Capturing intraoperative process deviations: The glitch method

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003519
Article Type:	Research
Date Submitted by the Author:	04-Jul-2013

Complete List of Authors:	Morgan, Lauren; University of Oxford, Nuffield Department of Surgical Sciences Robertson, Eleanor; University of Oxford, Nuffield Department of Surgical Sciences Hadi, Mohammed; Warwick Medical School, Warwick Orthopaedics Catchpole, Ken; Cedars-Sinai Medical Center, Department of Surgery Pickering, Sharon; Warwick Medical School, Warwick Orthopaedics New, Steve; University of Oxford, Saïd Business School Collins, Gary; University of Oxford, Centre for Statistics in Medicine McCulloch, Peter; University of Oxford, Nuffield Department of Surgical Sciences

Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research, Surgery
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SURGERY, patient safety, quality improvement, process of care

TITLE PAGE

Capturing intraoperative process deviations: The glitch method

Lauren Morgan, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom OX3 9DU
lauren.morgan@nds.ox.ac.uk
Tel: 01865 740870; Fax: 01865 768876

Eleanor Robertson, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

Mohammed Hadi, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Ken Catchpole, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles 90048, United States of America

Sharon Pickering, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Steve New, Saïd Business School, University of Oxford, Oxford, United Kingdom
OX1 1HP

Gary Collins, Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom
OX3 7LD

Peter McCulloch, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

KEYWORDS

Quality in healthcare
Surgery
Patient safety
Quality improvement
Process of care

Word count: 3695

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 NHS 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 speciality.

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 sensitivity over a single observer. 429 operations were observed and 5742 glitches recorded (mean 14 per operation, range 0-83). Speciality-specific glitch rates varied from 6.9 to 8.3 per hour of operating (ns). The distribution of glitch categories was strikingly similar across specialities, with distractions the commonest type in all cases. The difference in glitch rate between speciality teams operating at different sites was larger than that between specialities (range 6.3 – 10.5 per hour, p<0.001). 40% 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 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 operative process more strongly than speciality.

ARTICLE SUMMARY

Article focus

- to study interventions to improve safety and reliability of theatre processes, accurate methods are needed to identify process glitches
- a direct observation method using pairs of observers with different skill sets should provide high sensitivity in identifying and categorising process glitches
- such a developed and validated method can be used to study the factors which affect the frequency and nature of imperfections in theatre process

Key messages

- the glitch method appears reliable in use and highlights the areas in which theatre process is vulnerable to deviations and errors
- glitch rates per hour of operating are significantly different in different hospitals, but the glitch rate and distribution of glitch types is remarkably consistent across different surgical specialities
- glitches are concentrated in the first quarter of operations, and are least likely to occur in the last quarter

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 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

INTRODUCTION

The delivery of safe surgical care is technically challenging. It requires considered patient selection, pre-operative assessment, and the coordination of technical expertise and resource. These requirements result in the creation of a challenging work environment which puts both healthcare workers and the enveloping healthcare system under considerable stress. The resultant iatrogenic patient harm has been recently estimated at 6.2%, with 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 observations[7] 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]. Both 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 'intra-operative 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 specialities, techniques and settings, do not make assumptions about causality and do not attempt to estimate impact on outcome. Since other published methods do not fulfil these criteria we present a method developed based on the approach used in previous studies[4, 18]. We previously reported the initial development of this method including reliability assessment and taxonomy development[19]. In this study we report a large scale evaluation of its use to characterise the relationship of glitches to the context, in terms of speciality, site and operative duration, their temporal pattern during operations and the relative frequency of the different categories.

METHODS

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 entire operations from the time the patient entered the operating theatre to the time they left, by pairs of observers comprising of one clinical and one human factors (HF) researcher. 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[20]. 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 de-duplicated and summed to provide a total glitch count for an operation. We recorded the detail of the glitch (e.g. ‘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 one month in training and orientation to the data collection methods before any real-time data was collected. Alongside the collection of glitches, the observers also assessed the teams’ non-technical skills, WHO checklist adherence and recorded operative duration as part of a larger programme of work.

Patients were informed of the possibility of observations taking place and given 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. The study was approved by Oxford A Ethics Committee (REC:09/H0604/39).

