators of organisational climate* and *Factors influencing the indicators of safety climate*, and a summary is provided in *Table 43*.

At site level, the type of service (DU or ED) mediated error wisdom scores (i.e. the effectiveness of systems for reporting and handling errors; DUs perceived higher organisational error wisdom) and response rates mediated line management scores (departments with above-average response rates returned lower line management scores). There was no statistically significant residual variation at hospital level, but there was considerable residual variation at the level of respondents nested within research sites after allowing for the effects of individual-level characteristics.

**TABLE 43** Summary of mediating factors

Prevalidated scale	More favourable scores (points) from <sup>a</sup>
Organisation <sup>b</sup>	People working for the trust for < 2 years (1.577) or > 15 years (0.552) and managers (0.033)
Error wisdom <sup>b</sup>	DU staff (0.038), age over 40 years (0.017–0.020), having a mentor (0.018), managers (0.017), updating skills or knowledge (0.017), online training (0.017) and shadowing someone (0.016)
Overload <sup>b</sup>	Having support rather than direct care role (0.112), < 3 years' employment in trust (0.068–0.070), selecting Asian or Asian British ethnic group (0.065) and having a mentor (0.061)
Line management <sup>b</sup>	Below-average response rates (0.289), doctors (0.081), updating knowledge or skills (0.062), managers (0.054), women (0.050), receiving on-the-job training (0.048) and having a mentor (0.046)
Teamwork climate <sup>c</sup>	Doctors (0.077), women (0.056), managers (0.036), > 10 years' employment with trust (0.050–0.065), updating knowledge or skills (0.039), having a mentor (0.038) and doing online training (0.038)
Safety climate <sup>c</sup>	Managers (0.049), updating skills or knowledge (0.048), age over 40 years (0.027–0.048), women (0.040), on-the-job training (0.033) and having a mentor (0.026)

a See Chapter 5, *Factors influencing the indicators of organisational climate* and *Factors influencing the indicators of safety climate*, for details of the reference group for each comparison.

b Healthcare Commission (2006).<sup>175</sup>

c Sexton *et al.* (2006).<sup>25</sup>

This indicates that scores on the scales selected to elicit perceptions of aspects of organisational climate and safety climate predominantly varied at the level of the clinical department, not at hospital level, reflecting the findings of other studies.<sup>137,143</sup>

Our findings broadly support existing research that shows a more positive assessment of safety climate by managers than by clinical staff.<sup>37,102,103,229,230</sup> The teamwork climate<sup>25</sup> results, but not other components of the survey-based safety climate measure, accord with other studies that found nurses' assessments of aspects of safety climate more negative than doctors' assessments.<sup>106,218,228</sup>

## To what extent might the empirical results be considered predictable?

There are a number of reasons why one might argue a priori that the results of the three different research methods would *not* agree.

First, one could argue that observation and survey methods look at structures that support safety, whereas quality measures look at the process of care. Although structure and process are key dimensions of quality, and are linked,<sup>231</sup> they are conceptually distinct, although operationally the distinction may become fuzzy.<sup>232</sup>

Second, one could argue that the quality and the safety of care are themselves subtly different concepts. Thus, one could hypothesise a service where routine care was of a high standard (quality and one aspect of safety), but where staff were unable to respond appropriately and swiftly to untoward and unusual events (another contrasting aspect of safety). This is not to deny the close links between the two concepts.

Third, one could argue that the quality of a service is the result of contributions of a whole system rather than a discrete unit subjected to observation or survey; thus, an ACS patient may be given aspirin by the ambulance crew, his or her GP or in a community hospital before arriving in ED, and, therefore, no account of one unit can be complete. Furthermore, the units thus coupled may have very different safety cultures and climates, as there is evidence of considerable variation within hospitals and specialties.<sup>137,143</sup>

Fourth, it could be argued that the requirements of a 'safety culture' are in fact contradictory. For example, for routine clinical practice to be safe a teamwide acceptance of and adherence to evidence-based standards of care is required, whereas appropriate team responses to unusual events may require a team that can be creative in problem-solving and is able to disagree and debate constructively. In her analysis of social formations, Mary Douglas<sup>233</sup> demonstrated how 'grid' (rules) and 'group' (social cohesion), though linked, do not go hand in hand in a simple way. Thus, our audit assesses the role of 'grid', whereas our surveys and observation methods pay more attention to 'group'. More specifically, studies of organisational culture suggest that contrasting cultures support different aspects of safety: although bureaucratic hierarchies are comfortable with predictability such as that which comes from reliance on evidence-based rules of care, more entrepreneurial and/or group-based cultures may be more strongly associated with a safety climate.<sup>59,103</sup>

Fifth, the scope and limitations of each method may reduce the degree to which they can be said to be looking at the same aspects of safety culture and climate. *Table 44* illustrates this, using the framework proposed by Vincent and colleagues<sup>56</sup> for analysing risk and safety in clinical medicine. Cells are shaded and marked Y or Y? where different measures seem likely to yield data about each item in the framework, although no method is able to yield complete data about any item.

## Challenges identified during this study

### *Limitations of administrative systems*

Hospitals' administrative systems varied widely and were often not well suited for interrogation in relatively simple ways. For example, every research site struggled to produce a staff list for the staff survey. Locating records for the criterion-based audit varied in difficulty. At one ED, inconsistency of coding meant that some FNoF cases could not be differentiated from other fractures on the electronically generated list of cases, so records had to be manually sifted first. Similarly, most electronic lists did not contain the information needed to determine whether cases were eligible according to our criteria (see *Table 1*). One would not expect administrative systems to be tailored for research rather than clinical use, but some of the database coding deficiencies identified during this study were also sources of irritation and inefficiency during the routine work of clinical and administrative staff. Deficiencies that hampered the criterion-based audit for this study equally hamper trust-based monitoring for quality improvement.

### *Recruiting and deploying service user coresearchers*

It was difficult to recruit service users interested in working as co-observers in clinical departments that were geographically spread across six strategic health authorities in England (see *Chapter 3, Selecting the sample of research sites*, and *Chapter 4, Recruiting service user and service provider coresearchers*): only two were recruited, and one withdrew before completing research governance screening. Recruiting local service user coresearchers was the logical alternative, but this study's sequential recruitment of research sites and tight timetable for data collection did not allow sufficient time for interested individuals to be identified and complete trust-based research governance screening. At four of the hospitals included in this study, research governance screening for the service-user coresearcher could not be completed before the data collection period ended, preventing her participation. The service user coresearcher did not attend meetings of the project steering group, initially due to late recruitment.

### *Tensions and challenges in quantifying observational data*

The process of creating the required observation scoring mechanism (see *Chapter 4, Quantifying semistructured observations*, and *Appendix 6*) was challenging, and a major part of the methodological innovation in this study: it therefore warrants extended discussion. In this

**TABLE 44** What do this study's data collection methods measure?

	Survey	Observation <sup>a</sup>	Notes audit
<b><i>Institutional context</i></b>			
Economic and regulatory context	N	N	N
NHS	N	N	N
Clinical Negligence Scheme for Trusts (CNST)	N	N	N
<b><i>Organisational/management factors</i></b>			
Financial resources and constraints	N?	N?	N
Organisational structure	N?	N?	N
Policy standards and goals	Y?	N?	N?
Safety culture and priorities	Y	N?	N
<b><i>Work environment</i></b>			
Staffing levels and skill mix	Y	Y?	N
Workload and shift patterns	Y	Y	N
Design, availability and maintenance of equipment	N	Y?	N
Administrative and management support	Y	Y	N
<b><i>Team factors</i></b>			
Verbal communication	Y?	Y	N
Written communication	Y?	Y?	Y
Supervision and seeking help	Y	Y	N
Team structure	Y	Y	N?
<b><i>Individual (staff) factors</i></b>			
Knowledge and skills	N?	N?	N?
Motivation	Y	Y	N?
Physical and mental health	N	N?	N
<b><i>Task factors</i></b>			
Task design and clarity of structure	Y	Y?	N
Availability and use of protocols	N	Y?	Y
Availability and accuracy of test results	N?	N?	N?
<b><i>Patient characteristics</i></b>			
Condition (complexity and seriousness)	N	N	N
Language and communication	N	N?	N?
Personality and social factors	N	N?	N?

N, this method will not yield relevant data; N?, this method may yield some relevant data; Y, yes, this method will yield relevant data; Y?, this method will probably yield relevant data.

<sup>a</sup> Interpreted strictly, that is, not using answers to questions during orientation interview.

Framework from Vincent *et al.* (1998)<sup>56</sup>: population characteristics omitted.

section, we first review the process that led us to the adopted scoring system and then discuss the rationale for this method.

Scoring did not begin until observation data were available from half of the study sites to ensure that no site had undue influence on the emergent scoring framework. Initially, we explored the possibility of awarding separate scores to items on the observation data collection prompt sheet (see *Appendix 5*). We chose a scoring system modelled on Likert scales: the highest score for strong and consistent evidence for the attribute of interest; the lowest for weak, rare or

non-existent evidence. Such a scale has been used for team audit purposes for safety culture.<sup>84</sup> For each aspect of safety culture, we tried to score separately at the organisational and team levels (to mirror the survey data). This process helped us to understand which facets of safety culture would be amenable to data collection and scoring, and the nature of difficulties that would arise.

The research fellows carrying out the scoring found it difficult to reach judgements that they both trusted. The main reasons for this were researchers' caution and differences between data sets, of four kinds:

- differences between sites, in current practice, facilities, etc.
- differences in what we were able to observe (see *Chapter 4, Observable activities*)
- differences between what was seen by different observers at the same site and
- differences in how each scorer categorised similar material.

For example, it was difficult to give convincing comparative scores to sites in relation to their use of the whiteboard when some sites had none, although the absence did not mean that teams did not communicate effectively.

Other possible sources of difference that when investigated did not prove to be matters of concern for this study were:

- differences between the composition of observation teams (see *Chapter 4, Recruiting service user and service provider coresearchers*)
- differences over time as observers became more experienced.

In fact, observations by service user and service provider observers were highly congruent with those of research fellows; and when research fellows later reviewed their earlier notes, they did not find significant differences from those made later.

Many of the agreed scores at this stage were unsatisfactory to the scorers, who felt them to be indicators of compromise rather than of consensus. A more parsimonious framework was sought, both as a route to facilitating consensus and to make any emergent scheme more accessible and economical for future users. We experimented with grouping aspects of safety culture from the data collection prompt sheet (see *Appendix 5*) into a smaller number of domains, under two overarching headings: 'organisation and work environment factors' and 'team factors'. After several iterations and trials of scoring, the scoring system used (see *Appendix 6*) was agreed. The use of broader domains enabled the scorers to look at subsets of data clustering round domains rather than try to interpret small pieces of data separated from their context. For example, communication as a whole was scored rather than the use of the whiteboard in isolation. Using this method, each scorer found it easier to reach an individual judgement in which he or she had confidence, and together he or she found it quicker to reach consensus in those instances when he or she disagreed. This continued to be the case when applied to data from later sites. It is worth noting that the results were sometimes counterintuitive: systematic reference to the notes in these cases generated domain scores different from 'off-the-top-of-the-head' scores. Indeed, scorers sometimes felt the need to double-check the data records to ensure that a general impression had been mistaken. Such counterintuitive results strengthened our confidence that our scores were not impressionistic but grounded in recorded data.

Using a smaller number of domains also made agreement between scorers easier to achieve. First, the use of eight, rather than 32, categories reduced disagreement by removing the likelihood of disagreements about definition. Whereas scorers would debate the boundaries between two or more of the 32 categories, this did not happen when using eight. Second, it became much easier



to reach consensus when the categories were more inclusive; for example, disagreements at one site about 'unprofessional respect and collaboration' reflected different exposure of observers to the dominating behaviour of a single consultant. These were easier to resolve when this category was merged with others into the more inclusive 'respect/warmth/collegiality', because the scorers agreed about the other components of that domain. Disagreements were often to do with either scorer needing the other's help to see that particularly vivid experiences should not be given undue influence in determining the overall score. This was an important benefit of engaging two observers to work in partnership, in preference to extending the observation time of a single observer.

However, as already noted (see *Chapter 6, Observation-based evaluations of organisational, teamwork and safety cultures*), the four-level scale, described in *Chapter 4, Quantifying semistructured observations*, resulted in several tied scores, indicating that the scheme provided insufficient differentiation between research sites. Increasing sensitivity by increasing the number of levels in the scoring system, or increasing the number of domains scored, was investigated. After piloting, both approaches were felt to be reducing reliability to a greater extent than they were improving sensitivity. We prioritised reliable scoring for this study. Further research is needed to find ways to improve the sensitivity of scoring observations.

The evolution of the scoring framework in *Appendix 6* led to the exclusion of a number of potential items (see *Appendix 5*). These related to material that relied more on report (during conversation with staff) than on observation. The main problem with reported data was the variability and limited number of opportunities for conversations to illuminate and contextualise observations. Relying on a small number of opportunistic conversations did not allow observers to make informed judgements about scoring and the presence and nature of any bias in conversational reports. A more extensive ethnographic approach (or another multimethod approach) to data collection would have been required to retain the excluded facets of safety culture, such as action outcomes from incident reports.

Service user and service provider observers did not take part in the development of the scoring system or the final scoring: the time commitment would have been too great. However, all observers from each site met to discuss their findings in detail, and to ensure that any differences in observations and/or judgements were clearly explained and recorded in field notes that contributed to the scoring process.

For the final scoring, two research fellows (SA and MR) separately scored observations made at every site then discussed any differences and agreed joint scores. Owing to staff changes during the project (see *Chapter 4, Impact of staff changes in research team*), these two research fellows did not collect data at sites 1 and 2. At the scoring phase, EJB, who collected data at these sites, helped to resolve uncertainties about the scoring of two (from nine) scores for site 1, based on field notes relating to sites 1 and 2, which were written by EJB and NM (see *Table 27*). She also reviewed and agreed with the remaining 14 scores for sites 1 and 2. The difficulty in confidently assigning two scores highlighted that it may be important for observers to undertake scoring or for scorers to have recourse to discussion with observers. The ability to make confident and verified judgements based on other observers' field notes for 88% of scores indicated that, with sufficiently detailed or well-structured field notes, experienced researchers could score others' observations. This would mitigate the effects of staff changes, or the use of local researchers on multiple sites.

There were several reasons for quantification in this study. Most basically, the commissioning brief required not only that a 'holistic' method of data collection was used, comprising largely observation, but also that the results be available for statistical analysis in order to test for correlations with survey scores: a positivist view of triangulation at variance with the multiple

forms and several purposes of triangulation in the qualitative and multiple methods research literatures. However, despite some reservations stemming from our qualitative and multiple methods research backgrounds, in choosing to accept the commission we gave tacit endorsement to quantification as a valid method in this context.

Qualms about quantifying observations that are initially recorded as continuous field notes may arise from viewing narratives and grading as epistemologically incompatible. However, the gap between the two is not necessarily as wide as is often perceived: an important proportion of quantitative data are in any case a quantification of attitudes and judgements, rather than a measurement of physical phenomena. For example, survey questions (including our own) frequently ask respondents to review their knowledge and observations and to reach a judgement that can be scored, for example by Likert scales, although how these judgements are reached is invisible to the researchers. In the case of our 'observation' method, we have attempted to be more transparent about the method whereby these judgements are reached, and this has required a detailed description of how we made observations and subsequently scored them. The first part of this description may resemble that provided for a purely qualitative study, but this does not alter the fact that the rationale for the observation was ultimately quantitative.

We chose to record field notes at the point of data collection rather than rely solely on a checklist that would yield easily quantifiable data, in order to minimise the risk that we failed to record data that later turned out to be important. Although the initial detailed checklist was based on indicators of organisational and safety culture drawn from the wider literature (see *Chapter 3, Developing the observation prompt list*) and insights gained in our previous work in DUs,<sup>34</sup> and the method had been piloted in a local ED, we were, nonetheless, aware that we might encounter situations in which items might not be relevant and comprehensive in relation to a particular setting. An early example was the presence and use of a whiteboard at site 3, which was much more restricted than we had previously encountered. At site 3 computer screens, instead of a whiteboard, were the main mechanism for recording up-to-date patient data. Having made detailed notes of what we saw, informed by the prompt list (see *Appendix 5*), we were able to refer to field notes from this and other sites about other modes of information exchange and handover (informal, one-to-one, consultation of computers) to assist us in making comparative judgements. Had we scored a checklist in the field, we would have much more restricted information from which to formulate ways to cope with the unexpected. Similarly, by recording field notes, we made it easier for observers to discuss their, at times varying, interpretations of the same phenomena. For example, the research fellows initially differed in their assessment of the senior consultant at one site, one finding his leadership style authoritarian and hierarchical, the other noting his consistent willingness and availability to give support to junior staff. Each observer could check his or her own notes against the other's interpretation, and a consensus based on recorded data rather than on recall was quickly reached, that both impressions were in fact valid. In addition, by recording field notes, we made it possible to reconsider data in the light of any new perspectives from service user or professional observers, whom we were not able to recruit before site 6. Detailed field notes also mitigated the impact of staff changes.

For all these reasons, the recording of field notes was preferred; but at no point did we forget that the purpose of observation was to generate quantitative data relating to safety culture. Indeed, team members experienced some frustration in the early stages of the study in not being able to finalise the scoring method until there were sufficient data to test it.

### **Multiple delays to the data collection timetable**

Even for the experienced research team conducting this study it was unusually difficult to accommodate the number and variety of delays to the data collection timetable that were encountered. These were due to:

- protracted decisions by hospitals about whether or not to participate in the study
- very time-consuming processes for research governance at hospital level
- variations in the time taken by trust-based auditors.

Of these, the second was by far the most important.

### **Patient consent for notes audit**

The NIGB requirement that case notes can be audited only by the care team unless individual patients consent was a significant challenge to us. If this view is applied consistently from now on, the viability of the use of case note audit by non-hospital-based research teams is much reduced, given the likelihood of low response rates to letters requesting permission and the reluctance of hospitals to disclose names and addresses. In our experience, hospitals did not find it easy to provide their own staff for external research purposes, even though this work would be fully reimbursed. During the period of this study, arrangements for reimbursement of trusts' research costs improved (in value and simplicity). In other research we conducted over the same time period we noticed increased enthusiasm for research participation, which seemed to parallel the improved system for research support costs. However, our audit-based research seeking non-clinical auditors remained difficult to support. Perhaps there is a shortage of non-clinical auditors, or perhaps audit work is not attractive to non-clinical staff. It is to be hoped that recent guidance on improving access to clinical data for health services research<sup>196</sup> results in changes that better support audit-based research. However, the initial focus is on speeding recruitment to clinical trials, rather than health services research.<sup>222</sup>

### **Strengths of this study**

We chose to conduct this study in DUs and EDs, which are high-risk settings with unpredictably fluctuating workloads and high footfall. It was ambitious to work in relatively unbounded settings, characterised by 'distributed work',<sup>46</sup> but most patient safety studies to date have focused on operating theatres or other bounded settings, and bounded activities such as specific clinical procedures. Much of health care occurs in less bounded settings and involves interactions between activities; distributed work is common. By developing the methods used in this study in these challenging environments we can be more confident that they will be feasible in both bounded and unbounded contexts, and applicable to distributed work.

