

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,^{1*} J Sandall,² T Allan,³ F Warburton,⁴
EJ Berridge,³ N Mackintosh,⁵ M Rogers³ and
S Abbott³

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

²Division of Women's Health, School of Medicine, King's College London, London, UK

³School of Health Sciences, City University London, London, UK

⁴Department of Psychological Medicine, Institute of Psychiatry, King's College London, London, UK

⁵Patient Safety and Service Quality Research Centre, King's College London, London, UK

*Corresponding author



Executive summary

Health Technology Assessment 2012; Vol. 16: No. 25
DOI: 10.3310/hta16250

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk



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.

Publication

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).

NIHR Health Technology Assessment programme

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 needed 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

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), on behalf of NETSCC, HTA.

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