

BMJ Open Capturing intraoperative process deviations using a direct observational approach: the glitch method

Lauren Morgan,¹ Eleanor Robertson,¹ Mohammed Hadi,² Ken Catchpole,³ Sharon Pickering,² Steve New,⁴ Gary Collins,⁵ Peter McCulloch¹

To cite: Morgan L, Robertson E, Hadi M, *et al*. Capturing intraoperative process deviations using a direct observational approach: the glitch method. *BMJ Open* 2013;**3**:e003519. doi:10.1136/bmjopen-2013-003519

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-003519>).

Received 4 July 2013
Accepted 16 October 2013



CrossMark

¹Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

²Warwick Orthopaedics, Warwick Medical School, Coventry, UK

³Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, California, USA

⁴Saïd Business School, University of Oxford, Oxford, UK

⁵Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Correspondence to

Dr Peter McCulloch; peter.mcculloch@nds.ox.ac.uk

ABSTRACT

Objectives: To develop a sensitive, reliable tool for enumerating and evaluating technical process imperfections during surgical operations.

Design: Prospective cohort study with direct observation.

Setting: Operating theatres on five sites in three National Health Service Trusts.

Participants: Staff taking part in elective and emergency surgical procedures in orthopaedics, trauma, vascular and plastic surgery; including anaesthetists, surgeons, nurses and operating department practitioners.

Outcome measures: Reliability and validity of the glitch count method; frequency, type, temporal pattern and rate of glitches in relation to site and surgical specialty.

Results: The glitch count has construct and face validity, and category agreement between observers is good ($\kappa=0.7$). Redundancy between pairs of observers significantly improves the sensitivity over a single observation. In total, 429 operations were observed and 5742 glitches were recorded (mean 14 per operation, range 0–83). Specialty-specific glitch rates varied from 6.9 to 8.3/h of operating (ns). The distribution of glitch categories was strikingly similar across specialties, with distractions the commonest type in all cases. The difference in glitch rate between specialty teams operating at different sites was larger than that between specialties (range 6.3–10.5/h, $p<0.001$). Forty per cent of glitches occurred in the first quarter of an operation, and only 10% occurred in the final quarter.

Conclusions: The glitch method allows collection of a rich dataset suitable for analysing the changes following interventions to improve process safety, and appears reliable and sensitive. Glitches occur more frequently in the early stages of an operation. Hospital environment, culture and work systems may influence the operative process more strongly than the specialty.

INTRODUCTION

The delivery of safe surgical care is technically challenging. It requires considered

Strengths and limitations of this study

- The use of disparate observers in a very large prospective direct observation study is likely to have resulted in high sensitivity and power to detect the associations and differences between subgroups.
- Direct observation methods are vulnerable to the Hawthorne effect, although subjectively this did not appear important.
- Findings about glitch associations and patterns of occurrence have raised new questions about the underlying causes of process deviations in theatre and their relationships to adverse outcomes for patients.

patient selection, preoperative assessment and the coordination of technical expertise and resource. These requirements result in the creation of a challenging work environment which puts healthcare workers and the enveloping healthcare system under considerable stress. The resultant iatrogenic patient harm has been recently estimated at 6.2%, with a half of this occurring in surgical patients.¹

The retrospective identification of non-operative procedural undesirable events or glitches in surgery is often not possible. The loss of detail of the nature and occurrence of events may result in biased or unrepresentative incident reporting and analysis.^{2–3} The direct observation of a process provides an opportunity to gather prospective insight into areas of systematic weakness which may benefit from improvement interventions. Investigations of the origin of iatrogenic events have previously used non-technical skills rating scales^{4–6} and system event observations⁷ to frame quantifiable standard descriptions of these two aspects of surgical team performance.

In recent years, the analysis of undesirable surgical outcomes has revealed a number of contributory factors beyond the traditionalist

view of individual accountability.^{8,9} Human fallibility and underlying organisational failings contribute to clinical interface errors.¹⁰ Failures of compliance, communication and procedure design can all contribute to error in complex group tasks such as surgery.^{11,12} Previous observational studies have suggested that serious safety and quality issues may result from accumulation of small observable process deviations in higher risk procedures such as paediatric cardiac surgery.^{8,13} There is little systematic analysis of the magnitude of the impact of these events on patient outcome. The quantification of the influence of technical process and outcome has been further hampered due to lack of standardisation in data collection approaches. Various descriptive terms have been used to describe unintended events during the operative process including ‘minor problems’ and ‘operating problems’;¹⁴ ‘surgical flow disruptions’;^{15,16} and ‘intraoperative interference’.¹⁷ To conduct high-volume comparative evaluations of interventions to improve operating theatre safety and reliability, we needed direct observational methods to measure process fidelity which are applicable across a wide range of specialties, techniques and settings, not making assumptions about causality and not attempting to estimate the impact on outcome. Direct observation methods for evaluating theatre team technical performance have been published by others^{6,16,18,19} and the principles used are very similar in several of these to those on which we had based some of our own previous work.^{4,18} Rather than tackling the difficulties of learning, adapting and validating a new system, we decided to develop our own based

on our previous studies. We previously reported the initial development of this method including reliability assessment and taxonomy development.²⁰ In this study we report a large-scale evaluation of its use to characterise the relationship of glitches to the context, in terms of specialty, site and operative duration, their temporal pattern during operations and the relative frequency of the different categories.