A diverse sample of research sites was secured that reflected the purposive sampling criteria outlined in *Chapter 3, Selecting the sample of research sites*. This increases confidence in the trustworthiness of the results. The professional knowledge and networks of the researchers and service provider advisors proved important for recruiting the purposive sample of research sites because publicly available data proved inadequate to identify clinical or managerial leads for initial communication to introduce the study.

Prior to the data collection phase, senior clinicians from two of the research sites (2 and 7) participated in the process of identifying markers of the quality of care. This brought context-specific clinical wisdom into this process.

The staff survey questionnaire used prevalidated questions selected from sections of two well-established research instruments: the then most recent NHS national staff survey<sup>175</sup> and the teamwork and safety climate survey.<sup>25</sup> The NHS national staff survey has good theoretical underpinning<sup>112</sup> and was tested for internal consistency and inter-rater reliability.<sup>234</sup> It was becoming increasingly familiar to NHS staff through annual distribution to a sample of staff in all NHS trusts and the interest generated by aggregated results returned to trusts and summaries that

were made public. Year on year there are minor modifications to the NHS national staff survey but it remained in use and essentially intact throughout this study. The teamwork and safety climate survey<sup>25</sup> also benefits from strong theoretical underpinning, extensive psychometric testing, the availability of benchmarking data and good face validity. Although developed in the USA, it has been used extensively in England.<sup>148</sup> One small change was required to reflect professional titles in DUs. We used only the 13 questions that constitute the teamwork climate and safety climate scales. These have been shown to have construct validity and reliability.<sup>96</sup>

Multilevel analysis of survey responses has expanded the range of demographic and role-related factors shown to influence organisational safety climate scores.

This study required quantifiable observation-based assessments of culture to enable comparison with other quantified assessments. No suitable scheme existed, so a framework based on earlier research was developed and refined through use. A scoring scheme was developed (see *Appendix 6*). The observation-based assessments of culture built upon expertise developed by members of the study team in a longitudinal study which examined interactions between staff on four contrasting delivery units over a 2-year period. This included both ethnographic and highly structured observations.<sup>34,132</sup> The marginally participant observation undertaken for this study was economical and unobtrusive (see further discussion of observation as a method in *Objectives*). Service provider coresearchers made observations from the same vantage points as non-clinical research fellows at eight research sites (50%) and discussed the research fellows' observations from the remaining sites. Ten research sites took up the offer of feedback from observations and other data collection strands. Clinical staff consistently validated the observations made by researchers, which indicated that the strategic immersion strategy developed for this study had been successful and that the researchers were capturing core facets of culture in a manner that was recognisable (if not always attractive) to stakeholders. However, the observation method still requires development to increase sensitivity and needs to be tested in other contexts.

This is the first study to examine the nature and feasibility of comparisons, first, between quantitative observation-based evaluations of culture and survey-based measurements of staff perceptions of climate and, secondly, between observation-based evaluations of culture and audit-based evaluations of the quality of care. This presents epistemological, methodological and practical challenges which are discussed throughout this chapter.

The study established reasonably good agreement between survey-based measurement of safety climate and observation-based evaluation of teamwork culture, indicating that, using a suitable framework and time-limited semistructured observations to evaluate teamwork culture may offer an alternative to staff surveys in circumstances where survey data collection is thwarted by very low response rates or increasing social desirability bias. On the other hand, some differences between survey results and observation-based evaluations were illuminating, suggesting a role for time-limited, semistructured observations as an adjunct to staff surveys in the pursuit of deeper understanding of safety culture.

## Limitations of this study

The power calculation identifying the need for comparison of 16 pairs to test each hypothesis with 90% power at the 1% significance level assumed independent samples. For economy, two research sites were recruited at each hospital, the DU and the ED. These are clinically distinct departments, in separate directorates, and at all of the study sites geographically separated within the hospital. We found no overlap between staffing, although peripatetic staff such as social

workers, health-care advocates and interpreters may well work in both departments. At research site 6 the DU midwives were employed by a primary care trust rather than by the hospital trust. Each hospital had different key contacts for the DU and ED research sites and we were never aware of any interaction between these individuals. Multilevel modelling of survey responses to prevalidated scales found no statistically significant residual variance at hospital level, whereas significant variance was found at the level of respondents nested within the research sites, indicating that staff perceptions of aspects of organisational and safety climates predominantly varied among the 16 clinical departments rather than among the eight hospitals. Nevertheless, we cannot rule out a degree of clustering which may have reduced the power of this study.

The staff surveys yielded variable and low to moderate response rates (see *Chapter 4, Strand A: staff survey*), in line with similar UK studies.<sup>100,178</sup> Occasionally, very low response rates reduced the usefulness of local data for the clinical department: at site 7, for example, we judged the response rate to be so low that we did not include the results in local feedback. For the comparisons at the heart of this study, low response rates would be a problem only if non-respondents differed systematically from respondents, which is unknown. The study sites could not provide demographic and role-related data with the staff lists provided for the survey (see *Chapter 4, Samples and response rates*), so non-response analyses could not be performed. Comparison of respondents with wider NHS workforce data established that survey respondents were broadly representative of the NHS workforce with respect to gender and ethnicity (see *Chapter 5, Demographics and role-related characteristics*). The multilevel modelling in *Chapter 5* revealed that site-specific response rates mediated responses to only one of the prevalidated scales used in this study; research sites with higher than average response rates returned lower 'line management' scores. Thus, this analysis largely allayed fears that the variable response rates would preclude further analysis of the survey data in comparisons with observations and audit data.

Difficulties with recruitment and research governance procedures resulted in much lower participation from service user coresearchers than had been planned. The observation-based assessments of organisational and safety cultures mainly reflect the perspectives of the study research fellows and service provider coresearchers.

Observation-based assessment, though sensitive to team factors, was of limited usefulness in assessing key aspects of organisational culture, for example the existence and effectiveness of appropriate systems for ensuring CPD, and for reporting and acting on errors and 'near-misses'. Though in some sites these might, by chance, be mentioned in passing, we could not assume that no mention indicated no systems. Similarly, routine checking and maintenance of some equipment is periodic, and may happen out of observers' sight. Furthermore, as we did not seek access to computer-based resources (to avoid obstructing staff who wished to use computers, of which there were usually limited numbers), we were unable to assess the availability of electronic information to support clinical decision-making. Conversations with senior staff could throw some light on such issues, but, consciously or not, staff may paint an unduly positive or negative picture. For example, staff might report staff levels or ward size to be inadequate because they were currently lobbying internally for increases; or report them to be adequate because they preferred not to face the limitations of their service they work in (termed 'cognitive dissonance').<sup>27</sup> However, relying on a small number of opportunistic conversations did not allow us to make informed judgements about which sort of bias, if any, was at work. The extended periods of observation and supplementary data collection characteristic of traditional ethnography would have been more suitable for understanding organisational culture or, equally, a multimethod design which augmented observation in the style of this study with more extensive complementary data collection, such as interviews.

Criterion-based case notes audit is a time-consuming process, and we were unable to examine larger numbers of records than the 50 per condition planned at the outset. We were thus unable to compensate for the relatively high incidences of missing data by looking at extra cases.

Comparisons like those at the heart of this study will not be meaningful and useful if the measures they are based upon lack of veracity. While the three strands of this study were conducted diligently and reflexively, we have described some imperfections and concerns, for example not being able to conduct non-response analyses for demographic and role-related variables in the staff survey (see *Chapter 4, Samples and response rates*) when other studies have found age, gender and profession affect responses to the SAQ used in this study.<sup>228,235</sup> The combination of SAQ and questions from the annual NHS staff survey (see *Chapter 3, Developing the organisational and safety climate questionnaire*) is unique to this study and will need testing in other contexts. The observation prompt list (see *Appendix 5*) and scoring (see *Chapter 4, Quantifying semistructured observations*, and *Appendix 6*) were both developed and refined during the course of this study and will need testing in other contexts. Staffing issues with the audit data collection (see *Chapter 4, Strand C: criterion-based audit of clinical notes*) precluded double-checking of a random sample of audited case notes.

The comparisons between survey-based and observation-based assessments of culture and with audit-based assessments of the quality of care examined association (using the Pearson product-moment correlation coefficient) and agreement (using Bland-Altman plots). Although these comparisons have begun to illuminate how different culture measures and the quality of care might be related, further research is required, for example a multivariable linear regression analysis to look at the relationships between the different measures of culture, after taking into account other variables, such as type and size of unit or demographic and role-related variables. However, multivariable linear regression analysis would, to a certain extent, be limited by the degree that non-linearity affected the Pearson product-moment correlation scores ( $r$ ) in this study, producing mainly low correlations. This might imply that other modelling techniques will be required.

## Chapter 10

### Conclusions

The government report *An organisation with a memory*<sup>2</sup> recognised that organisational factors play a key role in patient safety and that the past decade witnessed a rapid expansion of research and practice development activities focused on changing cultures in health care to better support safety. Of course, an emphasis on changing cultures necessitates the development of ways to assess culture, and this has been approached both qualitatively and quantitatively. Ethnographic studies have provided thick description and important insights relating to key interactions within health care. Self-assessment frameworks have been developed to prompt and guide practice development within health-care teams. In addition, a range of questionnaires containing mainly Likert-scale items have become popular means of measuring facets of culture. Perhaps the best known and most extensively researched of these is the SAQ.<sup>24,25</sup> The popularity of these measures might eventually become their downfall as health-care professionals tire of completing them and knowledge develops about 'correct' answers. One driver behind the commissioning of this study was to investigate whether equally good assessments of culture might be made through limited periods of observation.

The second purpose of this study concerned a need for further research into the relationship between measurements of culture and measurements of the quality of care. Consequently, this study, sited in 16 clinical departments, had three equally important strands (A–C):

- A A postal questionnaire survey of staff perceptions of organisational and safety climates, representing the most commonly adopted approach to assessing culture. The survey included prevalidated questions and scales from the annual NHS staff survey<sup>175</sup> and two domains, teamwork climate and safety climate, from the SAQ.<sup>25</sup> A weighted average of responses to questions concerning aspects of organisational climate was used to form a summary survey-based measure of organisational climate. A summary survey-based safety climate measure was formed in a similar manner.
- B Semistructured non-participant observation of work in non-treatment, workload management areas of the clinical departments, such as in the vicinity of a whiteboard, was used to monitor occupancy and the progress of care. This strand examined the viability of assessing culture from approximately 12 visits to a clinical department, each lasting 4–6 hours and purposively selected to span Monday–Saturday and 6 AM to midnight. This approach to observation was termed strategic immersion. Non-clinical research fellows with prior experience of health services research conducted 12 observation visits per research site. They were advised by service provider coresearchers with clinical expertise relevant to the research sites and a service user coresearcher. The service provider coresearchers made observations from the same vantage points as the research fellows at eight (50%) research sites. The service user coresearcher similarly made observations at three research sites. Observation field notes were grouped into eight categories, three of which concerned organisation and work environment factors, whereas five categories concerned aspects teamwork (see *Appendix 6*). Observations within each category were then scored using a four-point scale developed during the study:
  - 0 consistent lack of features of a safety climate
  - 1 frequent lack of features of a safety climate

- 2 frequent presence of features of a safety climate
- 3 consistent presence of features of a safety climate.

The mean score for the three organisation and work environment factors became the observation-based organisational culture measure; the mean of the five team factors scores became the observation-based teamwork culture measure and the mean of all eight categories became the observation-based holistic safety culture measure.

- C A retrospective criterion-based audit of markers of the quality of care for three commonly encountered conditions in each type of clinical department. The audits were conducted by non-clinical staff to minimise hindsight bias. The markers selected for each condition were drawn from national guidelines. Some markers concerned treatment interventions, whereas others concerned making and recording important observations.

To facilitate aggregation within strand A and comparison between different strands of the study, normally distributed scores were standardised (scaled to fit the standard normal distribution, mean zero, SD 1).

To help distinguish between strands A and B, following a convention in patient safety studies, the measurements made in strand A were termed organisational *climate* and safety *climate*, whereas the measurements developed in strand B were termed organisational *culture* and safety *culture*. Broadly this reflects the argument that organisational culture is the collection of shared beliefs, values and norms of behaviour found in an organisation<sup>18</sup> and safety *culture* is the subset of those values, beliefs and norms that have particular relevance to the promotion of patient safety.<sup>19</sup> Safety *climate* refers to the aggregate of individual perceptions of practices, policies, procedures and routines about safety in an organization.<sup>14,136</sup> Thus *climate* focuses on consciously held perceptions, which may be viewed as the surface layers of *culture*. It is often argued that questionnaire surveys capture *climate* rather than *culture*.<sup>25</sup>

## Survey-based assessments of organisational and safety climate

Surveys elicit variable response rates, which depend on a complex web of factors. Postal surveys should be expected to have lower response rates than surveys that are conducted face to face, with perhaps dedicated time given to survey completion. In this study, response rates varied widely between research sites. If respondents differ systematically from non-respondents, non-response bias will occur. Concerns about non-response bias can be partially mitigated by comparison of respondents' demographic data with the demographic profile of those invited to participate in the study and the demographic profile of the wider health-care workforce. A further approach, trialled in this study, was to examine response rate as a possible explanatory variable in multilevel modelling of factors which may mediate scores on the prevalidated scales used to form the composite measures of organisational climate and safety climate used in this study. Response rate was not found to mediate scores for five out of six prevalidated scales. This finding echoes a recent study<sup>215</sup> which found very variable response rates to a hand-delivered follow-up administration of the SAQ in 63 surgical departments; regression analysis revealed that response rates had no significant effect on the results. Although it is prudent to be vigilant, the evidence from these two studies suggests that safety climate scores may not be greatly affected by variable response rates.

Responses to prevalidated scales drawn from the annual NHS staff survey<sup>175</sup> were normally distributed. Responses to two domains, teamwork climate and safety climate, drawn from the SAQ<sup>25</sup> were slightly skewed to the right, i.e. there was a high proportion of 'agree' and 'strongly



agree' responses to the five-point Likert-scale items forming these scales. The amount of skew was not large, and these scales were used in our summary survey-based measure of safety climate. Had the skew been larger, transformation would have been necessary before the scores from these scales could be analysed alongside the normally distributed results.<sup>205</sup> In *Chapter 9, Revisiting the study's original aims, objectives and hypotheses*, we discussed whether, on the one hand, teamwork and safety climates may be improving slightly after several years' investment in improvement initiatives or if, on the other hand, a small degree of 'social desirability bias'<sup>26</sup> could be creeping into responses to these scales. This suggests that future studies should be vigilant about the distribution of responses to the better-known climate scales, which have items closely reflecting key messages in educational and quality improvement initiatives. Indeed, the pressure to return 'the right' (socially desirable) answers will increase further if safety climate measures became drawn into current trends towards ranking performance and producing multifaceted league tables. Thus, the monitoring of the usefulness of established survey instruments needs to be mindful of their evolving usage.

Multilevel modelling of survey responses to the six prevalidated scales that contributed to this study's summary measures of organisational and safety climate found several demographic, role-related and professional development factors that mediated scores on these scales (see summary in *Table 43*). We found no previous studies that had investigated the mediating effects of a range of factors, although there is a good range of earlier work showing that the perceptions of managers and frontline staff may differ and some studies on gender differences and different professions' perspectives. More research is needed on the range of factors that mediate perceptions of safety culture and the consequences of any differences.

Residual variation in multilevel models was concentrated at clinical department level, not hospital level. This means that, after taking into account the known individual-level and department-level mediating factors, survey scores still varied between clinical departments within hospitals, and rather more so than between hospitals. Thus, staff perceptions of organisational and safety climate are mainly influenced by factors at clinical department level.

Clinical departments where staff rated safety climate favourably also returned favourable organisational climate scores. Further research would be needed to examine whether one favourable assessment is antecedent to the other. Zohar and Luria<sup>142</sup> have begun to examine this in the context of manufacturing plants.

## **Making observation-based assessments of organisational and safety culture**

In considering the feasibility of making observation-based assessments of culture the backgrounds of observers, the time dedicated to observation, the schedule of data collection and the scoring of observations need to be considered separately. First, the non-clinical researchers in this study were able to discuss their observations and emergent understandings with service provider coresearchers and, to a more limited extent (owing to difficulties obtaining research governance approvals), with a service user coresearcher. The researchers made extensive use of discussion with service provider coresearchers while they were learning how to observe and assess behaviours in clinical areas, subsequently observing and scoring with less recourse to clinicians' advice and interpretation. Debate between people from different backgrounds was useful throughout the study. The service provider coresearchers provided an important clinical perspective, though, when making observations themselves, one was less successful at making and/or recording observations of non-technical skills. The service user coresearcher provided a reality check so that the researchers did not become too closely aligned with the

clinical perspective and also proved an excellent observer of non-technical skills. The research governance screening of service user observers was experienced as onerous, leading to withdrawal from the study by one potential service user observer, and delays that prevented the other service user observer from making observations at most research sites. This suggests that non-clinical observers (particularly those who are inexperienced) ought to work in collaboration with, first, observers who have a clinical background that is relevant to the observed clinical environment and, second, where possible, a service user coresearcher. However, contributions from service user observers may be vulnerable to discouragement and delays from multiple research governance processes at different research sites. The planned streamlining and standardisation of research governance arrangements,<sup>196,222</sup> if enacted, should mitigate some current problems.

Turning to the time required for trustworthy assessments, this study used twelve purposively selected observation periods in each clinical department, lasting 4–6 hours (six visits each by two non-clinical research fellows). An additional observation period was conducted at some research sites by a service provider coresearcher or a service user coresearcher. After observations at the first four research sites, during which observers were honing their skills, the researchers felt that trustworthy evaluations could be made with fewer visits, perhaps as few as eight 4-hour visits, provided that purposive sampling over the week and 24-hour day continued.

For the observation-based assessments of culture, this study used sequential data collection periods at eight participating hospitals. Sequential observation of clinical areas allowed researchers to form, test and consolidate observation-based assessments of cultures within each research site with the minimum risk of contamination from observations at other sites. Sequential data collection also reduced the number of observers needed for the study. The design of observation-based studies ought to consider the possible advantages of sequential rather than parallel data collection in several research sites.

Quantifying semistructured observations was possible in eight domains (see *Appendix 6*) on a four-point scale (see introduction to this chapter). It was not possible to make more finely graded assessments that were consistent and in which the researchers had confidence. In addition, only three domains relating to organisational culture were observable in all research sites (staffing; premises and equipment; administrative, managerial and other support). This produced a narrow organisational culture measure that lacked discrimination, and we do not recommend this type of observation in clinical departments as a means to assess organisational culture. On the other hand, assessment of teamwork culture proved feasible and successful: feedback with research sites validated the researchers' observation-based evaluations. However, the instrument requires testing in other contexts. On a site-specific basis, the evaluation of teamwork culture is best presented as a profile of scores in the five domains (informal training and supervision; leadership and responsibility; respect, warmth and collegiality; information exchange within the team; mutual support). Comparisons between sites, and between different evaluations of culture and quality, promote the use of average scores, as in this study. For these purposes, further development of the domains and/or scoring would be useful to achieve greater discrimination in the scoring of teamwork culture.