Development and validation of methodology

The observational methods were based upon those developed by others [14-16]: the development of the glitch categories and testing of inter-observer reliability have been described previously[19]. Briefly, a sample set of 94 glitches were collected during the initial training phase, grouped in common themes, and assigned titles and definitions (Table 1). The reliability of the categorisation process was assessed using Cohen's kappa . Agreement was good between the four observers (0.70, 95% CI: 0.66 to 0.75). 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 was reviewed jointly by the observers (LM, MH, SP, ER). Glitches noted by both observers were categorised by consensus where there was a difference, and an overall glitch score was assigned comprising the sum of all unique glitches seen (i.e. those unique to observer A + those unique to observer B + those in common). 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 re-categorised during this consensus process.

Observer background, training and context

The clinical observers had a clinical qualification (Surgical Trainees and Operating Department Practitioners) and more than one years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HF. The HF observers were orientated to the operating theatre environment and learnt technical aspects of the operative process through observation and through coaching from clinical observers. The clinical observers were introduced to HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from HF observers whilst in theatre.

Large scale evaluation of glitch method

Intra-operative observations were made across five UK NHS sites and four specialities: elective orthopaedics, trauma orthopaedics, vascular surgery 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). The surgical specialities were chosen to provide homogeneity in operation type across sites, and also for their differing complexity and operative durations. We sought to capture the glitches that occur in the operating theatres across different specialities and sites to allow comparison of volume and profile of glitches across both sites and specialities.

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

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; 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 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 & Safety	Any observed physical risk to personnel	Mask violations; food/drink in theatre
Planning & Preparation	Instances that may otherwise been avoided with appropriate prior planning and preparation	Insufficient equipment resources; staffing levels; training
Patient Related	Issues relating to the physiological status of the patient	Difficulty in extracting previous implants
Process Deviation	Incomplete or re-ordered 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	De-sterilising of

	issues	equipment/scrubbed staff on environment

Data analysis

Differences in mean glitch rates per operation between the sites and specialities 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 theatre (Table 2). Of the 5742 glitches, 64% were observed by the HF observers, 76% by the clinical observers (p=<0.001). The clinical observers consistently noted more glitches per operation than the HF (Table 1) but the difference varied markedly between glitch categories. Clinical observers noted a much larger proportion of Environment, Training, Health & Safety and Patient Related glitches, whilst there was minimal difference between the observers for Absence, Slips and Equipment Maintenance.

Table 2 Difference in observed glitches between observer specialities

Glitch category	Total observed n (% of total)	Observed by both HF and clinical n (% of category)	HF observed n (% of category)	Clinical observed n (% of category)	Difference % (95% CI)	P
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 & 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 & 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

Observed glitches by site and speciality

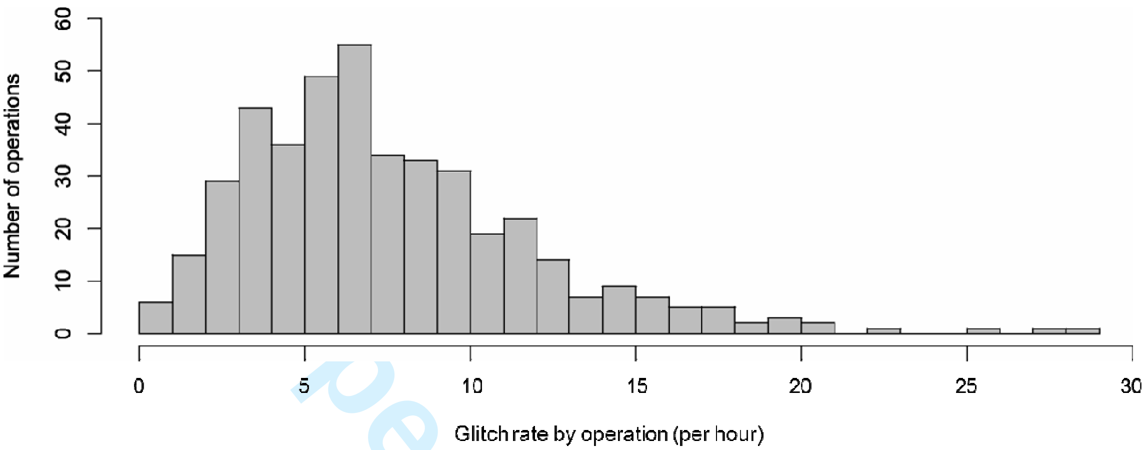
The number of procedures observed in different sites and specialities is shown in Table 3.