METHODS

Development and validation of methodology

The observational and categorisation methods were based on those developed by others:^{14–16} the development of the glitch categories and the testing of interobserver reliability have been described previously.¹⁹ The method involves two observers, one with human factors (HFs) and one with a surgical background, observing entire operative procedures and noting any deviations from the expected or planned course. To enable them to do this, they trained together and used a predesigned procedure template as a guide. A sample set of 94 glitches from 10 elective orthopaedic operations were collected during the initial 3-month training phase, grouped in common themes, and assigned titles and definitions (table 1). The reliability of the observers categorisation was assessed using Cohen’s κ and was good between the four observers (0.70, 95% CI 0.66 to 0.75).

Within a larger pilot sample of operations (42 elective orthopaedic procedures) the rate of glitches per operation ranged from 1 to 18, with an average of 8 per

Table 1 Glitch categories with definition and examples

Glitch category	Definition	Examples
Absence	Absence of theatre staff member, when required	Circulating nurse not available to get equipment
Communication	Difficulties in communication among team members	Repeat requests, incorrect terminology and misinterpretations
Distractions	Anything causing distraction from task	Phone calls/bleeps, loud music requiring to be turned down
Environment	Aspects of the working environment causing difficulty	Low lighting or variable temperature during operation causing difficulties
Equipment design	Issues arising from equipment design, that would not otherwise be corrected with training or maintenance	Compatibility problems with different implant systems; equipment blockage
Maintenance	Faulty or poorly maintained equipment	Battery depleted during use, blunt equipment
Health and safety	Any observed physical risk to personnel	Mask violations, food/drink in theatre
Planning and preparation	Instances that may otherwise been avoided with appropriate prior planning and preparation	Insufficient equipment resources, staffing levels and training
Patient related	Issues relating to the physiological status of the patient	Difficulty in extracting previous implants, unexpected anatomically related surgical difficulty and anaphylaxis
Process deviation	Incomplete or reordered completion of standard tasks	Unnecessary equipment opened
Slips	Psychomotor errors	Dropped instruments
Training	Repetition or delay of operative steps due to training	Consultant corrects assistants operating technique
Workspace	Equipment or theatre layout issues	Desterilising of equipment/scrubbed staff on environment

operation.²⁰ In contrast with previous methodologies,¹⁴ no immediate evaluation of the glitch significance was made, as the impact of a particular glitch on process or outcome is context dependent. Prior to the final analysis, the glitch data were reviewed jointly by the observers (LM, MH, SP and ER). Glitches noted by all observers were categorised by consensus where there was a difference. Some glitches were deleted (if the team considered this event was not a glitch), split (if the contextual data contained more than one glitch occurrence) or recategorised during this consensus process. An overall glitch score was assigned comprising the sum of all unique glitches seen (ie, those unique to observer A + those unique to observer B + those in common).

Observer background, training and context

The clinical observers had a clinical qualification (surgical trainees and operating department practitioners) and more than 1 years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HFs. The HF observers were orientated to the operating theatre environment and learnt the technical aspects of the operative process through observation and through coaching from the clinical observers. The clinical observers were introduced to the HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from the HF observers while in theatre.

Glitch count observation method

Glitches were defined as deviations from the recognised process with the potential to reduce quality or speed, including interruptions, omissions and changes, whether or not these actually affected the outcome of the procedure. To capture these, direct observations were made of all activities (surgical, nursing and anaesthetic) in the operating theatre from the time the patient entered to the time they left, by pairs of six observers comprising of one clinical and one HF researcher. Four of the six observers (MH, SP, ER and LM) were involved in the creation of the method, the remainder (Laura Bleakley and Julia Matthews) were introduced to the categorisation at a later date. Any process disruption which occurred in the pretheatre or post-theatre phase were not included in this method as it was thought that the collection of these events would not be as reliable as those collected in the intraoperative period. The clinical observers developed a process map of the main operation types to be observed, which took the form of a descriptive list of the operative process, including relevant procedures and steps. These process maps formed the basis for the training and subsequent structured observation.²¹ The glitches were collected independently by each observer, individually noting the time and detail of the glitch within data collection booklets. This results in a set of glitches captured by each observer. These are deduplicated and summed to provide a total glitch

count for an operation. We recorded the detail of the glitch (eg, 'diathermy not plugged in when surgeon trying to use it') along with the associated time point. All glitches were categorised post hoc and entered into a secure database. The observers spent a period of 1 month in training and orientation to the data collection methods before any real-time data were collected. Alongside the collection of glitches, the observers also assessed the teams' non-technical skills, WHO surgical safety checklist adherence and recorded the operative duration as part of a larger programme of work. Non-compliance with the WHO surgical safety checklist was not considered within the glitch scale.