The study research fellows changed after observations at the first two research sites and those in post at the end of the study were cautious about scoring from provisional ratings and supplementary field notes made by the initial two observers. This was resolved through discussion with one of the original observers. Wherever possible, scoring should be undertaken by the observers. If this is not possible, several strategies could help promote valid and reliable scoring, including: ongoing communication between scorers and observers; sufficiently

detailed and structured field notes; a formal hand-over discussion of field notes, in addition to a sufficiently structured scoring scheme.

This study found scant evidence of a Hawthorne effect or reactivity to observers. In any case, Monahan and Fisher<sup>223</sup> argue that much can be learnt when staged performances are noted. This suggests that observers should be vigilant for staged performances and other evidence of reactivity, but should not treat any such observations as 'contaminated' and therefore to be discarded.

## Making audit-based assessments of the quality of care

This study used a retrospective criterion-based audit of markers of the quality of care, conducted by non-clinical auditors. The retrospective nature addressed any concern about a possible Hawthorne effect. Criterion-based audit is important for inter-auditor reliability and facilitates the use of non-clinical auditors. The use of non-clinical auditors provides similar reliability to clinical auditors and reduces hindsight bias.<sup>149,186</sup> However, the use of non-clinical auditors is currently discouraged by the climate of research governance surrounding the use of clinical records for research. In addition, hospitals found it very difficult to identify non-clinical staff to undertake the audits, even though backfill or overtime costs could be reimbursed by the study. Delays of several months (up to a year) occurred before audit work could begin. Current conditions in NHS trusts and research governance processes strongly discourage audit-based research, particularly studies using non-clinical auditors. Recent guidance<sup>196</sup> regarding increased access to patient data in health services research, if enacted, should improve some aspects of audit-based research, but difficulties in providing clinical staff who are willing to extract anonymous data will remain. Unfortunately, facilitating the recruitment and deployment of non-clinical auditors to reduce hindsight bias was overlooked in the review and guidance, and needs to be revisited. Greater standardisation in the recording and coding of clinical records would reduce the time required to extract audit data.

Trust-based non-clinical auditors needed training and support from a person with experience of audit-based research. If the pool of sufficiently experienced people shrinks as a result of the current difficulties faced in conducting audit-based research, it may become difficult to maintain an adequate supply of mentors.

This study used audit to evaluate the quality of care and compare this evaluation with observation and survey-based assessments of culture/climate in a clinical department. However, compliance with some audit standards required contributions from other departments or teams within the hospital. Clinical departments also responded in the context of earlier interventions made at home or in pre-hospital care. Using audit to evaluate the quality of care provided by a clinical department (as opposed to the quality of a care pathway) overlooks the close-coupling of teams, departments and services contributing to care (see *Chapter 4, Close-coupling of departments and services*). This is unavoidable but ought to be noted where it occurs.

Audit results varied widely between different markers of the quality of care for the same clinical condition. First, this suggests that for an evaluation of the overall quality of care, audit results from several markers ought to be averaged to smooth variation between markers. Second, it is important not to lose sight of the variation between compliance with different markers of the quality of care because each marker is important and poor adherence could be clinically significant. As the choice of markers could influence research results, careful scrutiny of choices should occur. Repeated use of markers across studies supports comparisons and aggregation

of findings, provided the markers are valid indicators of the quality of care in each context. However, overuse of particular markers risks conversion into targets with a consequent loss of potency as indicators of the quality of care.

Audit markers that entailed repeated clinical observations (such as rechecking oxygen saturation or reviewing pain relief) exhibited poor compliance, suggesting that this area of clinical practice (or recording clinical practice) may need further support.

Audit markers were selected from national care standards. Those that required multiple observations sometimes yielded zero compliance, instead a subset of the required observations were made. Dixon-Woods and colleagues<sup>125</sup> discuss how rules in practice are so numerous that they quickly exceed individuals' ability to act on them. Staff then have to choose which rules to ignore, and may also ignore others without realising that they are doing so. This suggests that multifaceted national care standards may benefit from scrutiny to establish whether they could be streamlined.

Care records for repeated or multiple measurements were most complete in clinical departments where a suitable record sheet was provided. The design of site-specific record sheets promotes or inhibits the inclusion of certain clinical records and may also influence clinical actions.

### **Comparing different approaches to measuring climate, culture and the quality of care**

When different approaches to measuring climate, culture and the quality of care were examined using correlations, low correlations were found (except when comparing survey measurements of organisational and safety climate). This means that survey-based and observation-based assessments of culture did not have easily interpreted linear relationships either with each other or with summary audit results. However, Bland and Altman<sup>187</sup> argued that correlation should not be used to compare agreement between two measures, and this study found it more illuminating to compare measurements using Bland–Altman plots. These showed that this study's summary survey-based assessments of safety climate provided the closest levels of agreement with summary audit scores, suggesting that a questionnaire-based assessment, formed from prevalidated questions, can provide a useful indicator of the quality of care. However, there is a long way to go before researchers understand the link between safety climate and clinical performance, so direct measurements of performance must also continue. In studies in which an observation-based assessment of culture is preferred to a survey-based assessment of safety climate, the observation-based teamwork culture scores also provided good agreement with summary audit scores, with the same caveat of not abandoning direct evaluations of performance. Observation-based evaluation of culture is most likely to be preferred where survey response rates are low, there are concerns about social desirability bias or cognitive dissonance, or complementary data are desired to illuminate other measurements.

The process of making observation-based evaluations of culture generates valuable field notes. In this study, in the minority of cases where there was discrepancy between survey-based assessments of climate and audit-based evaluations of performance, the field notes from the observation strand identified likely explanations. Thus, staff surveys and semistructured observations have complementary roles and, together, may support a better understanding of variations in performance.

## Summary

This study developed survey-based and observation-based measures of organisational and safety climates/cultures and tested these for agreement with criterion-based audits of the quality of care, using evidence-based markers drawn from national care standards relating to three common clinical conditions encountered in the participating clinical departments. The principal conclusions from this study fall into three groups.

### *Questionnaire-based assessments of climate*

- Surveys elicit variable response rates and, although it is prudent to be vigilant, this may not negate the usefulness of the measurements of organisational and safety climates.
- Many factors mediate scores on prevalidated scales relating to organisational and safety climates, but studies to date have tended to focus only on the mediating effects of managerial responsibility, gender and professional background.
- Perceptions of organisational and safety climate are highly correlated and in close agreement (only one research site showed a large difference). Further research would be needed to examine whether perceptions of one aspect of climate are antecedent to perceptions of the other.
- Perceptions of climate vary more at the level of clinical departments than at hospital level.
- Survey-based assessments of safety climate exhibited close agreement with summary audit scores for most research sites, but large differences were found at two research sites. Field notes illuminated this difference, suggesting an important role for qualitative data alongside quantitative assessments.
- The four prevalidated scales drawn from the NHS annual staff survey, and used in this study to assess facets of organisational and safety climates, produced normally distributed results. On the other hand, the teamwork climate and safety climate scales from the extensively used SAQ produced slightly skewed results, owing to high levels of agreement. The items in these two scales closely reflect the main messages from numerous patient safety improvement initiatives. High levels of agreement may reflect the development of positive teamwork and safety climates. It is also possible that some degree of social desirability bias may be included in these scores.

### *Observation-based assessments of culture*

Based on 12 purposively sampled half-day visits by non-clinical observers:

- Observation-based assessments of organisational culture proved too limited and offered insufficient differentiation in scores. Observers at frontline level in clinical departments cannot make reliable evaluations of wider organisational culture, and questionnaire assessments of the perceptions of staff are a viable alternative, since the prevalidated scales used to measure facets of organisational climate appear to function well. Organisational ethnographies offer a second alternative.
- Safety-related facets of teamwork can be reliably observed and scored to obtain an observation-based measure of teamwork culture, which exhibits moderately good agreement with summary audit scores. The framework and scoring system developed during this study requires testing in other contexts and attention to greater discrimination in scoring. The current instrument is useful to augment survey assessments of safety climate. Following refinement, the instrument might be a suitable alternative to staff surveys of safety climate; particularly if survey response rates continue to fall, or concerns about social desirability bias in survey responses grow.

- Non-clinical observers benefit from collaboration with service provider observers to develop context-specific insights, and with service user observers to prevent insights becoming too closely aligned with clinical perspectives. In this study, research governance procedures discouraged the participation of service user observers.
- Reactivity to non-participant observers in clinical departments is minimal, possibly because work in these areas is constantly observed by a range of stakeholders.

*Criterion-based audit of the quality of care*

- Current conditions in NHS trusts and research governance processes strongly discourage audit-based research, particularly studies using non-clinical auditors. Recent guidance<sup>196</sup> regarding increased access to patient data in health services research, if enacted, should mitigate many current problems. However, the use of non-clinical auditors to minimise hindsight bias remains poorly supported.
- Averaging audit scores for a range of markers of the quality of care is necessary to smooth wide variations in performance on individual markers, but local quality improvement efforts must not lose sight of clinically important performance on individual markers.
- Markers requiring repeat observations showed poor compliance and this area might be a useful target for practice development. The design of clinical record forms supports or inhibits the recording of observations and may influence clinical actions.
- Markers requiring many repeated observations yielded very few cases of full compliance but many cases in which a subset of observations had been made, suggesting that it may be beneficial to examine the evidence for care standards of this type and whether any could be streamlined.
- As the selection of audit markers will affect performance results, careful choices must be made. It is important to avoid auditing only markers that are simple to assess, and it must be acknowledged that regularly audited markers will become targets, so lose their potency as markers of the quality of care. Use of markers used by other studies will support comparisons, provided that the markers are suitable for the studied contexts, but does risk overuse.

To summarise, this study was prompted by concerns that increasing awareness of ‘correct’ (socially desirable) answers may undermine any link between clinical performance and survey-based assessments of organisational and safety climates. Decreasing survey response rates were also a concern. The study was commissioned to formulate quantified assessments of safety culture, and if possible, organisational culture, derived from time-limited observations; compare the survey-based and observation-based assessments, then compare both with indicators of clinical performance and the quality of care, derived from criterion-based audits. We devised a feasible method for scoring teamwork culture, which exhibited reasonably good agreement with safety climate as measured by staff survey and quality of care as measured by evidence-based audit items: both observation-based teamwork culture and survey-based safety climate are partial indicators of performance. The observation scoring instrument developed during this study needs testing in other contexts and could benefit from increasing the discrimination of scoring. Nevertheless, this type of observation-based assessment of teamwork culture is useful to augment survey-based assessments, and, following refinement, could offer an alternative to staff surveys. However, time-limited observation in clinical departments provides too restricted a gaze to be a useful measure of organisational culture and staff surveys remain, as the results of this study indicate, the most effective way to gauge organisational climate. Field notes illuminated the lack of agreement between different measures that was found at a minority of sites, signalling the importance of qualitative data for understanding the complexity of influences upon and relationships between assessments of culture and performance.

## Future research

Findings from this study suggested that further research in the following areas would be beneficial.

### *Survey-based measurement of organisational or safety climate*

- Further examination of the impact of variable response rates between research sites is required.
- Further examination of individual, role-related and context-dependent variables that mediate scores on climate instruments would be beneficial.
- Established survey instruments should be monitored for signs of increasing proportions of positive and very positive responses and appropriate responses devised.

### *Measuring teamwork or safety culture using strategic immersion and semistructured non-participant observations*

- The data collection strategy, observational tool, categorisation and scoring system developed during this study need testing and further development in other contexts.
- The use of ethnographic methods to understand and assess safety culture requires more exploration.
- Further exploration of the possibilities and tensions associated with setting semistructured, broadly banded observational assessments alongside other approaches to assessing culture and performance is required.

### *Measuring the quality of care using criterion-based audit of clinical notes*

- Further research is needed into the best markers of quality in different contexts.
- Variation in compliance between different markers relating to the same clinical condition suggests the need for further investigation of which care standards are observed while others are not, and the reasons for these discrepancies.
- The influence of the design of physical and electronic record forms on clinical records, clinical actions, and the ease and accuracy of auditing should be further explored.
- Further exploration is required of ways to increase the use and availability of non-clinical auditors.





## Acknowledgements

We would like to thank the following for their contribution to the research:

- staff at the participating research sites: we are sorry that the promise of confidentiality prevents us from naming the many individuals who went out of their way to help us
- our service provider and service user coresearchers: Elaine Cole, Adela Hamilton, Antonia Lynch, Eva Menendez, Debra Penny
- our senior clinical advisors, for their sustained interest and advice throughout the lifetime of this project: Susan Bewley, Consultant Obstetrician, Maternal Fetal Medicine, Guy's and St Thomas's Hospitals NHS Trust and Matthew Cooke, National Clinical Director for Urgent and Emergency Care, Department of Health; Professor of Emergency Medicine, Warwick Medical School.

### Contributions of authors

Della Freeth led the design, implementation and reporting of the study.

Jane Sandall helped to design the study and contributed to the drafting of the final report.

Teresa Allan helped design the study and contributed to the analysis and reporting of findings.

Fiona Warburton contributed to the analysis and reporting of findings.

Emma-Jane Berridge and Nicola Macintosh helped design the study, carried out early field work (including pilot work) and contributed to the drafting of the final report.

Stephen Abbott and Mary Rogers carried out the majority of the field work, contributed to preliminary analysis of the data and helped to draft the final report.



## References

1. Lilford R. The English Patient Safety Research Programme: a commissioner's tale. *J Health Serv Res Policy* 2010;**15**(Suppl. 1):1–3.
2. Department of Health. *An organisation with a memory: report of an expert group on learning from adverse events in the NHS*. London: Stationery Office; 2000.
3. Department of Health. *Building a safer NHS for patients: implementing an organisation with a memory*. London: Department of Health; 2001.
4. Department of Health. *Safety first: a report for patients, clinicians and healthcare managers*. London: Department of Health; 2006.
5. National Audit Office. *A safer place for patients: learning to improve patient safety*. London: Department of Health; 2005.
6. Patient Safety Research Portfolio (PSRP). Birmingham: University of Birmingham. URL: [www.haps2.bham.ac.uk/publichealth/psrp/](http://www.haps2.bham.ac.uk/publichealth/psrp/) (accessed June 2011).
7. 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.
8. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, *et al*. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991;**324**:377–84.
9. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, *et al*. Incidence of adverse drug events and potential adverse drug events. *JAMA* 1995;**274**:29–34.
10. Kohn LT, Corrigan J, Donaldson MS. *To err is human: building a safer health system*. Washington, DC: National Academies Press; 2000.
11. Burke CS, Salas E, Wilson-Donnelly K, Priest H. How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. *Qual Saf Health Care* 2004;**13**(Suppl. 1):i96–104.
12. Flin R, Maran N. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Health Care* 2004;**13**(Suppl. 1):i80–4.
13. Vickers R. Training in human factor skills: lessons from aviation. *Clin Risk* 2009;**15**:8–10.
14. Singer SJ, Rosen A, Zhao S, Ciavarelli AP, Gaba DM. Comparing safety climate in naval aviation and hospitals: Implications for improving patient safety. *Health Care Manag Rev* 2010;**35**:134–46.
15. World Health Organization. *Patient safety research*. Geneva: World Health Organization, 2011. URL: [www.who.int/patientsafety/research/en/](http://www.who.int/patientsafety/research/en/) (accessed June 2011).
16. Mannion R, Davies HTO, Marshall M. *Cultures for performance in health care*. Maidenhead: Open University Press; 2005.
17. Cooper M. Towards a model of safety culture. *Safety Sci* 2000;**36**:111–36.
18. Davies HT, Nutley SM, Mannion R. Organisational culture and quality of health care. *Qual Health Care* 2000;**9**:111–19.

19. Pronovost PJ, Weast B, Holzmueller CG, Rosenstein BJ, Kidwell RP, Haller KB, *et al.* Evaluation of the culture of safety: survey of clinicians and managers in an academic medical center. *Qual Saf Health Care* 2003;**12**:405–10.
20. Singer SJ, Gaba DM, Geppert JJ, Sinaiko AD, Howard SK, Park KC. The culture of safety: results of an organization-wide survey in 15 California hospitals. *Qual Saf Health Care* 2003;**12**:112–18.
21. Flin R, Burns C, Mearns K, Yule S, Robertson EM. Measuring safety climate in health care. *Qual Saf Health Care* 2006;**15**:109–15.
22. National Reporting and Learning Service. *Patient safety resources*. London: National Patient Safety Agency. URL: [www.nrls.npsa.nhs.uk/resources/](http://www.nrls.npsa.nhs.uk/resources/) (accessed June 2011).
23. Patient Safety First Campaign. London. URL: [www.patientsafetyfirst.nhs.uk](http://www.patientsafetyfirst.nhs.uk) (accessed June 2011).
24. Sexton J, Thomas E, Helmreich RL, Neilands TB, Rowan K, Vella K, *et al.* *Frontline assessments of healthcare culture: Safety Attitudes Questionnaire norms and psychometric properties*. Technical report. Austin, TX: The University of Texas Centre of Excellence for Patient Safety; 2004.
25. Sexton JB, Helmreich RL, Neilands TB, Rowan K, Vella K, Boyden J, *et al.* The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
26. King M, Bruner G. Social desirability bias: a neglected aspect of validity testing. *Psychol Market* 2000;**17**:79–103.
27. Ovretveit J. Understanding and improving patient safety: the psychological, social and cultural dimensions. *J Health Organ Manag* 2009;**23**:581–96.
28. Oppenheim AN. *Questionnaire design, interviewing and attitude measurement*. London: Continuum; 2000.
29. Barclay S, Todd C, Finlay I, Grande G, Wyatt P. Not another questionnaire! Maximizing the response rate, predicting non-response and assessing non-response bias in postal questionnaire studies of GPs. *Fam Pract* 2002;**19**:105–11.
30. Cook JV, Dickinson HO, Eccles MP. Response rates in postal surveys of healthcare professionals between 1996 and 2005: an observational study. *BMC Health Serv Res* 2009;**9**:160.
31. Dixon-Woods M. Why is patient safety so hard? A selective review of ethnographic studies. *J Health Serv Res Policy* 2010;**15**(Suppl. 1):11–16.
32. Healthcare Commission. *Manual of clinical governance review/inspection practices*. 4th (web) edn. London: Healthcare Commission, 2004. URL: [www.healthcarecommission.org.uk/\\_db/\\_documents/04008515.pdf](http://www.healthcarecommission.org.uk/_db/_documents/04008515.pdf) (accessed June 2006).
33. Catchpole KR, de Leval MR, McEwan A, Pigott N, Elliott MJ, McQuillan A, *et al.* Patient handover from surgery to intensive care: using Formula 1 pit-stop and aviation models to improve safety and quality. *Paediatr Anaesth* 2007;**17**:470–8.
34. Mackintosh N, Berridge EJ, Freeth D. Supporting structures for team situation awareness and decision making: insights from four delivery suites. *J Eval Clin Pract* 2009;**15**:46–54.
35. Scott T, Mannion R, Marshall M, Davies H. Does organisational culture influence health care performance? A review of the evidence. *J Health Serv Res Policy* 2003;**8**:105–17.