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 to 13:32)	63 (2:02, 0:27 to 4:45)	96 (1:47, 0:22 to 3:58)	72 (1:46, 0:30 to 3:55)	24 (3:11, 0:43 to 7:28)
Elective orthopaedics	130 (1:49, 0:27 to 4:25)	63 (2:02, 0:27 to 4:45)	54 (1:49, 0:27 to 3:41)	51 (1:54, 1:01 to 3:55)	0
Trauma orthopaedics	0	0	42 (1:44, 0:22 to 3:58)	0	0
Plastic surgery	45 (2:17, 0:30 to 13:32)	0	0	0	0
Vascular surgery	0	0	0	21 (1:25, 0:30 to 2.24)	

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 1-83. The number of glitches per operation by speciality ranged from 1-63 in elective orthopaedic surgery, 1-35 in trauma orthopaedics, 2-49 in elective vascular surgery and 1-83 in elective plastic surgery. Due to the range of operation duration, both within and between specialities, a glitch rate is required to facilitate 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.

Figure 1 Distribution of glitch rate by operation

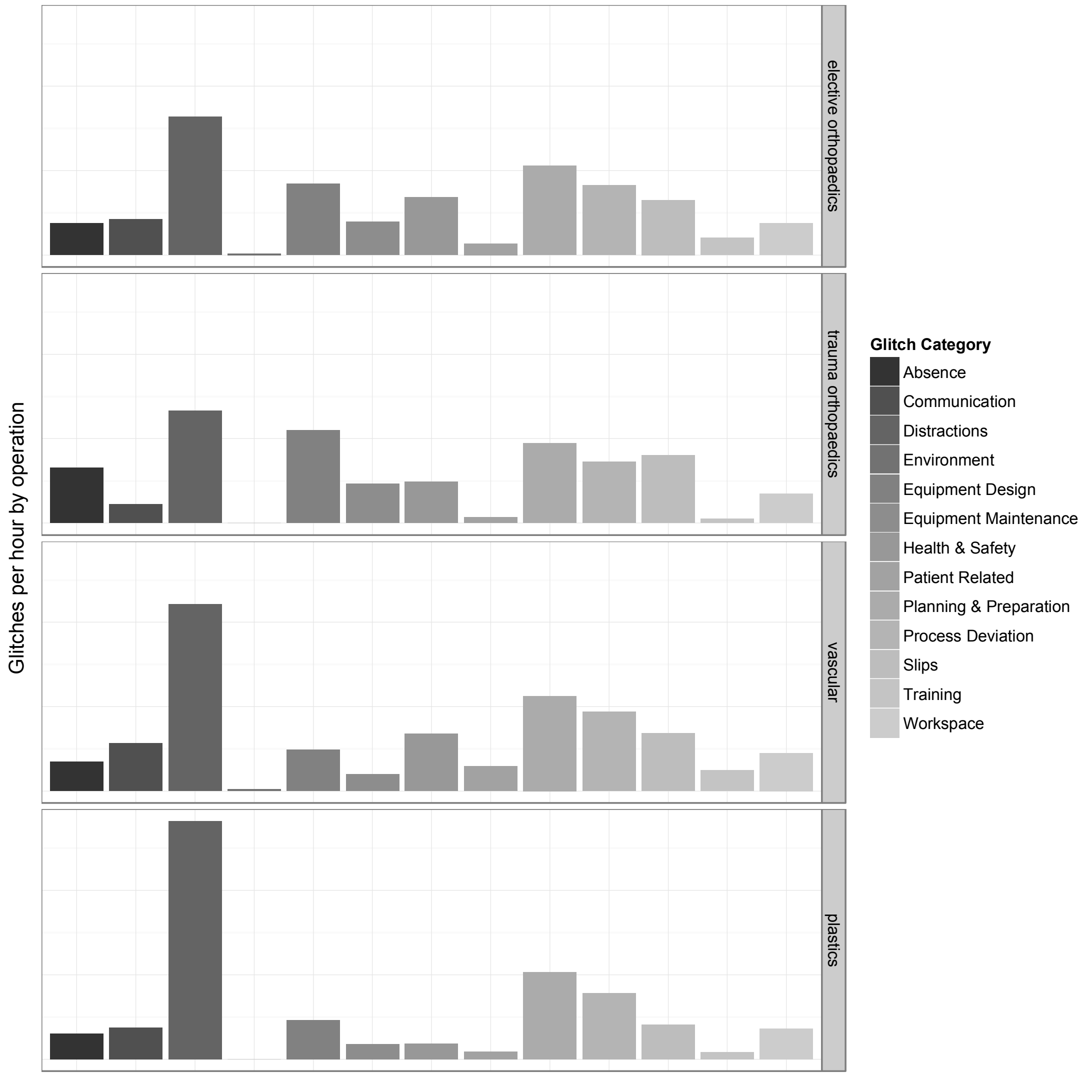


Although there is a strong clustering around the mean, the data are skewed, with a number of operations with high glitch rates at >20 per hour. The mean glitch rate for orthopaedics is 7.6 (range 0.4 to 28.4), trauma orthopaedics is 6.9 (range 1.3 to 15.3), vascular is 8.3 (range 1.5 to 20.6), and plastics is 7.1 (range 0.7 to 28). There was no statistically significant difference in average glitch rate across the four specialities (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 specialities (Figure 2).

Figure 2 Mean glitch rate by operation for each speciality



The profile of the glitch categories between the surgical specialities is strikingly similar. It can be seen that the most common glitches across all specialities 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 the greatest in plastics which relates anecdotally to the discursive and fluid nature of the teams involved. There is a higher rate of maintenance and absence glitches for trauma orthopaedics and a low level of slips in plastics in comparison to the other specialities.

Relationship between glitch rate and hospital site

Elective orthopaedic and vascular surgical procedures were observed in multiple sites, providing an opportunity for inter-site glitch rate comparison amongst teams performing the same types of surgery.