Patients were informed of the possibility of observations taking place and were given the opportunity to opt out if they wished. Staff in the theatres undergoing observation were given information on the study and asked for consent before observations took place.

Large scale evaluation of glitch method

Intraoperative observations were made across five UK National Health Service sites and four specialties: elective orthopaedics, trauma orthopaedics, vascular, general and plastic surgery. The sample included a tertiary referral centre (site A), two university teaching hospitals (sites C and E) and two district general hospitals (sites B and D). Elective orthopaedics was chosen as the main interhospital comparator specialty due to the homogeneity of operation technique and duration. As the Safer Delivery of Surgical Services (S3) study was designed to test the effectiveness of surgical improvement interventions, with both active and control groups from the same hospital site, there were occasions where it was not possible to adequately separate the theatre teams from within one specialty. On these occasions, other surgical specialties were recruited to the study, which in turn enabled the evaluation of the glitch and other intraoperative observational techniques across surgical specialties. This permitted the comparison of volume and profile of glitches across the sites and the specialties.

Whole operating lists were observed wherever practical, with lists being preferred if they contained standardised operations (eg, primary or revision total knee and hip arthroplasty). If a patient left the theatre during mid-operation (eg, to radiology), the observations were paused until the patient returned.

Data analysis

Differences in mean glitch rates per operation between the sites and specialties were examined by one-way analysis of variance and t tests. We considered p values <0.05 to be statistically significant (with no adjustment for multiple testing). All analyses were carried out using R-2.15.2.

RESULTS

A total of 429 operations were observed between November 2010 and July 2012, and 5742 glitches were

observed. The total number of glitches observed in a single operation ranged from 0 to 83 (mean 14).

We investigated possible differences in the profile of glitches that each observer collected in the theatre (table 2). Of the 5742 glitches, 64% were observed by the HF observers and 76% by the clinical observers ($p \leq 0.001$). The clinical observers consistently noted more glitches per operation than the HF (table 1) but the difference varied markedly between glitch categories. The clinical observers noted a much larger proportion of environment, training, health and Safety and patient related glitches, while there was a minimal difference between the observers for absence, slips and equipment maintenance.

Observed glitches by site and speciality

The number of procedures observed in different sites and specialities is shown in table 3.

Site A was the primary site of study and therefore more operations were observed there. The operative duration was similar across the orthopaedic operations, much more variable for plastic surgery and longer on average in vascular operations. The average total glitch count per operation was 14, range 0–83. The number of glitches per operation by specialty ranged from 1 to 63 in elective orthopaedic surgery, 1 to 35 in trauma orthopaedics, 2 to 49 in elective vascular surgery and 1 to 83 in elective plastic surgery. Owing to the range of operation duration, both within and between specialties, a glitch rate is required to facilitate the comparison. It is possible to determine a glitch rate per hour for each operation, calculated by the total number of glitches per operation divided by the length of the operation. The

distribution in glitch rates across all the operations observed can be seen in figure 1.

Although there is a strong clustering around the mean, the data are skewed, with a number of operations with high glitch rates at $>20/h$. The mean glitch rate for orthopaedics is 7.6 (range 0.4–28.4), trauma orthopaedics is 6.9 (range 1.3–15.3), vascular is 8.3 (range 1.5–20.6) and plastics is 7.1 (range 0.7–28). There was no statistically significant difference in average glitch rate across the four specialties ($p=0.453$).

Relationship between glitch category and speciality

As the glitches are categorised, it is possible to compare distribution across the categories for the different specialties (figure 2).

The profile of the glitch categories between the surgical specialties is strikingly similar. It can be seen that the most common glitches across all specialties are distractions, and planning and preparation. The rate of distractions is nearly twice that of any other category for all but trauma orthopaedics. The lowest frequency of glitches is that relating to the patient and the environment. The rate of distractions is higher in plastic surgery which relates anecdotally to the discursive and fluid nature of the teams involved. There is a higher rate of maintenance and absence of glitches for trauma orthopaedics and a low level of slips in plastics when compared with the other specialties.