36. Sheaff R. *Organisational factors and performance: a review of the literature: report for NHS service delivery and organisation research & development programme: programme of research on organisational form and function*. London: NCCSDO; 2004.
37. Singer S, Lin S, Falwell A, Gaba D, Baker L. Relationship of safety climate and safety performance in hospitals. *Health Serv Res* 2009;**44**:399–421.
38. Silva S, Lima ML, Baptista C. OSCI: an organisational and safety climate inventory. *Safety Sci* 2004;**42**:205–20.
39. Department of Health. *Reforming emergency care*. London: Department of Health, 2001. URL: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH\\_4008702](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_4008702) (accessed October 2010).
40. Department of Health. *Transforming emergency care in England*. London: Department of Health, 2004. URL: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH\\_4091775](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_4091775) (accessed October 2010).
41. Department of Health. *National Service Framework for children, young people & maternity services: maternity services*. London: Department of Health, 2004. URL: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH\\_4089101](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_4089101) (accessed October 2010).
42. National Patient Safety Agency. *Business plan 2004–05*. London: National Patient Safety Agency; 2004.
43. National Patient Safety Agency. *Understanding the patient safety issues for some groups of vulnerable women known to be at higher risk of maternal death or morbidity*. London: National Patient Safety Agency; 2005.
44. Hall KK, Schenkel SM, Hirshon JM, Xiao Y, Noskin GA. Incidence and types of non-ideal care events in an emergency department. *Qual Saf Health Care* 2010;**19**(Suppl. 3):i20–5.
45. Woloshynowych M, Davis R, R B, Wears R, Vincent C. *Enhancing safety in A&E care*. London: Imperial College, Clinical Safety Research Unit; 2006.
46. Wears RL, Woloshynowych M, Brown R, Vincent CA. Reflective analysis of safety research in the hospital accident & emergency departments. *Appl Ergon* 2010;**41**:695–700.
47. Ballard GM. *Nuclear safety after Three Mile Island and Chernobyl*. New York, NY: Elsevier Applied Science; 1988.
48. Vaughan D. *The Challenger launch decision: risky technology, culture, and deviance at NASA*. Chicago, IL: University of Chicago Press; 1996.
49. Silbey SS. Taming Prometheus: talk about safety and culture. *Annu Rev Sociol* 2009;**35**:341–69.
50. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* 2001;**322**:517–19.
51. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals. I. Occurrence and impact. *N Z Med J* 2002;**115**:U271.
52. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, *et al*. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ* 2004;**170**:1678–86.
53. Donaldson LJ, Gray JA. Clinical governance: a quality duty for health organisations. *Qual Health Care* 1998;**7**(Suppl.):37–44.

54. Kennedy I. *The inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary*. London: Department of Health; 2001.
55. Nieva VF, Sorra J. Safety culture assessment: a tool for improving patient safety in healthcare organizations. *Qual Saf Health Care* 2003;**12**(Suppl. 2):ii17–23.
56. Vincent C, Taylor-Adams S, Stanhope N. Framework for analysing risk and safety in clinical medicine. *BMJ* 1998;**316**:1154–7.
57. Reason JT, Carthey J, de Leval MR. Diagnosing ‘vulnerable system syndrome’: an essential prerequisite to effective risk management. *Qual Health Care* 2001;**10**(Suppl. 2):ii21–5.
58. Hann M, Bower P, Campbell S, Marshall M, Reeves D. The association between culture, climate and quality of care in primary health care teams. *Fam Pract* 2007;**24**:323–9.
59. Singer SJ, Falwell A, Gaba DM, Meterko M, Rosen A, Hartmann CW, *et al*. Identifying organizational cultures that promote patient safety. *Health Care Manag Rev* 2009;**34**:300–11.
60. Langfield-Smith K. Organisational culture and control. In Berry AJ, Broadbent J, Otley DT, editors. *Management control: theories, issues and practices*. Basingstoke: Macmillan; 1995.
61. Mannion R, Konteh FH, Davies HT. Assessing organisational culture for quality and safety improvement: a national survey of tools and tool use. *Qual Saf Health Care* 2009;**18**:153–6.
62. Hammersley M, Atkinson P. *Ethnography: principles in practice*. 3rd edn. London: Routledge; 2007.
63. Scott T, Mannion R, Davies H, Marshall M. The quantitative measurement of organizational culture in health care: a review of the available instruments. *Health Serv Res* 2003;**38**:923–45.
64. Pidgeon NF. Safety culture and risk management in organizations. *J Cross-Cult Psychol* 1991;**22**:129–40.
65. Wakefield JG, McLaws ML, Whitby M, Patton L. Patient safety culture: factors that influence clinician involvement in patient safety behaviours. *Qual Saf Health Care* 2010;**19**:585–91.
66. Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;**320**:781–5.
67. Page A. *Keeping patients safe: transforming the work environment of nurses*. Washington, DC: National Academies Press; 2004.
68. National Health Service Confederation. *Creating the virtuous circle: patient safety, accountability and an open and fair culture*. London: National Health Service Confederation; 2003.
69. Firth-Cozens J, Mowbray D. Leadership and the quality of care. *Qual Health Care* 2001;**10**(Suppl. 2):ii3–7.
70. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care* 2004;**13**(Suppl. 1):i85–90.
71. Boreham N. A theory of collective competence: challenging the neo-liberal individualisation of performance at work. *Br J Educ Stud* 2004;**52**:5–17.
72. Baker DP, Day R, Salas E. Teamwork as an essential component of high-reliability organizations. *Health Serv Res* 2006;**41**:1576–98.
73. Weaver SJ, Rosen MA, DiazGranados D, Lazzara EH, Lyons R, Salas E, *et al*. Does teamwork improve performance in the operating room? A multilevel evaluation. *Jt Comm J Qual Patient Saf* 2010;**36**:133–42.

74. Wright MC, Taekman JM, Endsley MR. Objective measures of situation awareness in a simulated medical environment. *Qual Saf Health Care* 2004;**13**(Suppl. 1):i65–71.
75. Miller K, Riley W, Davis S. Identifying key nursing and team behaviours to achieve high reliability. *J Nurs Manag* 2009;**17**:247–55.
76. 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 Saf Health Care* 2010;**19**:592–6.
77. Boreham NC, Shea CE, Mackway-Jones K. Clinical risk and collective competence in the hospital emergency department in the UK. *Soc Sci Med* 2000;**51**:83–91.
78. Armstrong K, Laschinger H, Wong C. Workplace empowerment and magnet hospital characteristics as predictors of patient safety climate. *J Nurs Care Qual* 2009;**24**:55–62.
79. Coyle IR, Sleeman SD, Adams N. Safety climate. *J Saf Res* 1995;**26**:247–54.
80. Elfering A, Semmer NK, Grebner S. Work stress and patient safety: observer-rated work stressors as predictors of characteristics of safety-related events reported by young nurses. *Ergonomics* 2006;**49**:457–69.
81. Halbesleben JR, Wakefield BJ, Wakefield DS, Cooper LB. Nurse burnout and patient safety outcomes: nurse safety perception versus reporting behavior. *West J Nurs Res* 2008;**30**:560–77.
82. Parker D, Lawrie M, Carthey J, Coultous M. The Manchester Patient Safety Framework: sharing the learning. *Clinical Risk* 2008;**14**:140–2.
83. Parker D. Managing risk in healthcare: understanding your safety culture using the Manchester Patient Safety Framework (MaPSaF). *J Nurs Manag* 2009;**17**:218–22.
84. National Patient Safety Agency. *Manchester Patient Safety Framework (MaPSaF)*. London: National Patient Safety Agency; 2006.
85. Harrison R. Understanding your organization's character. *Harv Bus Rev* 1972;**50**:119–28.
86. Litwinenko A, Cooper CL. The impact of trust status on corporate culture. *J Manag Med* 1994;**8**:8–17.
87. Quinn RE, Rohrbaugh J. A spatial model of effectiveness criteria – towards a competing values approach to organizational analysis. *Manag Sci* 1983;**29**:363–77.
88. Cameron KS, Quinn RE. *Diagnosing and changing organizational culture: based on the competing values framework*. Reading, MA: Addison-Wesley; 1998.
89. Cameron K, Freeman S. Culture, congruence, strength and type: relationship to effectiveness. *Res Organiz Change Develop* 1991;**5**:23–58.
90. Gerowitz MB, Lemieux-Charles L, Heginbothan C, Johnson B. Top management culture and performance in Canadian, UK and US hospitals. *Health Serv Manag Res* 1996;**9**:69–78.
91. Cooke R, Lafferty J. *Organizational Culture Inventory (OCI)*. Plymouth, MI: Human Synergistics; 1987.
92. Thomas C, Ward M, Chorba C, Kumiega A. Measuring and interpreting organizational culture. *J Nurs Admin* 1990;**20**:17–24.
93. Ingersoll GL, Kirsch JC, Merk SE, Lightfoot J. Relationship of organizational culture and readiness for change to employee commitment to the organization. *J Nurs Admin* 2000;**30**:11–20.

94. Shortell SM, Zazzali JL, Burns LR, Alexander JA, Gillies RR, Budetti PP, *et al.* Implementing evidence-based medicine: the role of market pressures, compensation incentives, and culture in physician organizations. *Med Care* 2001;**39**(Suppl. 1):162–78.
95. Pronovost P, Sexton B. Assessing safety culture: guidelines and recommendations. *Qual Saf Health Care* 2005;**14**:231–3.
96. Kho ME, Carbone JM, Lucas J, Cook DJ. Safety Climate Survey: reliability of results from a multicenter ICU survey. *Qual Saf Health Care* 2005;**14**:273–8.
97. Shtyenberg G, Sexton J, Thomas E. *Test retest reliability of the safety climate scale. Technical Report 01-05.* Austin: University of Texas Centre of Excellence for Patient Safety Research and Practice, 2006. URL: [www.utpatientsafety.org](http://www.utpatientsafety.org).
98. Waterson P, Griffiths P, Stride C, Murphy J, Hignett S. Psychometric properties of the Hospital Survey on Patient Safety Culture: findings from the UK. *Qual Saf Health Care* 2010;**19**:e2.
99. Blegen MA, Gearhart S, O'Brien R, Sehgal NL, Alldredge BK. AHRQ's hospital survey on patient safety culture: psychometric analyses. *J Patient Saf* 2009;**5**:139–44.
100. Sarac C, Flin R, Mearns K, Jackson J. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. *BMJ Qual Saf* 2011;**20**:842–8.
101. Helmreich RL, Schaefer H. Team performance in the operating room. In Bogner M, editor. *Human error in medicine.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1994: pp. 225–53.
102. Hartmann CW, Rosen AK, Meterko M, Shokeen P, Zhao S, Singer S, *et al.* An overview of patient safety climate in the VA. *Health Serv Res* 2008;**43**:1263–84.
103. Hartmann CW, Meterko M, Rosen AK, Shibe Zhao, Shokeen P, Singer S, *et al.* Relationship of hospital organizational culture to patient safety climate in the Veterans Health Administration. *Med Care Res Rev* 2009;**66**:320–38.
104. Singer S, Meterko M, Baker L, Gaba D, Falwell A, Rosen A. Workforce perceptions of hospital safety culture: development and validation of the patient safety climate in healthcare organizations survey. *Health Serv Res* 2007;**42**:1999–2021.
105. Singer SJ, Hartmann CW, Hanchate A, Zhao S, Meterko M, Shokeen P, *et al.* Comparing safety climate between two populations of hospitals in the United States. *Health Serv Res* 2009;**44**:1563–83.
106. Singer SJ, Gaba DM, Falwell A, Lin S, Hayes J, Baker L. Patient safety climate in 92 US hospitals: differences by work area and discipline. *Med Care* 2009;**47**:23–31.
107. Gershon R, Karkashian, Grosch J, Murphy L, Escamilla-Cejudo A, Flanagan P, *et al.* Hospital safety climate and its relationship with safe work practices and workplace exposure incidents. *Am J Infect Control* 2000;**28**:211–21.
108. Turnberg W, Daniell W. Evaluation of a healthcare safety climate measurement tool. *J Safety Res* 2008;**39**:563–8.
109. Westrum R. A typology of organisational cultures. *Qual Saf Health Care* 2004;**13**(Suppl. 2):ii22–7.
110. National Patient Safety Agency. *Seven steps to patient safety.* London: National Patient Safety Agency; 2003.
111. National Patient Safety Agency. *Team Climate Assessment Measure (TCAM) Programme.* London: National Patient Safety Agency, 2006. URL: [www.nrls.npsa.nhs.uk/resources/?entryid45=59884](http://www.nrls.npsa.nhs.uk/resources/?entryid45=59884) (accessed June 2011).



112. Michie S, West MA. Managing people and performance: an evidence based framework applied to health service organizations. *Int J Manag Rev* 2004;**5**–6:91–111.
113. Ginsburg L, Gilin D, Tregunno D, Norton PG, Flemons W, Fleming M. Advancing measurement of patient safety culture. *Health Serv Res* 2009;**44**:205–24.
114. Milne JK, Bendaly N, Bendaly L, Worsley J, Fitzgerald J, Nisker J. A measurement tool to assess culture change regarding patient safety in hospital obstetrical units. *J Obstet Gynaecol Can* 2010;**32**:590–7.
115. Olsen E. Exploring the possibility of a common structural model measuring associations between safety climate factors and safety behaviour in health care and the petroleum sectors. *Accid Anal Prev* 2010;**42**:1507–16.
116. Currie L, Watterson L. Measuring the safety climate in NHS organisations. *Nurs Stand* 2010;**24**:35–8.
117. de Wet C, Spence W, Mash R, Johnson P, Bowie P. The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care* 2010;**19**:578–84.
118. Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, *et al.* Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care* 2004;**13**:330–4.
119. Cole E, Crichton N. The culture of a trauma team in relation to human factors. *J Clin Nurs* 2006;**15**:1257–66.
120. Catchpole KR, Giddings AE, de Leval MR, Peek GJ, Godden PJ, Utley M, *et al.* Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;**49**:567–88.
121. McDonald R, Waring J, Harrison S. At the cutting edge? Modernization and nostalgia in a hospital operating theatre department. *Sociology* 2006;**40**:1097–115.
122. Waring J. Adaptive regulation or governmentality: patient safety and the changing regulation of medicine. *Sociol Health Illness* 2007;**29**:163–79.
123. Waring JJ. Doctors' thinking about 'the system' as a threat to patient safety. *Health (London)* 2007;**11**:29–46.
124. Waring J, Harrison S, McDonald R. A culture of safety or coping? Ritualistic behaviours in the operating theatre. *J Health Serv Res Policy* 2007;**12**(Suppl. 1):1–3–9.
125. Dixon-Woods M, Suokas A, Pitchforth E, Tarrant C. An ethnographic study of classifying and accounting for risk at the sharp end of medical wards. *Soc Sci Med* 2009;**69**:362–9.
126. Robson C. *Real world research: a resource for users of social research methods in applied settings*. 3rd edn. Oxford: Wiley-Blackwell; 2011.
127. Smith V. Ethnographies of work and the work of ethnographers. In: Atkinson P, editor. *Handbook of Ethnography*. London: Sage, 2001; pp. 220–33.
128. Willis EM. The problem of time in ethnographic health care research. *Qual Health Res* 2010;**20**:556–64.
129. Fitzpatrick JM, While AE, Roberts JD. Shift work and its impact upon nurse performance: current knowledge and research issues. *J Adv Nurs* 1999;**29**:18–27.
130. Pound P, Sabin C, Ebrahim S. Observing the process of care: a stroke unit, elderly care unit and general medical ward compared. *Age Ageing* 1999;**28**:433–40.

131. Booth J, Davidson I, Winstanley J, Waters K. Observing washing and dressing of stroke patients: nursing intervention compared with occupational therapists. What is the difference? *J Adv Nurs* 2001;**33**:98–105.
132. Berridge EJ, Mackintosh NJ, Freeth DS. Supporting patient safety: examining communication within delivery suite teams through contrasting approaches to research observation. *Midwifery* 2010;**26**:512–19.
133. Spradley JP. *Participant observation*. New York, NY: Holt, Rinehart and Winston; 1980.
134. Hammersley M. *Reading ethnographic research: a critical guide*. 2nd edn. New York, NY: Longman; 1998. pp. 13–34.
135. Holy L. Theory, methodology and the research process. In Ellen R, editor. *Ethnographic research: a guide to general conduct*. London: Academic Press; 1984.
136. Griffin MA, Neal A. Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation. *J Occup Health Psychol* 2000;**5**:347–58.
137. Huang DT, Clermont G, Sexton JB, Karlo CA, Miller RG, Weissfeld LA, *et al*. Perceptions of safety culture vary across the intensive care units of a single institution. *Crit Care Med* 2007;**35**:165–76.
138. Halligan M, Zecevic A. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. *Qual Saf Health Care* 2011. doi:10.1136/bmjqs.2010.040964
139. Gaba D, Singer S, Rosen A. 2007. Safety culture: is the ‘unit’ the right ‘unit of analysis’? *Crit Care Med* 2007;**35**:314–16.
140. Colla JB, Bracken AC, Kinney LM, Weeks WB. Measuring patient safety climate: a review of surveys. *Qual Saf Health Care* 2005;**14**:364–6.
141. Mohr JJ, Batalden PB. Improving safety on the front lines: the role of clinical microsystems. *Qual Saf Health Care* 2002;**11**:45–50.
142. Zohar D, Luria G. A multilevel model of safety climate: cross-level relationships between organization and group-level climates. *J Appl Psychol* 2005;**90**:616–28.
143. Smits M, Wagner C, Spreeuwenberg P, van der Wal G, Groenewegen PP. Measuring patient safety culture: an assessment of the clustering of responses at unit level and hospital level. *Qual Saf Health Care* 2009;**18**:292–6.
144. Waring J, McDonald R, Harrison S. Safety and complexity: inter-departmental relationships as a threat to patient safety in the operating department. *J Health Organ Manag* 2006;**20**:227–42.
145. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;**335**:648–50.
146. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, *et al*. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. *Qual Saf Health Care* 2004;**13**:145–51; discussion 51–2.
147. Hutchinson A, McIntosh A, Anderson J, Gilbert C, Field R. Developing primary care review criteria from evidence-based guidelines: coronary heart disease as a model. *Br J Gen Pract* 2003;**53**:690–6.
148. Hutchinson A, Cooper KL, Dean JE, McIntosh A, Patterson M, Stride CB, *et al*. Use of a safety climate questionnaire in UK health care: factor structure, reliability and usability. *Qual Saf Health Care* 2006;**15**:347–53.