Relationship between glitch rate and hospital site

Elective orthopaedic and vascular surgical procedures were observed in multiple sites (4 and 2 sites, respectively), providing an opportunity for intersite glitch rate

Table 2 Difference in observed glitches between observer specialties

Glitch category	Total observed n (% of total)	Observed by both HFs and clinical n (% of category)	HFs observed n (% of category)	Clinical observed n (% of category)	Difference % (95% CI)	p Value
Absence	292 (5.1)	123 (42.1)	202 (69.2)	213 (72.9)	3.8 (–3.9 to 11.5)	0.362
Communication	334 (5.8)	128 (38.3)	218 (65.3)	244 (73.1)	7.8 (0.5 to 15.1)	0.036
Distractions	1342 (23.4)	585 (43.6)	887 (66.1)	1039 (77.4)	11.3 (7.9 to 14.8)	<0.001
Environment	15 (0.3)	5 (33.3)	8 (53.3)	12 (80.0)	26.7 (–12.4 to 65.7)	0.245
Equipment design	595 (10.4)	224 (37.6)	379 (63.7)	440 (73.9)	10.3 (4.9 to 15.7)	<0.001
Equipment maintenance	278 (4.8)	146 (52.5)	206 (74.1)	218 (78.4)	4.3 (–3.1 to 11.7)	0.273
Health and safety	423 (7.4)	171 (40.4)	243 (57.4)	350 (82.7)	25.3 (19.1 to 31.5)	<0.001
Patient related	120 (2.1)	36 (30.0)	49 (40.8)	107 (89.2)	48.3 (37.1 to 59.6)	<0.001
Planning and preparation	789 (13.7)	304 (38.5)	495 (62.7)	596 (75.5)	12.8 (8.2 to 17.4)	<0.001
Process deviation	614 (10.7)	227 (37.0)	375 (61.1)	465 (75.7)	14.7 (9.4 to 20.09)	<0.001
Slips	508 (8.8)	256 (50.4)	386 (76.0)	377 (74.2)	–1.8 (–7.3 to 3.7)	0.562
Training	154 (2.7)	36 (23.4)	70 (45.5)	120 (77.9)	32.5 (21.6 to 43.4)	<0.001
Workspace	278 (4.8)	67 (24.1)	165 (59.4)	180 (64.7)	5.4 (–3.0 to 13.8)	0.221
Overall	5742	2308 (40.2)	3683 (64.1)	4361 (75.9)	11.8 (10.1 to 13.5)	<0.001

HFs, human factors.

Table 3 Sample characteristics of observations

	Site A	Site B	Site C	Site D	Site E
Number of operations (mean, range of operating duration; hh:mm)	175 (1:56, 0:27–13:32)	63 (2:02, 0:27–4:45)	96 (1:47, 0:22–3:58)	72 (1:46, 0:30–3:55)	24 (3:11, 0:43–7:28)
Elective orthopaedics	130 (1:49, 0:27–4:25)	63 (2:02, 0:27–4:45)	54 (1:49, 0:27–3:41)	51 (1:54, 1:01–3:55)	0
Trauma orthopaedics	0	0	42 (1:44, 0:22–3:58)	0	0
Plastic surgery	45 (2:17, 0:30–13:32)	0	0	0	0
Vascular surgery	0	0	0	21 (1:25, 0:30–2:24)	24 (3:11, 0:43–7:28)

comparison among teams performing the same types of surgery.

In elective orthopaedics, the mean operating duration varies by 14 min between the sites, with the glitch rate varying between 6 and 8 glitches/hour. There was a statistically significant heterogeneity in mean glitch rates per operation between the four sites ($p < 0.001$). This was explained in 1–1 comparisons by significant differences between individual sites A (mean of 8.1 glitches/hour) and B (mean of 6.0 glitches/hour; difference=2.1; 95% CI 0.9 to 3.3; $p = 0.001$), sites D (mean of 8.7 glitches/hour) and B (mean of 6.0 glitches/hour; difference=2.7; 95% CI 1.4 to 4; $p < 0.001$), and sites C (mean of 7.3 glitches/hour) and B (mean of 6.0 glitches/hour; difference=1.2; 95% CI 0.01 to 2.6; $p = 0.047$). No other statistically significant differences between centres were observed.

In vascular surgery, the difference in the glitch rate between site E (mean of 6.3 glitches/hour) and site D (mean of 10.5 glitches/hour) was highly significant, $p = 0.0003$.

Relationship between glitch occurrence and stage of operation

Glitches were recorded alongside the time at which they occurred, which can be referenced to the start (patient enters theatre) and end time (patient leaves theatre) of the operation. To enable the cross-comparison between operations of different lengths, the glitch timings were normalised to operative duration so that the operative duration total is 100%, halfway through is 50% and so on. Analysing the spread of the occurrence of glitches across each operation allows for interpretation of any trends. The graphical representation illustrates a flattened sigmoid relationship between the glitches and the duration of operation, suggesting a reduction in glitch occurrence in the last 20% of operation duration (figure 3).

Elective orthopaedics and trauma orthopaedics both follow a similar linear trend in the first half of any operation: slightly more than 60% of glitches occur in this time (figure 3). Vascular and plastic surgery appear to have more glitches in the earlier stages of the operation, with nearly 40% of their total glitches occurring within the first 25% of the operation. For vascular the early accumulation of glitches continues with 75% of the glitches occurring by the halfway point of the operation. The accrual of glitches reduces markedly during the last 25% of the operation, with only 10% of the total glitches happening in this period.

DISCUSSION

There is an increasing acceptance within the healthcare community that to achieve safe and reliable systems of care requires the same scrutiny that has previously been directed at healthcare professionals' behaviour and technical skill.^{22–24} The prospective collection of information about process imperfections or deviations enables healthcare researchers to analyse intraoperative events

Figure 1 Distribution of glitch rate by operation.



so that the system and the operative technique can be evaluated for quality and risk. The observed events include different subclasses such as distractions, process deviations, equipment design problems and slips (table 1). These glitches are not necessarily associated with immediate consequences or to any failure in the surgical team. However, they reflect the additional, unplanned and often unnecessary activity within the operating theatre. Lowering the total number of imperfections in the process may be advantageous for the patient safety, as it may preserve the team’s capacity for dealing with the unexpected events.²⁵ Although not

tested in this study, it has previously been suggested that the accumulation of ‘minor’ events predisposes to a ‘major’ event associated with the potential for serious patient harm.¹⁴ We did not seek to prove the causality of glitches as we felt that it would be unwise to attempt to link a glitch with what could be multiple upstream factors. We consider that some glitch categories may correlate with patient harm events more than others; however, we did not test this hypothesis in this study.

Measuring the prevalence of glitches provides a quantitative practical insight into the effects of system malfunctions on the process and on the healthcare

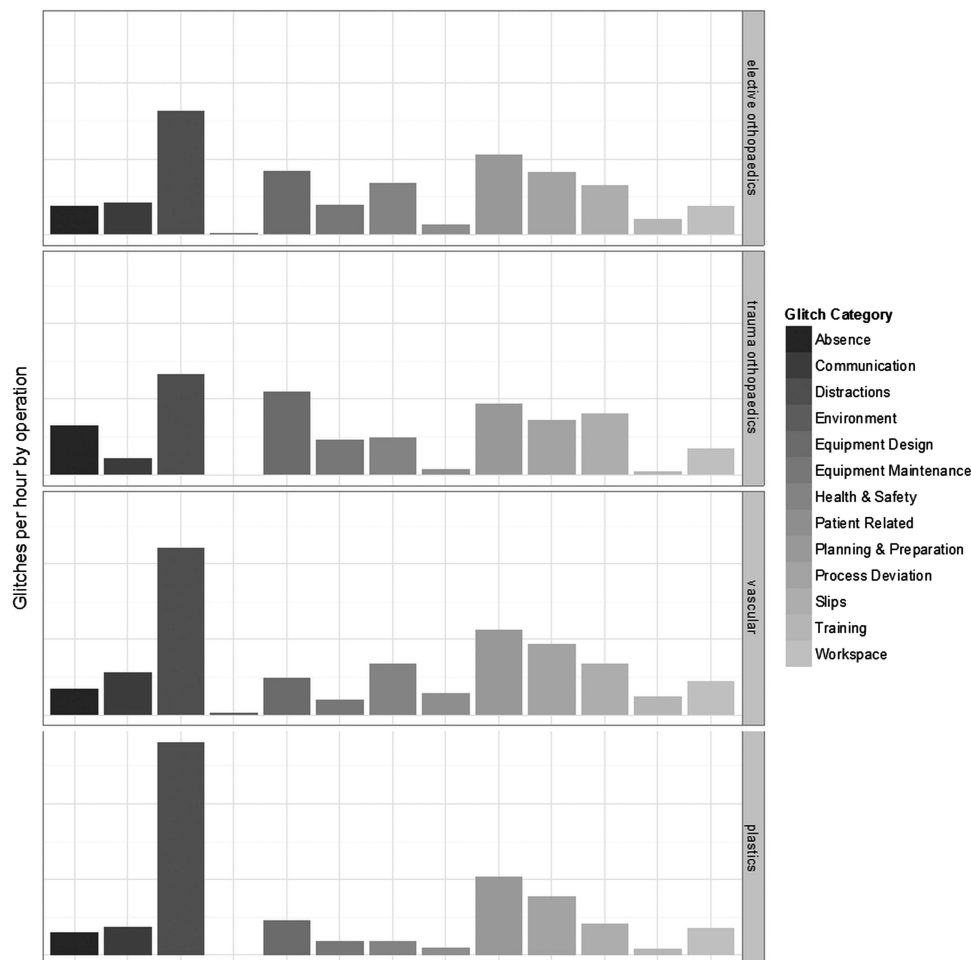
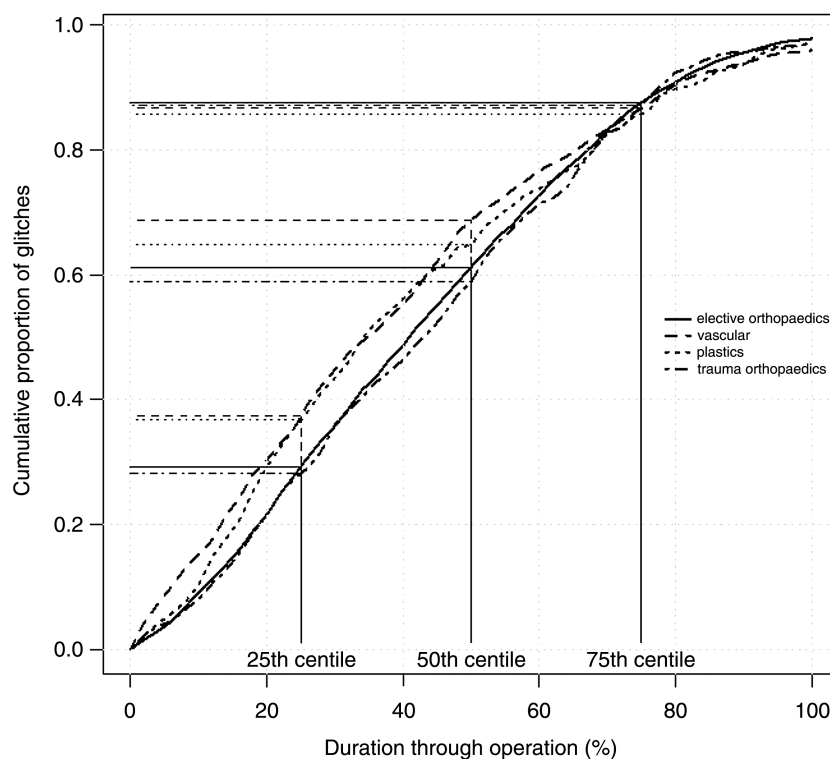