149. Hutchinson A, Coster JE, Cooper KL, McIntosh A, Walters SJ, Bath PA, *et al.* Assessing quality of care from hospital case notes: comparison of reliability of two methods. *Qual Saf Health Care* 2010;**19**:e2.
150. Woloshynowych M, Neale G, Vincent C. Case record review of adverse events: a new approach. *Qual Saf Health Care* 2003;**12**:411–15.
151. Hofer TP, Asch SM, Hayward RA, Rubenstein LV, Hogan MM, Adams J, *et al.* Profiling quality of care: is there a role for peer review? *BMC Health Serv Res* 2004;**4**:9.
152. Hulka B, Romm F, Parkerson, Jr. G, Russell I, Clapp N, Johnson F. Peer review in ambulatory care: use of explicit criteria and implicit judgements. *Med Care* 1979;**17**(Suppl.):1–73.
153. Fischhoff B. Hindsight not equal to foresight: the effect of outcome knowledge on judgement under uncertainty. *J Exp Psychol* 1975;**1**:e99.
154. Ashton CM, Kuykendall DH, Johnson ML, Wray NP. An empirical assessment of the validity of explicit and implicit process-of-care criteria for quality assessment. *Med Care* 1999;**37**:798–808.
155. Rudd AG, Lowe D, Irwin P, Rutledge Z, Pearson M, Party ISW. National stroke audit: a tool for change? *Qual Health Care* 2001;**10**:141–51.
156. Hutchinson A, Coster JE, Cooper KL, McIntosh A, Walters SJ, Bath PA, *et al.* Comparison of case note review methods for evaluating quality and safety in health care. *Health Technol Assess* 2010;**14**(10).
157. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;**363**:1147–54.
158. Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Brauholtz D, *et al.* The Leeds University Maternity Audit Project. *Int J Qual Health Care* 2002;**14**:175–81.
159. Neale G, Woloshynowych M, Vincent C. Exploring the causes of adverse events in NHS hospital practice. *J R Soc Med* 2001;**94**:322–30.
160. Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, *et al.* Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy* 2007;**12**:173–80.
161. Ehrenberg A, Ehnfors M. The accuracy of patient records in Swedish nursing homes: congruence of record content and nurses' and patients' descriptions. *Scand J Caring Sci* 2001;**15**:303–10.
162. Zegers M, de Bruijne MC, Spreeuwenberg P, Wagner C, Groenewegen PP, van der Wal G. Quality of patient record keeping: an indicator of the quality of care? *BMJ Qual Saf* 2011;**20**:314–18.
163. Flood M, Small R. Researching labour and birth events using health information records: methodological challenges. *Midwifery* 2009;**25**:701–10.
164. Hewett DG, Watson BM, Gallois C, Ward M, Leggett BA. Intergroup communication between hospital doctors: implications for quality of patient care. *Soc Sci Med* 2009;**69**:1732–40.
165. Werner RM, Bradlow ET, Asch DA. Hospital performance measures and quality of care. *LDI Issue Brief* 2008;**13**:1–4.
166. Hansen LO, Strater A, Smith L, Lee J, Press R, Ward N, *et al.* Hospital discharge documentation and risk of rehospitalisation. *BMJ Qual Saf* 2011;**20**:773–8.

167. Weingart SN, Pagovich O, Sands DZ, Li JM, Aronson MD, Davis RB, *et al.* What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med* 2005;**20**:830–6.
168. Weingart SN, Price J, Duncombe D, Connor M, Sommer K, Conley KA, *et al.* Patient-reported safety and quality of care in outpatient oncology. *Jt Comm J Qual Patient Saf* 2007;**33**:83–94.
169. Bowker G. Dismembering and remembering: classification and organizational memory. Locating design, development and use, Oksnoen Symposium, 13–18 May 1994.
170. Berg M. Practices of reading and writing: the constitutive role of the patient record in medical work. *Sociol Health Illness* 1996;**18**:499–524.
171. Fitzgerald L, Ferlie E, Wood M, Hawkins C. Interlocking interactions, the diffusion of innovations in health care. *Hum Relat* 2002;**55**:1429–49.
172. Ferlie E, Fitzgerald L, Wood M, Hawkins C. The (non) spread of innovations: the mediating role of professionals. *Acad Manag J* 2005;**48**:117–34.
173. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, *et al.* Changes in safety attitude and relationship to decreased postoperative morbidity and mortality following implementation of a checklist-based surgical safety intervention. *BMJ Qual Saf* 2011;**20**:102–7.
174. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
175. Healthcare Commission. *Staff survey reports*. London: Healthcare Commission; 2006.
176. Boud D, Garrick J. *Understanding learning at work*. London: Routledge; 1999.
177. Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, *et al.* Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;**324**:1183.
178. Catchpole K, Godden P, Giddings A, Hirst G, Dale T, Utley M, *et al.* *Identifying and reducing errors in the operating theatre: final report*. London: TSO; 2005. URL: [www.haps2.bham.ac.uk/publichealth/psrp/documents/PS012\\_Final\\_Report\\_DeLaval.pdf](http://www.haps2.bham.ac.uk/publichealth/psrp/documents/PS012_Final_Report_DeLaval.pdf)
179. Statistical Toolbox: Hong Kong: Department of Obstetrics and Gynaecology, Chinese University of Hong Kong, 2010.
180. Rasbash J, Steele F, Browne W, Golstein H. *A users' guide to MLwiN version 2.10*. Bristol: Centre for Multilevel Modelling, University of Bristol; 2009.
181. Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, *et al.* An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care* 2008;**17**:170–7.
182. College of Emergency Medicine. *Clinical standards for emergency departments*. London: College of Emergency Medicine; 2008.
183. Scottish Intercollegiate Guidelines Network. *Hip fracture. Quick reference guide*. Edinburgh: Scottish Intercollegiate Guidelines Network; 2002.
184. National Institute for Health and Clinical Excellence. *Intrapartum care. Care of healthy women and their babies during childbirth*. London: National Institute for Health and Clinical Excellence; 2007.
185. NHS Information Centre for Health and Social Care. *Caesarean section clinical guideline*. London: RCOG Press; 2004.

186. Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. *JAMA* 1991;**265**:1957–60.
187. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
188. Freeth D, Ayida G, Berridge EJ, Mackintosh N, Norris B, Sadler C, *et al.* Multidisciplinary obstetric simulated emergency scenarios (MOSES): promoting patient safety in obstetrics with teamwork-focused interprofessional simulations. *J Contin Educ Health Prof* 2009;**29**:98–104.
189. Hospital Episode Statistics. *Maternity data*. London: Department of Health, 2009. URL: [www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=19378categoryID=1011](http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=19378categoryID=1011) (accessed 23 November 2010).
190. National maternity statistics and trends. BirthChoice. URL: [www.birthchoiceuk.com/](http://www.birthchoiceuk.com/) (accessed June 2011).
191. Maskell NA, Jones EL, Davies RJ, committee BMMs. Variations in experience in obtaining local ethical approval for participation in a multi-centre study. *QJM* 2003;**96**:305–7.
192. Bedward J, Davison I, Field S, Thomas H. Audit, educational development and research: what counts for ethics and research governance? *Med Teach* 2005;**27**:99–101.
193. Thompson AG, France EF. One stop or full stop? The continuing challenges for researchers despite the new streamlined NHS research governance process. *BMC Health Serv Res* 2010;**10**:124.
194. Swan J, Robertson M, Evans S. *Managing clinical research in the UK*. Warwick: Warwick Business School; 2009.
195. McDonach E, Barbour R, Williams B. Reflections on applying for NHS ethical approval and governance in a climate of rapid change: prioritising process over principles. *Int J Soc Res Methodol* 2009;**12**:227–41.
196. Academy of Medical Sciences. *A new pathway for the regulation and governance of health research*. London: Academy of Medical Sciences; 2011.
197. Goffman E. *The presentation of self in everyday life*. London: Penguin; 1990.
198. Smart-Survey.<sup>TM</sup> URL: [www.smart-survey.co.uk/](http://www.smart-survey.co.uk/) (accessed June 2011).
199. Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care* 2004;**13**(Suppl. 1):i33–40.
200. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. *Med Educ* 2006;**40**:1098–104.
201. Thompson HC, Osborne CE. Office records in the evaluation of quality of care. *Med Care* 1976;**14**:294–314.
202. Starfield B, Steinwachs D, Morris I, Bause G, Siebert S, Westin C. Concordance between medical records and observations regarding information on coordination of care. *Med Care* 1979;**17**:758–66.
203. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ* 2000;**320**:50–2.
204. NHS ICHSC. *NHS hospital and community health services non-medical workforce census, England*. London: NHS Information Centre for Health and Social Care; 2009.
205. Bland JM, Altman DG. Transforming data. *BMJ* 1996;**312**:770.

206. Lane D. *Hyperstat online statistics textbook*, 1999. URL: <http://davidmlane.com/hyperstat/A14043.html> (accessed June 2011).
207. British Association for A&E Medicine Clinical Effectiveness Committee. *Standards for A&E departments*. London: British Association for A&E Medicine; 2006.
208. Sexton J. *A matter of life and death: social psychological and organizational factors related to patient outcomes in the intensive care unit*. Austin, TX: University of Texas; 2002.
209. Crouch S, Robinson P, Pitts M. A comparison of general practitioner response rates to electronic and postal surveys in the setting of the National STI Prevention Program. *Aust N Z J Public Health* 2011;**35**:187–9.
210. Dixon-Woods M. What can ethnography do for quality and safety in health care? *Qual Saf Health Care* 2003;**12**:326–7.
211. Gawande A. *The checklist manifesto: how to get things right*. New York, NY: Metropolitan Books, 2010.
212. Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. Reality check for checklists. *Lancet* 2009;**374**:444–5.
213. Brewer J, Hunter A. *Foundations of multimethod research: synthesizing styles*. 2nd edn. London: Sage Publications; 2005.
214. Creswell JW, Plano Clark VL. *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications; 2007.
215. Brannen J. Combining qualitative and quantitative approaches: an overview. In Brannen J, editor. *Mixing methods: qualitative and quantitative research*. Aldershot: Avebury; 1992. pp. 3–37.
216. Watts BV, Percarpio K, West P, Mills PD. Use of the Safety Attitudes Questionnaire as a measure in patient safety improvement. *J Patient Saf* 2010;**6**:206–9.
217. Sexton JB, Holzmueller CG, Pronovost PJ, Thomas EJ, McFerran S, Nunes J, *et al*. Variation in caregiver perceptions of teamwork climate in labor and delivery units. *J Perinatol* 2006;**26**:463–70.
218. Holden RJ, Scanlon MC, Patel NR, Kaushal R, Escoto KH, Brown RL, *et al*. A human factors framework and study of the effect of nursing workload on patient safety and employee quality of working life. *BMJ Qual Saf* 2011;**20**:15–24.
219. Carthy J, Clarke J. *Implementing human factors in healthcare*. London: Patient Safety First Campaign; 2010.
220. Gaba D, Howard S, Fish K, Smith B, Sowb Y. Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. *Simulation and Gaming* 2001;**32**:175–93.
221. WHO (World Health Organization). *Patient safety curriculum guide for medical schools*. Geneva: WHO; 2009.
222. HM Treasury and Department for Business, Innovation and Skills. *The plan for growth*. London: HM Treasury; 2011.
223. Monahan T, Fisher JA. Benefits of ‘Observer effects’: lessons from the field. *Qual Res* 2010;**10**:357–76.
224. Iedema R. New approaches to researching patient safety. *Soc Sci Med* 2009;**69**:1701–4.

225. Amalberti R, Vincent C, Auroy Y, de Saint Maurice G. Violations and migrations in health care: a framework for understanding and management. *Qual Saf Health Care* 2006;**15**(Suppl. 1):i66–71.
226. Osman M. *Controlling uncertainty: decision making and learning in complex worlds*. Oxford: Wiley-Blackwell; 2010.
227. Sheldon H, Rasul F. *Increasing response rates amongst BME and other hard to reach groups – a review of literature relevant to the national acute patients' survey*. Oxford: Picker Institute; 2006.
228. Carney BT, West P, Neily J, Mills PD, Bagian JP. Differences in nurse and surgeon perceptions of teamwork: implications for use of a briefing checklist in the OR. *AORN J* 2010;**91**:722–9.
229. Singer SJ, Falwell A, Gaba DM, Baker LC. Patient safety climate in US hospitals: variation by management level. *Med Care* 2008;**46**:1149–56.
230. Parand A, Burnett S, Benn J, Pinto A, Iskander S, Vincent C. The disparity of frontline clinical staff and managers' perceptions of a quality and patient safety initiative. *J Eval Clin Pract* 2011;**17**:1184–90.
231. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;**260**:1743–8.
232. Jha AK, Prasopa-Plaizier N, Larizgoitia I, Bates DW, Research Priority Setting Working Group of the WHO World Alliance for Patient Safety. Patient safety research: an overview of the global evidence. *Qual Saf Health Care* 2010;**19**:42–7.
233. Douglas M. *Natural symbols: explorations in cosmology*. London: Routledge; 1996.
234. CHI. *NHS national staff survey 2003: appendix 4 – psychometric properties of scales*. London, 2004: Publication by the former Commission for Health Improvement now available from the Healthcare Commission. URL: [www.cqc.org.uk/publications.cfm?fde\\_id+2003](http://www.cqc.org.uk/publications.cfm?fde_id+2003) (accessed October 2010).
235. Carney BT, Mills PD, Bagian JP, Weeks WB. Sex differences in operating room care giver perceptions of patient safety: a pilot study from the Veterans Health Administration Medical Team Training Program. *Qual Saf Health Care* 2010;**19**:128–31.





# Appendix 1

## Extracts from the tender specification

### RM05/JH33 – comparison of culture tools with holistic evaluation of an organisation's culture

#### Introduction

Management researchers have long been concerned with exploring the role that organisational culture may play in influencing the efficiency and effectiveness of organisational performance. More recently, this research has begun to be adapted and applied to a health-care context.<sup>2</sup> There is increasing interest in the idea that one way of improving health-care performance factors such as quality, efficiency and patient safety could be through influencing professional and organisational culture.<sup>2,3</sup>

#### Background

While business researchers seek to measure culture (and all NHS staff in England and Wales take part in a yearly staff attitude survey), many top business people, such as Harvey Jones and Archie Norman, believe they can perceive the culture of an organisation within a short period of time. In the more specific context of patient safety, the importance of creating a 'safety culture' is underscored in two key documents of the patient safety movement.<sup>1,4</sup> These reports suggest that such an outcome could be achieved by moving from a culture of blame to one of learning from mistakes. Some preliminary work has already been attempted to correlate the construct of 'organisational culture' with quality and performance. However, empirical evidence in support of a link between organisational culture and clinical performance (i.e. to test criterion validity) is weak.<sup>5,6</sup> Scott *et al.*<sup>6</sup> reviewed the evidence regarding organisational culture and health-care performance and identified several methodological difficulties. One problem is the fact that the studies that have compared culture measurements with criterion-based outcome or clinical standards are not only limited in number, but also in size. In order to test the validity of culture as a marker for quality/safety, it is necessary to have access to reliable measurements of both quality and culture.

As far as measuring quality of the process of clinical care is concerned, some steps have been taken to develop appropriate criterion-based indicators.<sup>7,8</sup> For example, Eric Thomas<sup>9</sup> has reviewed different methods to measure safety and Lilford and colleagues<sup>10</sup> have analysed their strengths and weaknesses and have articulated the importance of measuring error rates in terms of opportunity for error rather than patients treated. In addition, the NHS Research Methodology Programme is currently sponsoring a large study into the measurement of error/quality<sup>11</sup> on behalf of the Patient Safety Research Programme and the Health Care Commission. The challenge now is to find out how safety culture correlates with criterion based measurements of the quality of clinical care. For the purpose of this study, an error is any violation of a clinical standard, whether an act of omission or commission. As far as culture is concerned, this can be measured at a general level, for example using the Organisational Culture Instrument currently used by the Health Care Commission.<sup>12</sup> More specific constructs of culture can also be measured, for example safety culture.

Like organisational culture, safety culture is an abstract concept, as the following widely accepted definition of safety culture indicates:

The safety culture of an organisation is the product of individual and group values, attitudes, perceptions, competencies and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's health and safety management. Organisations with a positive safety culture are characterised by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventive measures.

This definition provided by the Advisory Committee on the Safety of Nuclear Installations<sup>13</sup> underscores the conceptual breadth of safety culture. However, the fact that safety culture is a broad and integrative concept also makes it difficult to operationalise. In addition, its global nature has meant that definitions of safety culture come from a variety of disciplines and perspectives ranging from social anthropology through sociology and psychology to management. Not surprisingly this has led to ongoing debate concerning several issues, and key amongst these is the definition of safety culture itself:

- Is culture integral to an organisation or is it merely an attribute of an organisation?<sup>14,15</sup>
- So, can culture be actively managed producing predictable outcomes? Or does it just happen?
- Are 'culture' and 'climate' distinct concepts?<sup>16</sup>
- What components are encompassed within the concept of 'culture'?<sup>17</sup>

The conceptual breadth of the construct of culture and the lack of consensus regarding its definition has meant that there is not one generic culture or safety culture tool.<sup>18</sup> This means that the process of deciding upon the most appropriate tool, selecting potential respondents and choosing the context in which to apply it are complex tasks and will depend on several factors. Some of these are outlined below:

- **Psychometric properties:** In addition to ensuring that a tool captures as many elements of the multifaceted construct of culture as possible, a successful tool must fulfil a number of psychometric criteria. A psychometrically sound tool must have construct and content (face) validity, it must be reliable in that the results are stable and it must be useable.<sup>19–21</sup>
- **Respondents:** Respondents should be selected to provide a representative and unbiased sample. However, this is not an easy task. Should culture be measured from the perspective of management or from that of the front-line workers or both? Furthermore, hierarchies exist even within these two groups and perceptions will not be homogeneous. For example, frontline workers might not necessarily share the same perceptions of culture, particularly if these respondents are different members of a multidisciplinary team. Finally, in a health-care setting, it may also be useful to obtain the patients' perspective of culture.
- **Generalisability:** If tools have been developed in a particular context or country they may require modification before they can be applied in new contexts.
- **Choice of setting:** Adverse events occur more frequently in some settings than others (e.g. accident and emergency vs primary care contexts). In such settings particular aspects of culture, such as safety may be more developed than in an area such as mental health where other dimensions may be more important.

Some aspects of culture, such as safety<sup>17,22–24</sup> may correlate poorly with generic culture – a point this call is designed to investigate. Even when tools have adequate psychometric properties some people believe that gathering culture information quantitatively by means of questionnaire, is too superficial. Furthermore, self-report tools might, at least in theory, be subject to social desirability biases.<sup>18</sup> Under self-report conditions individuals are often tempted to answer questions according to the way they think they should be responding rather than providing an accurate account of their true behaviour. A suggested way to address these issues would be to supplement the information gathered by these quantitative self-report tools with insights gleaned from more

holistic assessment.<sup>6,25</sup> However, to date very few standard tools have been compared with more holistic information. We therefore wish to 'triangulate' information obtained from two different methods of assessing culture: (1) using explicit quantitative culture tools (questionnaires) and (2) using holistic (implicit) methods. We leave it to the researchers to select (within the budget) suitable tools and holistic methods to compare. However, both the explicit and implicit methods should quantify general and safety culture separately, acknowledging that the holistic method might not be able to distinguish these dimensions.

Simply comparing people's responses and how they appear to behave can only tell us whether, or to what extent they agree or differ. We also need to know the way in which these measures relate to the quality and safety of care. We therefore wish to explore the extent to which standard safety culture tools and holistic approaches to the measurement of culture relate to criterion-based measures of error rates/the criterion-based quality of process of clinical care. We are aware that some clinical quality measures such as hand washing or infusion errors are based on observational methods. However, researchers developing/using holistic methods should use methods which stop short of actually observing compliance with clinical care standards. Separate and discrete measurements of compliance with clinical care standards will be needed so that the results of measuring culture can be correlated with the results of measuring quality/safety.

## Research required

We want to commission work to triangulate one or more generic and safety culture tools with holistic approaches to the assessment of generic and safety culture, and to determine the extent to which culture (as assessed by each method) and criterion-based measurements of the quality of clinical care (e.g. as reflected in error rates) are related. It is envisaged that the research will consist of the following phases:

### Pre-evaluation

1. **Selection of the general and safety culture tools:** The successful applicants must supply convincing evidence to support their choice of culture tool.
2. **Choice of holistic approach:** In order to address concerns that current culture tools may miss the most important aspect of culture, it will be important to compare them with the results of implicit assessment. This should be one in which observers who understand the underlying concepts of good health care immerse themselves more deeply in practice and make in-depth qualitative observations by means of direct observation and possibly internal conversations with staff. Knowledgeable and sensitive investigators or inspectors should 'get under the skin' of different institutions. More than one person should observe the same setting, so that interobserver agreement can be measured.
3. **Measuring the agreement between the culture tool(s), holistic methods and criterion-based measurements of clinical process:** Applicants must outline the way they will measure the level of agreement between the quantitative safety culture tools, criterion-based measurements of clinical process and the holistic approach.  
The results of the holistic approach will need to yield a continuous or ordinal assessment of culture as seen through the eyes of the observer.
4. **Choice of clinical context:** The research team will wish to select settings where clinical process measures can be made using established methods (e.g. acute medical wards). Preference will be given to research teams who can save resources by using sites in which they anyway propose to make measurement of one sort or another. For example, a research group might already be deploying staff to measure error rates across organisations and would simply have to graft the other two methods onto this infrastructure.

5. **Selection of intended respondents:** The research team must decide upon the composition of the sample of intended respondents in terms of management, frontline workers and service-users/patients. They should provide justification for their choice. For example, it might be appropriate to compare and contrast the perspectives of management and frontline workers. We are also interested in obtaining the perspectives of patients and this information can be obtained from the biannual patient survey in England and Wales.
6. **Methods to obtain criterion-based measurements of the quality of care:** The applicants should outline and defend the way in which they intend to obtain detailed, accurate and reliable measures of the quality of clinical care.

### Evaluation

**Triangulation of culture tools, holistic approaches and criterion-based measurements of the quality of care:** In the evaluation phase, information regarding culture obtained via the selected culture tool(s) will be triangulated with that gathered by holistic approaches to culture. At the same time, the extent to which culture as captured by both of these approaches relates to criterion-based measurements of the quality of clinical care will also be explored. This phase will involve using the culture tool(s) alongside holistic approaches to safety culture and criterion-based measurements of the quality of clinical care in each of the identified settings. In this way it will be possible to determine the extent to which the culture tool(s) reflect(s) the observed culture of each setting. Furthermore, the inclusion of criterion-based measurements of the quality of clinical care will provide a valuable opportunity to examine the criterion validity of the two approaches for measuring culture. It will be possible to see whether or to what extent safety culture can be separated from culture in general.

### Data protection/ethics

The research team will need to specify how they will deal with the necessary data protection and ethical issues.

### References

1. Department of Health. An organisation with a memory: Report of an expert group of learning from adverse events in the NHS chaired by the Chief Medical Officer. London: HMSO; 2000.
2. Scott T, Mannion R, Davies HTO, Marshall MN. Implementing culture change in health care: theory and practice. *International Journal for Quality in Health Care* 2003;**15**:111–18.
3. Nieva VF, Sorra J. Safety culture assessment: a tool for improving patient safety in healthcare organizations. *Quality & Safety in Health Care* 2003;**12**:II17–23.
4. Institute of Medicine. To err is human: Building a safer health system. Washington, DC: National Academy Press; 1999.
5. Wilderom CPM, Glunk U, Maslowski R. Organizational culture as a predictor of organisational performance. In Ashanasy NM, Wilderom CPM, Peterson MF, editors. *Handbook of organisational culture and climate*. Thousand Oaks: Sage Publications; 2000. pp. 193–209.
6. Scott T, Mannion R, Marshall M, Davies H. Does organisational culture influence health care performance? A review of the evidence. *Journal of Service Research and Policy* 2003;**8**:105–17.

7. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a Modified-Delphi approach. *Academic Emergency Medicine* 2002;**9**:1131–9.
8. Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. *Quality in Health Care* 2001;**10**:152–8.
9. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *Journal of General Internal Medicine* 2003;**18**:61–7.
10. Lilford RJ, Mohammed MA, Braunholtz D, Hofer TP. The measurement of active errors: methodological issues. *Quality & Safety in Health Care* 2003;**12**:II8–12.
11. Hutchinson, A. RM03/JH08/AH: Comparative Study of Different Methods to Assess Quality of Care/Safety. 2004.
12. Healthcare Commission, 2003 NHS staff survey report. Healthcare Commission. 2004.
13. ACSNI. *Organising for safety* (Advisory Committee for the Safety of Nuclear Installations: Human factors study group third report). Sheffield: HSE Books; 1993.
14. Elsmore P. *Organisational culture: Organisational change?* Hampshire: Gower Publishing Limited; 2001.
15. Huczynski A, Buchanan D. *Organisational behaviour: An introductory text*. Essex: Pearson Education Limited; 2001.
16. Guldenmund FW. The nature of safety culture: a review of theory and research. *Safety Science* 2000;**34**:215–57.
17. Singer SJ, Gaba DM, Geppert JJ, Sinaiko AD, Howard SK, Park KC. The culture of safety: results of an organization-wide survey in 15 California hospitals. *Quality & Safety in Health Care* 2003;**12**:112–18.
18. Cooper MD. Towards a model of safety culture. *Safety Science* 2000;**36**:111–36.
19. De Vellis R. *Scale development: Applications and theory*. Newbury Park, CA: Sage; 1991.
20. Kline P. *The handbook of psychological testing*. London: Routledge; 1999.
21. Oppenheim A. *Questionnaire design, interviewing and attitude measurement*. London: 1992.
22. Parker D, Kirk S, Claridge T, Esmail A, and Marshall M. The Manchester Patient Safety Assessment Tool. 2002.
23. Sexton JB, Thomas EJ, Helmreich RL. Error, stress and teamwork in medicine and aviation: Cross sectional surveys. *British Medical Journal* 2000;**320**:745–9.
24. Sorra, J. S. and Nieva, V. F. Hospital Survey on Patient Safety Culture. AHRQ Publication No. 04–0041. 1–9–2004. Rockville, MD: Agency for Healthcare Research and Quality.
25. Marshall M, Parker D, Esmail A, Kirk S, Claridge T. Culture of safety. *Quality & Safety in Health Care* 2003;**12**:318.
26. Savage J. Ethnography and health care. *British Medical Journal* 2000;**321**:1400–2.



# Appendix 2

## Study protocol

### Introduction

Patient safety is a core professional value that derives ultimately from the concept ‘*do no harm*’. However, in recent years concern has grown that the *culture* of healthcare can operate to the detriment of patient safety, for example pressure on resources, unwillingness to admit fallibility and difficulties in reporting concerns across professional boundaries and organisational hierarchies. Alongside technological advances to support safer care, increasing attention has fallen on promoting *cultural change* to promote patient safety and place more emphasis on mechanisms for *learning* from errors. In the UK the National Patient Safety Agency has taken forward the issues raised in *An Organisation with a Memory* and *Building a Safer NHS for Patients* stressing a systems-led approach to changing those aspects of NHS culture that contribute to organisational disincentives to learning from errors.

Interest in organisational and safety cultures is predicated on assumptions that culture is related to performance; that positive aspects of cultures can be identified; and that, over time, with appropriate interventions cultures can be changed without inadvertently creating dysfunctional consequences. None of these assumptions is unproblematic. Nevertheless the multifaceted concept of safety culture has become an increasingly popular focus for policy and managerial attention. Because culture is a fusion of values, attitudes, perceptions, competencies and behaviour it is very difficult to measure. The purpose of this study is to test the relationships between organisational and safety cultures and the quality of care, employing two contrasting methods to measure culture.

### Aims

- to compare<sup>a</sup> questionnaire and holistic<sup>b</sup> assessments of organisational<sup>c</sup> and safety culture
- to compare assessments of organisational and safety culture with criterion-based assessment of the quality of care.

### Notes

- a The tender specification (Appendix 1) used the term ‘triangulate’ but there are several types of triangulation defined in the research methods literature and different forms of triangulation might be considered to apply to different parts of this study. The central purpose of the study was to *compare* different assessments of culture with each other and with markers of the quality of care, using quantified assessments. For clarity, *compare* is used in this report in preference to triangulate.
- b The tender specification used the term holistic evaluation and this is reflected in the aims. The term holistic evaluation is ambiguous and, for clarity, in this report we will reflect the dominant data collection method by substituting the term observation-based assessment.
- c The tender specification used the term generic where we will use organisational.

## Objectives

1. To work *with* staff in the participating trusts such that their organisational and professional knowledge is respected, the study is understood and supported within participating departments, and prompt feedback to participant departments allows local development in advance of the study reporting to the wider health and research communities.
2. To use questionnaires to obtain quantitative assessments of the organisational and safety climate at each site.
3. To generate quantified holistic evaluations of organisational and safety culture for each site using observation.
4. To obtain criterion-based measurements for the quality of care at each site.
5. To compare levels of agreement between the questionnaire and holistic measurements of culture.
6. To compare organisational culture with safety culture.
7. To compare culture measurements and criterion-based measurements of the quality of care.
8. To collect data such that, where sufficient respondents exist within a category to protect anonymity, data can be explored by: stakeholder group (e.g. managers, midwives, nurses, doctors, allied health professionals, support staff) and level (e.g. management responsibility).

## Hypotheses

Comparing questionnaire assessments with holistic (observation-based) assessments:

- H1a There will be a strong correlation and good agreement between questionnaire-based and observation-based evaluations of *organisational* culture;
- H1b There will be a strong correlation and good agreement between questionnaire-based and observation-based evaluations of *safety* culture.

Testing the relationship between organisational and safety climate/culture:

- H2a There will be a strong correlation and good agreement between *questionnaire*-based evaluations of organisational and safety climates (see *Chapter 2, Safety culture and climate*, for discussion of the convention regarding the terms culture and climate).
- H2b There will be a strong correlation and good agreement between *holistic* evaluations of organisational and safety cultures.

Comparing culture assessments with the quality of care:

- H3a There will be a moderately strong correlation and reasonably good agreement between criterion-based measurements of the quality of care and, first, questionnaire-based and, secondly, holistic evaluations of *organisational* climate/culture;
- H3b There will be strong correlations and good agreement between criterion-based measurements of the quality of care and, first, questionnaire-based and, secondly, holistic observation-based evaluations of *safety* culture.

For this study strong correlation is defined as  $\geq 0.7$ .



## Study design

This study will assess organisational and safety climates/cultures in the labour ward and A&E department of eight hospitals (i.e. 16 research sites) so that these can be compared and, further, compared with evaluations of the quality of care in these clinical settings.

Organisational and safety climate/culture will be assessed in two ways: with a survey of staff (clinical, managerial and support staff) in each site using pre-validated questionnaires, and with a 'holistic assessment' using ethnographic methods and documentary evidence. Experienced clinicians and service-users will assist with the holistic measurement of organisational and safety culture.

Quality of care will be assessed from retrospective audits of patients' notes focused on three commonly occurring conditions for each clinical setting (labour ward and A&E). Three evidence-based markers of good care will be evaluated for each condition. Hospitals will be purposively selected and researched sequentially; each hospital will receive feedback at the end of its data collection period.

## Sampling

### Number of research sites

Silva *et al.* (2004) obtained a strong correlation,  $r = 0.72$  between organisational and safety climate, and showed organisational climate and safety climate to be inversely correlated with the incidence of accidents ( $-0.955 \leq \rho \leq -0.865$ ). Setting a similar threshold correlation level for this study is reasonable. It acknowledges that not only *can* these measures be highly correlated but indeed they *should* be if safety climate measures are meaningful. For our study we will seek a threshold correlation of 0.7. For 90% power at 5% significance this requires 16 pairs of measurements (Cohen, 1988). Two research sites (labour ward and A&E) will be studied in each participating hospital, requiring the recruitment of eight hospitals. These will be purposively selected.

### Purposive selection of hospitals

Eight hospitals will be identified. Purposive selection of hospitals is designed to ensure:

- Two research sites per hospital for economy in conducting the study
- Geographical spread (four out of the ten Strategic Health Authorities, with the North, East and South of England represented)
- Variability in size
- Varying staffing issues
- Type of Trust – the inclusion of single-site general hospitals, major tertiary centres and split-site provision.

### Research participants

Midwives, nurses, doctors, support workers, administrators and managers (and, where appropriate, allied health professionals) will be invited to participate. The staff to be invited will be determined through negotiation with key contacts to enable local definition of who belongs to the labour ward and A&E teams. Invitations to participate will be restricted to individuals who commenced work with the relevant team at least four weeks before data collection (at least eight weeks for staff working less than 0.5wte) so that perceptions of organisational and safety climates have been informed by adequate experience of local practice. All grades of non-medical staff

will be invited to participate in the research. Medical staff invitations will exclude Foundation Programme House Officers whose rotation makes it unlikely they will have sufficient consistent experience in the relevant clinical setting. Similarly, students will be excluded due to the limited duration of their clinical placements. At all the sites we anticipate surveying all members of the locally-defined team.

### **Planned time sampling for observation**

Three two-day visits will be undertaken at each hospital (see weeks 3–5 in data collection table below). This allows a reasonable balance between sustained observation to see how issues pan out over a number of hours and sampling sufficient time points to gauge the range of work and activity levels in the participating clinical areas. In recognition of fluctuating workloads and staffing levels the visits will be distributed throughout the week. Time sampling within visits will ensure that data are collected in each of the six hour periods 8am–2pm, 2pm–8pm, 8pm–2am, 2am–8am.

### **Planned quality of care measures for retrospective audit of notes**

For each clinical area we will select three common conditions for which there is a strong evidence base and national clinical guidelines on appropriate actions and care processes. We will focus on conditions associated with raised mortality and morbidity. The notes will look for three key markers of safe, evidence-informed care per condition. For labour wards we will focus on induction of labour, emergency caesarean section and premature labour (3 markers for each). In A&E, we will focus on fractured neck of femur, acute coronary syndrome, retention of urine and acute severe asthma.

Markers selected for each condition should be present in records and reliably recorded. They will reflect processes that will have a significant impact on subsequent diagnosis and/or care delivery. They will demonstrate levels of adherence to national guidelines; and have the potential to demonstrate error recovery, or interprofessional working and communication.

50 sets of notes will be sampled per condition, yielding 450 observations per research site.

## **Methods**

### **1. Organisational and safety climate questionnaires**

The Healthcare Commission NHS national staff survey (Health Care Commission, 2006) will be used to measure organisational climate. The instrument has good theoretical underpinning (Michie and West 2002) and is becoming increasingly familiar to NHS staff.

The Teamwork and Safety Climate Survey (Sexton *et al*, 2006), will be used to elicit safety climate. This 30-item questionnaire boasts strong theoretical underpinning, extensive psychometric testing, the availability of benchmarking data and good face validity.

The two instruments have been merged and the single questionnaire will be distributed via key contacts and local audit clerks. Reminders will be sent by post to protect the anonymity of non-responders.

### **2. Holistic measurement of safety culture**

Holistic measurement of safety culture at each site will be formed from ethnographic methods: direct non-participant observation of staff at work, brief conversations with key informants and some recourse to naturally occurring data. Data collection will be structured by a checklist which directs attention to key processes (e.g. triage and handover) and pertinent documentary evidence

(e.g. local protocols and vacancy rates). The checklist contains ordinal rating scales to facilitate comparison across sites and between measures.

Data collection will be undertaken by research fellows, service-provider coresearchers (doctors, nurses, midwives) and service-user coresearchers.

Retrospective audit of patients' notes to evaluate the quality of care

The research fellows will work with audit clerks from each hospital to identify cases and retrieve notes, and the research fellows will code the notes. The audit will be taken from 50 consecutive cases of the sampled conditions, commencing four weeks before 'week 1' in the data collection table below and working backwards in time. This will reduce the possibility of capturing a Hawthorne effect, with quality of care (or at least quality of recording care) potentially improving when researchers begin visiting the site. The start point will also increase the likelihood of the relevant records having been returned to medical records storage.

## Data collection

Data will be collected from each hospital sequentially; within hospitals, data will be collected in each research site (Labour ward and A&E) concurrently. Each hospital will require three months data collection and processing (outlined in the table below).

Researchers will liaise with key contacts within each hospital (a senior midwife, a lead nurse for A&E and a consultant doctor in each clinical area) to negotiate visits and other access issues, agree means of communication and ways of working, and discuss preliminary findings. The key contacts will help the RFs to comply with local research governance requirements, assist with identification of team members for the staff survey, and help to identify the holders of key information for the research. Service-user and clinical service-provider advisors will contribute to data collection for the holistic measurement of culture. An audit clerk at each site will conduct the retrospective audit of patients' notes. The RFs will co-ordinate data collection and processing, supporting key contacts, audit clerks and service-provider and service-user coresearchers.

### Planned data collection timetable (each hospital, sequentially)

Week	Activity
1	Two day visit (1 RF) to present forthcoming research to relevant staff, liaise with key contacts, begin extraction of notes, initiate compilation of staff lists for questionnaires, initiate search for key artefacts (e.g. local protocols).
2	Telephone and email support for audit clerk and key contacts
3	Mon-Tue visit (2 RF) semistructured observations in labour ward and A&E including some joint observation to establish inter-rater reliability. Liaison with key contacts. Preliminary data processing.
4	Wed-Thur visit (1 RF 2 days; service-user and clinical service-provider 1 day each). Data collection, preliminary processing, support and liaison as in week 3.
5	Fri-Sat visit (1 RF) data collection, preliminary processing, support and liaison as in week 3
6	Questionnaires distributed to staff.
7-8	Data analysis
9	Preliminary discussion of findings with key contacts, possible visit (1 RF, 1 day) to clarify any issues that cannot be resolved by email or telephone.
10	Questionnaire reminders
11	Data analysis
12	Visit (1 RF, 1 day) to provide feedback to staff
throughout	Processing and preliminary analysis of accumulating data set from this and previous hospitals. Negotiation of visits etc. with subsequent hospitals.

## Analysis

Data from each dimension of the safety climate scores will be matched with an appropriate measure from the audit of notes and suitable categories or variables created from the holistic measurement of culture. These data will be analysed using a random effects model to examine the levels of agreement between the three measures for each of the sites. This will be carried out for agreement between all the domains in the questionnaire and other important ones measured during the study. It will allow variations between the different aspects of climate and practice to be elicited between and within the sites.