Figure 2 Mean glitch rate by operation for each specialty.

Figure 3 Distribution of glitch occurrence across operative duration.



professionals who are delivering care. When observing theatre teams in action, there is some overlap between the assessment of non-technical skills and the recording of the technical process imperfections what we have called as glitches. There may be circumstances where there is blending of the system and HFs. For example, a planning and preparation glitch may arise due to a last-minute change of plan, giving an impression of generalised low situational awareness. However, this situation may arise due to a lack of allocated prelist briefing time forced by time constraints. The interplay of non-technical skills and systems issues is as yet not fully understood, and some measurement systems have attempted to incorporate both.⁶

We describe the development of an operating theatre whole-system assessment method focused on technical performance, and present the results from its initial use in a range of environments in five hospitals, and across a variety of surgical specialties in the emergency and elective settings. The glitch rate can be used to detect the similarities and differences in the volume and distribution of process imperfections among operating sites and specialties. This novel method builds on previous experience and has resulted in a tool which is transferrable between surgical disciplines. We consider that the method has been shown to be sufficiently robust to prove to be of use in the assessment of most intraoperative settings. However, differences in personnel, procedures and equipment in different types of surgery are likely to result in systematic differences in median baseline glitch rates. We therefore suggest that the principal use of the method should be to follow change within a team in response to influences

such as stressors or training, rather than comparisons between the operation types. The collection and analysis of glitches could facilitate the development of targeted systems improvement interventions. We suggest that expression as a glitch rate per hour is appropriate, as it accounts for the varying length of operations, and facilitates the interspecialty/site comparisons. The use of dual observation in the challenging environment of an operating theatre, with multiple demands on observer attention is a deliberate choice and is integral to this method, since we have noted (and confirmed here) that one observer identifies only between 40% and 75% of total glitch events. There were clear differences in the event detection profiles between clinical and HF observers, which might have been expected, but the clinical observers consistently collected significantly more glitches (table 2). This finding is at odds with previous research where the HF observers were found to be more efficient at recording deviations.