For the larger professional groups (medicine, nursing and midwifery) it may be possible to detect profession-specific differences in organisational and safety climate perceptions from questionnaire responses. We will also be able to compare questionnaire responses of people who have joined teams recently with longer-serving members.

### Planned research schedule

Time	Research activity
Sep–Oct 2007	<p>The work notionally allocated to this period will commence as soon as funding is secured, since it will require intermittent effort over a period of five to six months.</p> <p>Research ethics and governance requirements</p> <p>Confirmation of access to research sites (currently agreed in principle) and negotiation of sequence of data collection among hospitals.</p> <p>Arrangements for piloting A&amp;E holistic measurement</p> <p>Arrangements for piloting audit of notes</p> <p>Revisit current selection of research instruments and quality of care measures in light of any new or updated literature and instruments</p>
Nov–Dec 07	<p>Advisory panel meeting</p> <p>Recruit and train service-user and clinical service-provider coresearchers</p> <p>Pilot A&amp;E holistic measurement</p> <p>Pilot audit clerk training and notes audit</p> <p>Finalise negotiation of access with Hospital 1</p>
Jan–Mar 08	Hospital 1 data collection (see details in data collection section)
Apr–Jun 08	Hospital 2 data collection and processing
Jul–Oct 08	<p>Hospital 3 data collection and processing</p> <p>Conference poster to raise awareness of project and elicit feedback on work in progress</p> <p>Interim report to funder on work completed to date and any proposed changes to the study</p> <p>Advisory panel meeting to elicit feedback</p>
Nov 09	Synthesis of data from hospitals 1–3. Pilot final data analysis.
Dec 08–Feb 09	Hospital 4 data collection and processing
Mar–May 09	Hospital 5 data collection and processing
Jun–Sep 09	<p>Hospital 6 data collection and processing</p> <p>Conference poster/paper to raise awareness of project and elicit feedback on work in progress</p> <p>Interim report to funder on work completed to date and any proposed changes to the study</p> <p>Advisory panel meeting to elicit feedback</p>
Oct 09–Dec 09	Hospital 7 data collection and processing
Jan 10–Mar 10	Hospital 8 data collection and processing
April 10–Aug 10	<p>Final data analysis</p> <p>Comparison of measures – see hypotheses above</p> <p>Additional comparison of data sets -</p> <p>Study sites' responses to the NHS staff survey (selected organisational climate measure) and publicly available NHS staff survey for relevant directorates, trusts and the wider sector</p> <p>Exploration of organisational and safety climate data by profession and level.</p> <p>Preparation of final report</p> <p>Presentation of conference paper</p> <p>Preparation of papers for peer reviewed journals and articles for professional journals</p>



## Appendix 3

### Staff survey

*The first page of the questionnaire identified the universities participating in the study and outlined the purpose of the research, instructions for completing the questionnaire and the handling of responses as follows:*

#### **What is this survey and why are we asking you to complete it?**

This survey is part of a research study being carried out by City University London and King's College, London (KCL). The research is looking at different ways of measuring organisational and safety culture in the NHS – of which this questionnaire is one. By measuring organisational and safety culture from different perspectives (questionnaires and observation), and looking at how these measurements relate to quality of care (measured by auditing patients' notes), we can investigate how robustly these different tools measure culture. Ultimately, this will help to provide better care for patients.

#### **How do I complete this survey?**

Please complete the survey for your current job, or the job you do most of the time. If you work across two or more NHS employers, please answer in relation to the trust at which you received the questionnaire. Please read each question carefully, but give your immediate response by ticking the box which best matches your personal view.

#### **Who will see my answers?**

The survey is being conducted by City University London and KCL, funded by the MRC-NIHR Methodology Research Programme. **Your answers will be treated in confidence. Only the research team will be able to identify individual responses. No-one in your trust will know who has taken part or have access to individual responses.** The survey findings will be analysed by City University London and KCL, and the results will be presented in a summary report in which no individual's answers can be identified.

The name, postal and e-mail addresses and telephone number of a City University Research Fellow were provided to both receive returned questionnaires and as the first point of contact for enquiries. Recipients were encouraged to return the questionnaire within 2 weeks by advertising a draw for one of ten £10 high street shop vouchers. The second and subsequent pages of the questionnaire follow. In questionnaires for ED teams question 10c contained the word *nurses*, which was substituted by *midwives* in questionnaires for DU teams.

## YOUR JOB

## 1. Working hours

## a. How many hours a week are you contracted to work?

1 Up to 29 hours2 30 or more hours a week

2. To what extent do you agree or disagree with the following?		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a	I cannot meet all the conflicting demands on my time at work.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b	I am asked to do work without adequate resources to complete it,	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c	I am required to do unimportant tasks which prevent me completing more important ones.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d	Relationships at work are strained.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

3. To what extent do you agree with the following statements about your immediate manager?		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
My immediate manager...						
a	...encourages those who work for her/him to work as a team.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b	...can be counted on to help me with a difficult task at work.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c	...gives me clear feedback on my work.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d	...asks for my opinion before making decisions that affect my work.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
e	...is supportive in a personal crisis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

## 4. The following questions are about team working and relate to the group of people that you work with most closely.

a	Do you have to work closely with other team members to achieve the team's objectives?	<input type="checkbox"/> 1 Yes	<input type="checkbox"/> 2 No
b	Does the team meet regularly to discuss its effectiveness and how it could be improved?	<input type="checkbox"/> 1 Yes	<input type="checkbox"/> 2 No

## TRAINING, LEARNING AND DEVELOPMENT

5. In the past 12 months, have you taken part in any of the following types of training, learning or development, paid for or provided by your Trust?		Yes	No
<i>Please tick Yes or No for each type of training</i>			
a	Taught courses (internal or external)	<input type="checkbox"/> 1	<input type="checkbox"/> 2
b	Any supervised on-the-job training	<input type="checkbox"/> 1	<input type="checkbox"/> 2
c	Having a mentor	<input type="checkbox"/> 1	<input type="checkbox"/> 2
d	Shadowing someone	<input type="checkbox"/> 1	<input type="checkbox"/> 2
e	E-learning/Online training	<input type="checkbox"/> 1	<input type="checkbox"/> 2
f	Keeping up to date with developments in your type of work (e.g. by reading books or journals, or by attending seminars or workshops)	<input type="checkbox"/> 1	<input type="checkbox"/> 2
g	Other methods of training, learning or development (please specify)		
_____			

## YOUR ORGANISATION

6. To what extent do you agree or disagree with the following?		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a	Managers here try to involve staff in important decisions	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b	Communication between management and staff is effective.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c	Managers encourage staff to suggest new ideas for improving services.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d	On the whole, the different parts of the Trust communicate effectively with each other.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
e	Care of patients / service users is my Trust's top priority.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
f	If I were a patient of this Trust, I would be happy with the standard of care provided.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5



7. Whistle-blowing		Yes	No	Don't know
a	If you were concerned about negligence or wrongdoing by staff in this Trust, would you know how to report your concerns?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	
b	Is there a system to report such concerns confidentially?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 9

#### ERRORS, NEAR MISSES AND INCIDENTS

8. Please tick yes or no for the following statement		Yes	No
	I know how to report errors, near misses or incidents	<input type="checkbox"/> 1	<input type="checkbox"/> 2

9. To what extent do you agree or disagree with the following statements?		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a	My Trust treats fairly staff who are involved in an error, near miss or incident.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b	My Trust encourages us to report errors, near misses or incidents.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c	My Trust treats reports of errors, near misses or incidents confidentially.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d	My Trust blames or punishes people who make errors, near misses or incidents.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
e	When errors, near misses or incidents are reported, my Trust takes action to ensure that they do not happen again.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
f	We are informed about errors, near misses and incidents that happen in the Trust.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
g	We are given feedback about changes made in response to reported errors, near misses and incidents.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

#### TEAMWORK AND SAFETY CLIMATE IN YOUR TEAM

10. Teamwork climate		Disagree Strongly	Disagree Slightly	Neutral	Agree Slightly	Agree Strongly	Not Applicable
a	Nurse input is well received in this clinical area	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b	In this clinical area, it is difficult to speak up if I perceive a problem with patient care	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c	The doctors and nurses here work together as a coordinated team	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d	Disagreements in this clinical area are resolved appropriately (i.e., not who is right, but what is best for the patient)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e	It is easy for staff here to ask questions when there is something they don't understand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f	I have the support I need from other staff to care for patients	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. Safety climate		Disagree Strongly	Disagree Slightly	Neutral	Agree Slightly	Agree Strongly	Not Applicable
a.	I would feel safe being treated here as a patient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b.	I am encouraged by my colleagues to report any patient safety concerns I may have	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c.	The culture in this clinical area make it easy to learn from the errors of others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d.	I receive appropriate feedback about my performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e.	Medical errors are handled appropriately here	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f.	I know the proper channels to direct questions regarding patient safety in this clinical area	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g.	In this clinical area, it is difficult to discuss errors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## BACKGROUND DETAILS

We would like to know some of your background details. Answers will be treated in confidence and no one outside the research team will be able to identify individual responses.

## 12. About you

## a Gender:

- 1 Male      2 Female

## b Age:

- 1 16-20    2 21-30    3 31-40    4 41-50    5 51-65    6 66+

## 13. What is your ethnic group?

**Prefer not to say** 17

**White**

- 01 British  
02 Irish  
03 Other White background

**Mixed**

- 04 White and Black Caribbean  
05 White and Black African  
06 White and Asian  
07 Any other mixed background

**Asian/Asian British**

- 08 Indian  
09 Pakistani  
10 Bangladeshi  
11 Other Asian background

**Black/Black British**

- 12 Caribbean  
13 African  
14 Any other Black background

**Chinese and other ethnic group**

- 15 Chinese  
16 Any other ethnic group  
 (please state) \_\_\_\_\_

## 14. Do you manage others within the Trust?

- 1 Yes      2 No

## 15. How many years have you worked for this Trust?

*If your Trust has merged with another or changed its name, please include in your answer all the time you have worked with this Trust and its predecessors*

- 1 Less than 1 year      3 3-5 years      5 11-15 years  
2 1-2 years      4 6-10 years      6 More than 15 years

**16. What is your occupational group?***Please tick one box only***Admin & Clerical**01 Admin & Clerical**Allied Health Professionals**07 (e.g. physiotherapy, occupational therapy, radiography, podiatry, clinical psychology, etc.)08 Scientific and Technical or Healthcare Scientists or support staff (e.g. haematology, clinical biochemistry, microbiology)09 Support to Allied Health Professionals (e.g. support worker, therapy helper, therapy assistant or student)**Medical and Dental**11 Medical - Consultant12 Medical - In Training (e.g. Foundation Y1, SHO / Foundation Y2, SpRs)13 Medical – Other (e.g. Staff and Associate Specialists / Non-consultant career grade)**NHS Infrastructure**15 Maintenance / Ancillary (e.g. housekeeping, domestic staff, maintenance, facilities, estates)**Registered Nurses and Midwives**16 Adult / General nurse17 Children's nurse18 Midwife**Nursing or Healthcare Assistants**19 Nursing auxiliary / Nursing assistant / Healthcare assistant / Midwifery assistant**Social Care**20 Social workers (qualified or unqualified)**General Management**23 General Management*(N.B. If you are a manager and can choose a group from elsewhere in the list, please select that other occupational group)***Other**24 Other occupational group (please specify)**17. If you have any additional comments about working in this organisation, please write them below and/or use a separate page**

*Thank you for completing the survey. Your time and participation are greatly appreciated.*  
Please return to Stephen Abbott, City University, 20 Bartholomew Close, London EC1A 7QN



## Appendix 4

### New variables created for summary survey-based organisational climate scores

Proportion of respondents answering 'yes' to whistle-blowing and reporting errors questions

Site	Knows how to report concerns	Believes concerns can be reported confidentially	Knows how to report errors, near-misses or incidents	<i>n</i>
1	0.855	0.667	0.889	36
2	0.667	0.519	0.778	27
3	0.855	0.675	0.964	83
4	0.947	0.789	1.000	38
5	0.800	0.629	0.886	35
6	0.929	0.821	0.964	28
7	0.765	0.647	0.882	17
8	1.000	0.857	0.714	7
9	0.865	0.595	0.973	37
10	0.933	0.767	0.867	30
11	0.920	0.640	0.960	25
12	0.815	0.704	0.852	27
13	0.952	0.714	0.952	21
14	0.880	0.840	1.000	25
15	0.925	0.725	0.975	40
16	0.892	0.782	0.964	55



## Appendix 5

### Strand B data collection prompt sheet

---

#### **Safety culture: observation checklist**

##### *Work environment*

Justification of staffing levels/skill mix for workload  
 Review of staffing/skill mix ratios  
 Staffing levels maintained at recommended level  
 Skill mix maintained at preset level  
 Systems in place to report instances where staffing falls below accepted level  
 Annual review of number/type and appropriateness of equipment for workload on unit  
 Action results from incident reports  
 Accessible documentation of list of unit's equipment  
 Equipment maintenance checks carried out regularly  
 Administrative support [visible, given and used appropriately]  
 Managerial support [visible, given, used appropriately]  
 Coordinator/consultant demonstrating team situational awareness, able to manage workload/staffing, etc.

##### *Team factors*

Supervision explicitly documented for junior staff on rotas  
 Prioritisation of teaching and learning/CPD (e.g. regular timetabled teaching sessions)  
 Regular multidisciplinary team handover  
 Decisions/plans reviewed at regular time points and when required  
 Inter professional respect/collaboration  
 Uniprofessional respect and collaboration  
 Staff prioritisation of social as well as task factors  
 Effective use of whiteboard for team situational awareness  
 Inclusively and ownership of handover amongst all staff  
 Inclusively and ownership of whiteboard amongst all staff  
 Staff seeking and receiving help from seniors/colleagues  
 Staff demonstrating error wisdom

##### *Individual staff factors*

Staff use of appropriate knowledge and skills to care for clients  
 Use of management skills to organise, anticipate and prioritise workload on unit  
 High levels of motivation amongst staff  
 Support for one another – pulling together

##### *Task factors*

Recently reviewed, locally agreed protocols relating to selected conditions  
 Safety policies and standards pertinent to the clinical area (e.g. local risk management policies, equipment maintenance policies, controlled drug checking policies, transfer standards)  
 Regular mortality and morbidity/clinical governance/audit meetings  
 Protocols, policies and guidelines; easily accessible  
 Whistle-blowing mechanism in place  
 Staff aware of existence of protocols, etc., and where to seek guidance  
 Public display of important safety information  
 Staff reading notices, updating

---

---

*Patient characteristics*

Access to appropriate support services (e.g. interpreters)

Infrastructure to support patient workload (i.e. size of unit, size and location of whiteboard)

---



## Appendix 6

# Scoring guide for observation-based evaluation of culture

---

### ***Domains for scoring***

Scores were as follows;

0 = consistent lack of features of a safety culture

1 = frequent lack of features of a safety culture

2 = frequent presence of features of a safety culture

3 = consistent presence of features of a safety culture

### ***1. Organisation and work environment factors***

#### *Staffing*

Documentary or key informant data on establishment, recruitment and sickness

Degree to which staff are busy/say that they are busy; use of bank staff/locums/overtime/need to call staff in or rebalance skill mix of team at short notice

#### *Premises/equipment*

Size of unit (space, capacity); equipment for tests and treatment; IT and telephone systems; public safety measures (emergency exits, hand-washing)

Information about premises, equipment, maintenance and upgrades

#### *Administrative, managerial and other support*

Availability and use of administrative and portering support, and of allied health professionals; presence of senior management (i.e. those working at level above frontline clinical team)

Support deficits of any sort

### ***2. Team factors***

#### *Informal training and supervision*

Informal teaching and learning especially for junior staff and students

#### *Leadership and responsibility*

Leadership offered by shift coordinator and consultants, and whether accepted

Situational awareness

Staff take responsibility for their own work

#### *Respect/warmth/collegiality*

Quality of interactions within and between professions

#### *Information exchange within the team*

Frequency and content of interactions about workload and patients/clients within and between professions

#### *Mutual support*

Personal and task support offered and received within and across professions

---



## Appendix 7

# Data extraction sheet for fractured neck of femur (audit)

Study case number		Presenting complaint		Date:	Earliest time noted		Initial assessment	
		Diagnosis			Allergies recorded		Medic assessment	

#	Standard	Looking for:				Comments
1	Pain score recorded on arrival	Pain score	Value:	Time:		
	If pain assessed, but not scored:	Pain assessed with descriptor <i>e.g. mild/moderate/severe pain?</i> Y/N		Details:	Time:	
	OR:	Pain assessed without descriptor <i>e.g. in pair?</i> Y/N		Details:	Time:	
	If no pain score or assessment:	Reason recorded in notes? Y/N		Details:	Time:	
2	Time from arrival to receive analgesia	Analgesia given Y/N	Time:	If analgesia not given, was it prescribed? Y/N		
	Analgesia examples: IV morphine, ketarolac, diclofenac, codeine phosphate, co-codamol. (Standard specifies 'If pain score >3', but record details requested here in <i>all</i> cases)					
3	Time from arrival to X-ray	X-ray done within 60 mins?	Yes: No:	Time:		
	If X-ray not ordered:	Reason recorded in notes?		Yes: No:		
4	If in pain, evidence of re-evaluation of pain <i>Aim for 30 mins if severe, 60 if moderate pain</i>	Pain re-evaluated?	Yes/No (circle one)	Time:		
	Evidence might be record of re-evaluation, or pain score or descriptor, or additional analgesia prescribed and/or given					
	If no pain re-assessment:	Reason recorded in notes?		Yes: No:	Time:	
	Further action if required	Additional analgesia given Y/N	Time:	If analgesia not given, was it prescribed? Y/N		
5	Time to ward	Time:				
<b>Comments</b>						



# Health Technology Assessment programme

**Director,**  
**Professor Tom Walley, CBE,**  
 Director, NIHR HTA programme,  
 Professor of Clinical Pharmacology,  
 University of Liverpool

**Deputy Director,**  
**Professor Hywel Williams,**  
 Professor of Dermato-Epidemiology,  
 Centre of Evidence-Based Dermatology,  
 University of Nottingham

## Prioritisation Group

### Members

**Chair,**  
**Professor Tom Walley, CBE,**  
 Director, NIHR HTA  
 programme, Professor of Clinical  
 Pharmacology, University of  
 Liverpool

Professor Imti Choonara,  
 Professor in Child Health,  
 Academic Division of Child  
 Health, University of Nottingham  
 Chair – Pharmaceuticals Panel

Dr Bob Coates,  
 Consultant Advisor – Disease  
 Prevention Panel

Dr Andrew Cook,  
 Consultant Advisor – Intervention  
 Procedures Panel

Dr Peter Davidson,  
 Director of NETSCC, Health  
 Technology Assessment

Dr Nick Hicks,  
 Consultant Adviser – Diagnostic  
 Technologies and Screening Panel,  
 Consultant Advisor–Psychological  
 and Community Therapies Panel

Ms Susan Hird,  
 Consultant Advisor, External  
 Devices and Physical Therapies  
 Panel

Professor Sallie Lamb,  
 Director, Warwick Clinical Trials  
 Unit, Warwick Medical School,  
 University of Warwick  
 Chair – HTA Clinical Evaluation  
 and Trials Board

Professor Jonathan Michaels,  
 Professor of Vascular Surgery,  
 Sheffield Vascular Institute,  
 University of Sheffield  
 Chair – Interventional Procedures  
 Panel

Professor Ruairidh Milne,  
 Director – External Relations

Dr John Pounsford,  
 Consultant Physician, Directorate  
 of Medical Services, North Bristol  
 NHS Trust  
 Chair – External Devices and  
 Physical Therapies Panel

Dr Vaughan Thomas,  
 Consultant Advisor –  
 Pharmaceuticals Panel, Clinical  
 Lead – Clinical Evaluation Trials  
 Prioritisation Group

Professor Margaret Thorogood,  
 Professor of Epidemiology, Health  
 Sciences Research Institute,  
 University of Warwick  
 Chair – Disease Prevention Panel

Professor Lindsay Turnbull,  
 Professor of Radiology, Centre for  
 the MR Investigations, University  
 of Hull  
 Chair – Diagnostic Technologies  
 and Screening Panel

Professor Scott Weich,  
 Professor of Psychiatry, Health  
 Sciences Research Institute,  
 University of Warwick  
 Chair – Psychological and  
 Community Therapies Panel

Professor Hywel Williams,  
 Director of Nottingham Clinical  
 Trials Unit, Centre of Evidence-  
 Based Dermatology, University of  
 Nottingham  
 Chair – HTA Commissioning  
 Board  
 Deputy HTA Programme Director

## HTA Commissioning Board

**Chair,**  
**Professor Hywel Williams,**  
 Professor of Dermato-Epidemiology,  
 Centre of Evidence-Based Dermatology,  
 University of Nottingham

**Deputy Chair,**  
**Professor Jon Deeks,**  
 Department of Public Health and  
 Epidemiology,  
 University of Birmingham

**Professor Tom Walley, CBE,**  
 Professor of Clinical Pharmacology,  
 Director, NIHR HTA programme,  
 University of Liverpool

### Members

Professor Ann Ashburn,  
 Professor of Rehabilitation and  
 Head of Research, Southampton  
 General Hospital

Professor Judith Bliss,  
 Director of ICR-Clinical Trials  
 and Statistics Unit, The Institute of  
 Cancer Research

Professor Peter Brocklehurst,  
 Professor of Women's Health,  
 Institute for Women's Health,  
 University College London

Professor David Fitzmaurice,  
 Professor of Primary Care  
 Research, Department of Primary  
 Care Clinical Sciences, University  
 of Birmingham

Professor John W Gregory,  
 Professor in Paediatric  
 Endocrinology, Department of  
 Child Health, Wales School of  
 Medicine, Cardiff University

Professor Steve Halligan,  
 Professor of Gastrointestinal  
 Radiology, University College  
 Hospital, London

Professor Angela Harden,  
 Professor of Community and  
 Family Health, Institute for  
 Health and Human Development,  
 University of East London

Dr Martin J Landray,  
 Reader in Epidemiology, Honorary  
 Consultant Physician, Clinical  
 Trial Service Unit, University of  
 Oxford

Dr Joanne Lord,  
 Reader, Health Economics  
 Research Group, Brunel University

Professor Stephen Morris,  
 Professor of Health Economics,  
 University College London,  
 Research Department of  
 Epidemiology and Public Health,  
 University College London

Professor Dion Morton,  
 Professor of Surgery, Academic  
 Department of Surgery, University  
 of Birmingham

Professor Gail Mountain,  
 Professor of Health Services  
 Research, Rehabilitation and  
 Assistive Technologies Group,  
 University of Sheffield

Professor Irwin Nazareth,  
 Professor of Primary Care and  
 Head of Department, Department  
 of Primary Care and Population  
 Sciences, University College  
 London

Professor E Andrea Nelson,  
 Professor of Wound Healing and  
 Director of Research, School of  
 Healthcare, University of Leeds

Professor John David Norrie,  
 Chair in Clinical Trials and  
 Biostatistics, Robertson Centre for  
 Biostatistics, University of Glasgow

Dr Rafael Perera,  
 Lecturer in Medical Statistics,  
 Department of Primary Health  
 Care, University of Oxford

## HTA Commissioning Board *(continued)*

Professor Barney Reeves,  
Professorial Research Fellow  
in Health Services Research,  
Department of Clinical Science,  
University of Bristol

Professor Peter Tyrer,  
Professor of Community  
Psychiatry, Centre for Mental  
Health, Imperial College London

Professor Martin Underwood,  
Professor of Primary Care  
Research, Warwick Medical  
School, University of Warwick

Professor Caroline Watkins,  
Professor of Stroke and Older  
People's Care, Chair of UK  
Forum for Stroke Training, Stroke  
Practice Research Unit, University  
of Central Lancashire

Dr Duncan Young,  
Senior Clinical Lecturer and  
Consultant, Nuffield Department  
of Anaesthetics, University of  
Oxford

### Observers

Dr Tom Foulks,  
Medical Research Council

Dr Kay Pattison,  
Senior NIHR Programme  
Manager, Department of Health

## HTA Clinical Evaluation and Trials Board

### Chair,

**Professor Sallie Lamb,**  
Director,  
Warwick Clinical Trials Unit,  
Warwick Medical School,  
University of Warwick and Professor of  
Rehabilitation,  
Nuffield Department of Orthopaedic,  
Rheumatology and Musculoskeletal Sciences,  
University of Oxford

### Deputy Chair,

**Professor Jenny Hewison,**  
Professor of the Psychology of Health Care,  
Leeds Institute of Health Sciences,  
University of Leeds

### Programme Director,

**Professor Tom Walley, CBE,**  
Director, NIHR HTA programme,  
Professor of Clinical Pharmacology,  
University of Liverpool

### Members

Professor Keith Abrams,  
Professor of Medical Statistics,  
Department of Health Sciences,  
University of Leicester

Professor Martin Bland,  
Professor of Health Statistics,  
Department of Health Sciences,  
University of York

Professor Jane Blazeby,  
Professor of Surgery and  
Consultant Upper GI Surgeon,  
Department of Social Medicine,  
University of Bristol

Professor Julia M Brown,  
Director, Clinical Trials Research  
Unit, University of Leeds

Professor Alistair Burns,  
Professor of Old Age Psychiatry,  
Psychiatry Research Group, School  
of Community-Based Medicine,  
The University of Manchester &  
National Clinical Director for  
Dementia, Department of Health

Dr Jennifer Burr,  
Director, Centre for Healthcare  
Randomised trials (CHART),  
University of Aberdeen

Professor Linda Davies,  
Professor of Health Economics,  
Health Sciences Research Group,  
University of Manchester

Professor Simon Gilbody,  
Prof of Psych Medicine and Health  
Services Research, Department of  
Health Sciences, University of York

Professor Steven Goodacre,  
Professor and Consultant in  
Emergency Medicine, School of  
Health and Related Research,  
University of Sheffield

Professor Dyfrig Hughes,  
Professor of Pharmacoeconomics,  
Centre for Economics and Policy  
in Health, Institute of Medical  
and Social Care Research, Bangor  
University

Professor Paul Jones,  
Professor of Respiratory Medicine,  
Department of Cardiac and  
Vascular Science, St George's  
Hospital Medical School,  
University of London

Professor Khalid Khan,  
Professor of Women's Health and  
Clinical Epidemiology, Barts and  
the London School of Medicine,  
Queen Mary, University of London

Professor Richard J McManus,  
Professor of Primary Care  
Cardiovascular Research, Primary  
Care Clinical Sciences Building,  
University of Birmingham

Professor Helen Rodgers,  
Professor of Stroke Care, Institute  
for Ageing and Health, Newcastle  
University

Professor Ken Stein,  
Professor of Public Health,  
Peninsula Technology Assessment  
Group, Peninsula College  
of Medicine and Dentistry,  
Universities of Exeter and  
Plymouth

Professor Jonathan Sterne,  
Professor of Medical Statistics  
and Epidemiology, Department  
of Social Medicine, University of  
Bristol

Mr Andy Vail,  
Senior Lecturer, Health Sciences  
Research Group, University of  
Manchester

Professor Clare Wilkinson,  
Professor of General Practice and  
Director of Research North Wales  
Clinical School, Department of  
Primary Care and Public Health,  
Cardiff University

Dr Ian B Wilkinson,  
Senior Lecturer and Honorary  
Consultant, Clinical Pharmacology  
Unit, Department of Medicine,  
University of Cambridge

### Observers

Ms Kate Law,  
Director of Clinical Trials,  
Cancer Research UK

Dr Morven Roberts,  
Clinical Trials Manager, Health  
Services and Public Health  
Services Board, Medical Research  
Council

## Diagnostic Technologies and Screening Panel

### Members

<p><b>Chair,</b> <b>Professor Lindsay Wilson Turnbull,</b> Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary</p> <p>Professor Judith E Adams, Consultant Radiologist, Manchester Royal Infirmary, Central Manchester &amp; Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester</p> <p>Mr Angus S Arunkalaivanan, Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham</p> <p>Dr Diana Baralle, Consultant and Senior Lecturer in Clinical Genetics, University of Southampton</p>	<p>Dr Stephanie Dancer, Consultant Microbiologist, Hairmyres Hospital, East Kilbride</p> <p>Dr Diane Eccles, Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital</p> <p>Dr Trevor Friedman, Consultant Liaison Psychiatrist, Brandon Unit, Leicester General Hospital</p> <p>Dr Ron Gray, Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford</p> <p>Professor Paul D Griffiths, Professor of Radiology, Academic Unit of Radiology, University of Sheffield</p> <p>Mr Martin Hooper, Public contributor</p>	<p>Professor Anthony Robert Kendrick, Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton</p> <p>Dr Nicola Lennard, Senior Medical Officer, MHRA</p> <p>Dr Anne Mackie, Director of Programmes, UK National Screening Committee, London</p> <p>Mr David Mathew, Public contributor</p> <p>Dr Michael Millar, Consultant Senior Lecturer in Microbiology, Department of Pathology &amp; Microbiology, Barts and The London NHS Trust, Royal London Hospital</p> <p>Mrs Una Rennard, Public contributor</p>	<p>Dr Stuart Smellie, Consultant in Clinical Pathology, Bishop Auckland General Hospital</p> <p>Ms Jane Smith, Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds</p> <p>Dr Allison Streetly, Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine</p> <p>Dr Matthew Thompson, Senior Clinical Scientist and GP, Department of Primary Health Care, University of Oxford</p> <p>Dr Alan J Williams, Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital</p>
--	--	---	---

### Observers

<p>Dr Tim Elliott, Team Leader, Cancer Screening, Department of Health</p> <p>Dr Joanna Jenkinson, Board Secretary, Neurosciences and Mental Health Board (NMHB), Medical Research Council</p>	<p>Professor Julietta Patnick, Director, NHS Cancer Screening Programme, Sheffield</p> <p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>	<p>Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health</p>
--	--	---	---

## Disease Prevention Panel

### Members

<p><b>Chair,</b> <b>Professor Margaret Thorogood,</b> Professor of Epidemiology, University of Warwick Medical School, Coventry</p> <p>Dr Robert Cook, Clinical Programmes Director, Bazian Ltd, London</p> <p>Dr Colin Greaves, Senior Research Fellow, Peninsula Medical School (Primary Care)</p> <p>Mr Michael Head, Public contributor</p>	<p>Professor Cathy Jackson, Professor of Primary Care Medicine, Bute Medical School, University of St Andrews</p> <p>Dr Russell Jago, Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol</p> <p>Dr Julie Mytton, Consultant in Child Public Health, NHS Bristol</p>	<p>Professor Irwin Nazareth, Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London</p> <p>Dr Richard Richards, Assistant Director of Public Health, Derbyshire County Primary Care Trust</p> <p>Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene &amp; Tropical Medicine</p>	<p>Dr Kenneth Robertson, Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow</p> <p>Dr Catherine Swann, Associate Director, Centre for Public Health Excellence, NICE</p> <p>Mrs Jean Thurston, Public contributor</p> <p>Professor David Weller, Head, School of Clinical Science and Community Health, University of Edinburgh</p>
---	--	--	--

### Observers

<p>Ms Christine McGuire, Research &amp; Development, Department of Health</p>	<p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>
---	---	---

## External Devices and Physical Therapies Panel

### Members

<b>Chair,</b> <b>Dr John Pounsford,</b> Consultant Physician North Bristol NHS Trust	Dr Dawn Carnes, Senior Research Fellow, Barts and the London School of Medicine and Dentistry	Dr Shaheen Hamdy, Clinical Senior Lecturer and Consultant Physician, University of Manchester	Mr Jim Reece, Public contributor
<b>Deputy Chair,</b> <b>Professor E Andrea Nelson,</b> Reader in Wound Healing and Director of Research, University of Leeds	Dr Emma Clark, Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol	Professor Christine Norton, Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust	Professor Maria Stokes, Professor of Neuromusculoskeletal Rehabilitation, University of Southampton
Professor Bipin Bhakta, Charterhouse Professor in Rehabilitation Medicine, University of Leeds	Mrs Anthea De Barton-Watson, Public contributor	Dr Lorraine Pinnigton, Associate Professor in Rehabilitation, University of Nottingham	Dr Pippa Tyrrell, Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester
Mrs Penny Calder, Public contributor	Professor Nadine Foster, Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University	Dr Kate Radford, Senior Lecturer (Research), University of Central Lancashire	Dr Nefyn Williams, Clinical Senior Lecturer, Cardiff University

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
---	---	---	---

## Interventional Procedures Panel

### Members

<b>Chair,</b> <b>Professor Jonathan Michaels,</b> Professor of Vascular Surgery, University of Sheffield	Mr Seumas Eckford, Consultant in Obstetrics & Gynaecology, North Devon District Hospital	Dr Fiona Lecky, Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust	Professor Jon Moss, Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust
<b>Deputy Chair,</b> <b>Mr Michael Thomas,</b> Consultant Colorectal Surgeon, Bristol Royal Infirmary	Professor Sam Eljamel, Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee	Dr Nadim Malik, Consultant Cardiologist/Honorary Lecturer, University of Manchester	Dr Simon Padley, Consultant Radiologist, Chelsea & Westminster Hospital
Mrs Isabel Boyer, Public contributor	Dr Adele Fielding, Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School	Mr Hisham Mehanna, Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust	Dr Ashish Paul, Medical Director, Bedfordshire PCT
Mr Sankaran Chandra Sekharan, Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust	Dr Matthew Hatton, Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust	Dr Jane Montgomery, Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust	Dr Sarah Purdy, Consultant Senior Lecturer, University of Bristol
Professor Nicholas Clarke, Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust	Dr John Holden, General Practitioner, Garswood Surgery, Wigan		Dr Matthew Wilson, Consultant Anaesthetist, Sheffield Teaching Hospitals NHS Foundation Trust
Ms Leonie Cooke, Public contributor			Professor Yit Chiun Yang, Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
---	---	---	---



## Pharmaceuticals Panel

### Members

<b>Chair,</b> <b>Professor Imti Choonara,</b> Professor in Child Health, University of Nottingham	Dr James Gray, Consultant Microbiologist, Department of Microbiology, Birmingham Children's Hospital NHS Foundation Trust	Dr Maria Kouimtzi, Pharmacy and Informatics Director, Global Clinical Solutions, Wiley-Blackwell	Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool
<b>Deputy Chair,</b> <b>Dr Yoon K Loke,</b> Senior Lecturer in Clinical Pharmacology, University of East Anglia	Dr Jurjees Hasan, Consultant in Medical Oncology, The Christie, Manchester	Professor Femi Oyeboode, Consultant Psychiatrist and Head of Department, University of Birmingham	Professor Donald Singer, Professor of Clinical Pharmacology and Therapeutics, Clinical Sciences Research Institute, CSB, University of Warwick Medical School
Dr Martin Ashton-Key, Medical Advisor, National Commissioning Group, NHS London	Dr Carl Heneghan, Deputy Director Centre for Evidence-Based Medicine and Clinical Lecturer, Department of Primary Health Care, University of Oxford	Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician and Gynaecologist, The Rosie Hospital, University of Cambridge	Mr David Symes, Public contributor
Dr Peter Elton, Director of Public Health, Bury Primary Care Trust	Dr Dyfrig Hughes, Reader in Pharmacoeconomics and Deputy Director, Centre for Economics and Policy in Health, IMSCaR, Bangor University	Ms Amanda Roberts, Public contributor	Dr Arnold Zermansky, General Practitioner, Senior Research Fellow, Pharmacy Practice and Medicines Management Group, Leeds University
Dr Ben Goldacre, Research Fellow, Epidemiology London School of Hygiene and Tropical Medicine		Dr Gillian Shepherd, Director, Health and Clinical Excellence, Merck Serono Ltd	

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Heike Weber, Programme Manager, Medical Research Council	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
Mr Simon Reeve, Head of Clinical and Cost- Effectiveness, Medicines, Pharmacy and Industry Group, Department of Health	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	

## Psychological and Community Therapies Panel

### Members

<b>Chair,</b> <b>Professor Scott Weich,</b> Professor of Psychiatry, University of Warwick, Coventry	Mrs Val Carlill, Public contributor	Dr Jeremy J Murphy, Consultant Physician and Cardiologist, County Durham and Darlington Foundation Trust	Dr Paul Ramchandani, Senior Research Fellow/Cons. Child Psychiatrist, University of Oxford
<b>Deputy Chair,</b> <b>Dr Howard Ring,</b> Consultant & University Lecturer in Psychiatry, University of Cambridge	Dr Steve Cunningham, Consultant Respiratory Paediatrician, Lothian Health Board	Dr Richard Neal, Clinical Senior Lecturer in General Practice, Cardiff University	Dr Karen Roberts, Nurse/Consultant, Dunston Hill Hospital, Tyne and Wear
Professor Jane Barlow, Professor of Public Health in the Early Years, Health Sciences Research Institute, Warwick Medical School	Dr Anne Hesketh, Senior Clinical Lecturer in Speech and Language Therapy, University of Manchester	Mr John Needham, Public contributor	Dr Karim Saad, Consultant in Old Age Psychiatry, Coventry and Warwickshire Partnership Trust
Dr Sabyasachi Bhaumik, Consultant Psychiatrist, Leicestershire Partnership NHS Trust	Dr Peter Langdon, Senior Clinical Lecturer, School of Medicine, Health Policy and Practice, University of East Anglia	Ms Mary Nettle, Mental Health User Consultant	Dr Lesley Stockton, Lecturer, School of Health Sciences, University of Liverpool
	Dr Yann Lefeuvre, GP Partner, Burrage Road Surgery, London	Professor John Potter, Professor of Ageing and Stroke Medicine, University of East Anglia	Dr Simon Wright, GP Partner, Walkden Medical Centre, Manchester
		Dr Greta Rait, Senior Clinical Lecturer and General Practitioner, University College London	

### Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---





### **Feedback**

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

***We look forward to hearing from you.***