We had expected that HF observers might be more aware of some categories of glitch than the clinical observers, but this did not appear to be the case. The clinical observers did not appear to 'overcall' glitches, as the calls were confirmed by consensus discussion, but did appear more sensitive to particular categories of glitch. However, the extensive exposure of our clinical observers to HF theory and practice should be noted, and it cannot be assumed that the clinical observers without this background could perform to this level. To perform this kind of study without observers with demonstrable HF expertise would have made interpretation of our data difficult for others. Therefore, dual observation increased the sensitivity of event detection by up to

60% by incorporating all observations from the two overlapping, non-identical domains of expertise. We suggest that this approach, which maximises sensitivity, is more likely to be generally valuable in operating theatres than one based on high levels of interoperator agreement which sacrifices sensitivity for specificity.

As indicated in the introduction, several groups have independently developed approaches similar to ours,^{6 16 18 19 26} all with differing taxonomies with analogous loci of focus. This 'convergent evolution' has, we believe, been driven by the need to develop a tool which preserves the rich data collection possibilities of direct observation without being impossibly unwieldy for live use in clinical settings. There are strengths and weaknesses in the various existing methods, and a clear opportunity exists for unification among them.

By analysing the content of the glitch, a richer understanding on the recurring problems within the system can be gained. As can be seen from a variety of glitches in the operative process, it would be unlikely that an intervention focusing on only one category of glitch (eg, distractions) would have as significant an effect as one which focused on the wider range of issues that our study identifies in the operating theatre. The methodology allows many layers of analysis, from a basic arithmetic evaluation to a richer contextual analysis. The analysis of the categories of glitches enables the consideration of a system-targeted intervention, with a focus on preventing the creation and propagation of additional works in the operating theatre.

A common criticism of observational research is the bias created as a result of human participants altering their behaviour when aware of being observed, that is, the Hawthorne effect,²⁷ although some doubt the importance of this phenomenon.²⁸ While we cannot exclude the bias of this type, the nature of several glitch categories excludes the possibility of mitigation by altered staff behaviour (eg, the occurrence of a phone call or the dropping of an instrument). Second, due to a number of observations over a prolonged period of time, the observers quickly became well known to the theatre staff and as such became 'part of the furniture', following which the behaviours of the staff did not appear to change when the observers were present. Throughout the large data sample, the same patterns of types and the rates of glitches were repeated, suggesting that Hawthorne effects were not prominent. The evolution of the glitch count method through observation, analysis and consensus discussion in an appropriately skilled team has given it an important degree of construct and face validity, while our data show adequate reliability within the method—notwithstanding the fact that the discordant observations within observer pairs actually strengthen sensitivity as discussed above.

The observed number of events per operation in our study is lower than that observed in the previous studies of this type using other direct observation methods.^{14 16} However, a direct comparison is difficult due to

methodological differences.^{14 16} Although developed as standardised methods, all of these approaches require calibration between observers, and suffer from potential problems in attempting to combine the observations from the teams where this has not occurred. Despite this, it is possible to find areas of close agreement between the studies in the high prevalence of some categories, such as distraction events.^{13 17} The similarities in the methods developed independently by different groups^{16 17 25} suggest that harmonisation and development of a standard methodology may be possible. Clearly this would have potential benefits for research, training and assessment, but would require a substantial cooperative work to achieve congruence and validation.

The use of a glitch rate to normalise for operative duration allowed interesting observations of the possible effects of specialty and hospital environment and culture on glitch rates. It might be expected that the different specialties would have different rates and types of glitches, but in fact, types seemed to show a remarkably consistent pattern and the specialty glitch rates do not significantly vary relative to each other ($p=0.453$). Hospital environment, on the other hand, may be important, as suggested by the 40% difference between sites for vascular surgery. Further work is needed to explore this.

A new finding from this study is the relationship between the accumulation of glitches and the phases of the operation. It appears that the majority of the glitches are clustered around the beginning of the operation, with 50% occurring in the first 30–40% of the operation. This important specialty-spanning finding from a large sample indicates that the highest rate of glitches occurs during one of the busiest parts of an operation, in which multiple activities (positioning, preparation, confirmation of anaesthesia and surgical incision) are occurring in parallel or in quick succession. The implication that safety and reliability might therefore be improved by an ergonomic approach in analysing and reducing the glitches during this phase deserves further study.

CONCLUSION

We propose the glitch methodology as a practical and sensitive methodology for evaluating technical performance during operations, which can be used to gain rich insights into the workings of the operating theatre teams. Our expansive data collection approach has been developed with two independent observers each collecting between 40% and 75% of all glitch occurrences. The majority of glitches occur within the first half of the operation which coincides with a number of safety critical steps. There seems to be a greater difference between hospital sites than surgical specialties in the frequency of glitches. Through analysing the frequency and context around the frequently occurring glitches it is proposed that a suite of targeted interventions could be developed in order to improve the safety and reliability of the operating theatre environment.

Acknowledgements The authors would like to thank Julia Matthews and Laura Bleakley for their assistance in intraoperative data collection for the study.

Contributors The glitch count method was based on earlier work by authors' group, starting with a taxonomy developed by KC, and progressing through a revised system derived by KC, PM and others. LM, KC and PM led the development of the new method. ER, LM, SP, MH, Laura Bleakley and Julia Matthews captured the glitch observations in theatre. ER, LM, SP and MH contributed to reliability studies. All authors contributed to discussions about the development of the method. GC led the statistical analysis. LM wrote the first and subsequent drafts and PM wrote the final draft of the article. All authors contributed to the writing process, including redrafting and editing of the text, and agreed with the final version.

Funding This article presents an independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research programme (Reference Number RP-PG-0108-10020). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None.

Ethics approval All theatre staff included in the study were consented for participation under ethical clearance from the Oxford A Ethics Committee [REC:09/H0604/39].

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Access to full datasets with all anonymised data will be made available to other researchers after the programme has been terminated and the principal manuscripts from it have been published. The authors will maintain this dataset for 3 years and then assign it to the Oxford Research Depository.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Baines RJ, Langelaan M, de Bruijne MC, *et al.* Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. *BMJ Qual Saf* 2013;22:290–8.
- Hu YY, Arriaga AF, Roth EM, *et al.* Protecting patients from an unsafe system: the etiology and recovery of intraoperative deviations in care. *Ann Surg* 2012;256:203–10.
- Roese NJ, Vohs KD, Bias H. Perspectives on psychological. *Science* 2012;7:411–26.
- Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care* 2009;18:104–8.
- Flin R, Patey R. Non-technical skills for anaesthetists: developing and applying ANTS. *Best Pract Res Clin Anaesthesiol* 2011;25:215–27.
- Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care* 2004;13(Suppl 1):i33–40.
- Carthey J. The role of structured observational research in health care. *Qual Saf Health Care* 2003;12(Suppl 2):ii13–16.
- de Leval MR, Carthey J, Wright DJ, *et al.* Human factors and cardiac surgery: a multicenter study. *J Thorac Cardiovasc Surg* 2000;119(4 Pt 1):661–72.
- Lingard L, Espin S, Whyte S, *et al.* Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care* 2004;13:330–4.
- Reason JT. *Managing the risks of organizational accidents*. Aldershot: Ashgate, 1997.
- Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;320:781–5.
- de Vries EN, Prins HA, Crolla RM, *et al.* Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med* 2010;363:1928–37.
- Catchpole KR, Giddings AE, de Leval MR, *et al.* Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;49:567–88.
- Catchpole KR, Giddings AE, Wilkinson M, *et al.* Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;142:102–10.
- Parker WH. Understanding errors during laparoscopic surgery. *Obstet Gynecol Clin North Am* 2010;37:437–49.
- Wiegmann DA, ElBardissi AW, Dearani JA, *et al.* Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery* 2007;142:658–65.
- Healey AN, Sevdalis N, Vincent CA. Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics* 2006;49:589–604.
- Schraagen JM, Schouten T, Smit M, *et al.* Assessing and improving teamwork in cardiac surgery. *Qual Saf Health Care* 2010;19:e29.
- Sevdalis N, Forrest D, Undre S, *et al.* Annoyances, disruptions, and interruptions in surgery: the Disruptions in Surgery Index (DiSI). *World J Surg* 2008;32:1643–50.
- Morgan LJ, Pickering SP, Catchpole KC, *et al.* Observing and categorising process deviations in orthopaedic surgery. *Proc Hum Factors Ergon Soc Annu Meeting* 2011;55:685–9.
- Robertson E, Hadi M, Morgan L, *et al.* The development of process maps in the training of surgical and human factors observers in orthopaedic surgery. *Int J Surg* 2011;9:1.
- Moss S. *Human Factors Reference Group, Interim Report*. Department of Health, 2012:56.
- Great Britain. *Dept. of H., an organisation with a memory: report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer*. The Stationery Office, 2000.
- Kohn LT, Corrigan J, Donaldson MS. *To err is human: building a safer health system*. Washington, DC: National Academy Press, 2000:287, xxi.
- Catchpole K. Observing failures in successful orthopaedic surgery. In: Flin RH, Mitchell L. eds. *Safer surgery: analysing behaviour in the operating theatre*. Ashgate: Aldershot, 2009:456, xxvi.
- Catchpole K, Wiegmann D. Understanding safety and performance in the cardiac operating room: from 'sharp end' to 'blunt end'. *BMJ Qual Saf* 2012;21:807–9.
- Adair J. Hawthorne effect. In: Lewis-Beck M, Bryman A, Liao TF. eds. *Encyclopedia of social science research methods*. Thousand Oaks, CA: Sage Publications, Inc, 2004:453.
- Levitt SD, List JA. Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *Am Econ J Appl Econ* 2011;3:224–38.