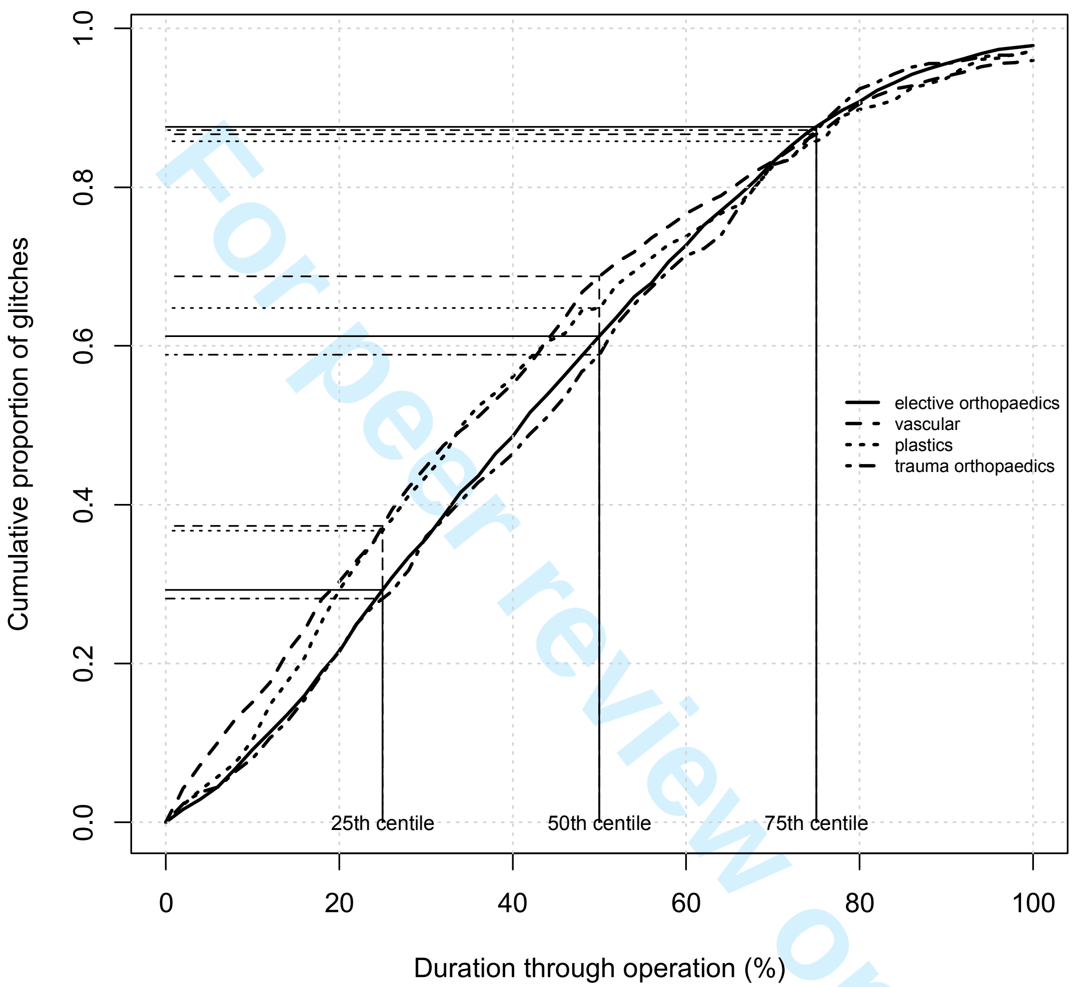
In elective orthopaedics, the mean operating duration varies by 14 minutes between the sites, with glitch rate varying between 6 and 8 glitches per hour. There was a statistically significant differences in mean glitch rates per operation between the four sites (p<0.001), with significant differences between individual sites A (mean of 8.1 glitches per hour) and B (mean of 6.0 glitches per hour) (difference = 2.1; 95%CI 0.9 to 3.3; p=0.001), sites D (mean of 8.7 glitches per hour) and B (mean of 6.0 glitches per hour) (difference = 2.7; 95%CI 1.4 to 4; p<0.001), and sites C (mean of 7.3 glitches per hour) and B (mean of 6.0 glitches per hour) (difference=1.2; 95% CI 0.01 to 2.6; p=0.047).

In vascular surgery, the difference in glitch rate between Site E (mean of 6.3 glitches per hour) and Site D (mean of 10.5 glitches per 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 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 glitches and duration of operation, suggesting a reduction in glitch occurrence in the last 20% of operation duration (Figure 3).

Figure 3 Distribution of glitch occurrence across operative duration



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 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[21-23]. The prospective collection of information about process imperfections or deviations enables healthcare researchers to analyse intra-operative events so that the system and the operative technique can be evaluated for quality and risk. The observed events include different sub-classes such as distractions, process deviations, equipment design problems, and slips (Table 1). These glitches are not necessarily associated with immediate consequences or due to any failure in 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 patient safety, as it may preserve the team's capacity for dealing with unexpected events[18]. Although not tested in this study, it has previously been suggested that the accumulation of 'minor' events predisposes to a 'major' event associated with potential for serious patient harm[14].

Measuring the prevalence of glitches provides a quantitative practical insight into the effects of system malfunctions on the process and on the healthcare 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 we have called glitches. There may be circumstances where there is blending of both system and human factors. 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 pre-list 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 specialities in emergency and elective settings. The glitch rate can be used to detect similarities and differences in the volume and distribution of process imperfections amongst operating sites and specialities. 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 inter-speciality/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 40-75% of total glitch events. There were clear differences in event detection profiles between clinical and HF observers, which might have been expected, but clinical observers consistently collected significantly more glitches (Table 2).

We had expected that HF observers might be more aware of some categories of glitch than clinical observers, but this did not appear to be the case. 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 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 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 inter-operator agreement which sacrifices sensitivity for specificity.

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 the variety of glitches in the operative process, it would be unlikely that an intervention focussing on only one category of glitch (e.g. distractions) would have as significant an effect as one which focussed on the wider range of issues that our study identifies in the operating theatre. The methodology allows many layers of analysis, from basic arithmetic evaluation to richer contextual analysis. Analysis of the categories of glitches enables the consideration of a system-targeted intervention, with the focus on preventing the creation and propagation of additional work in the operating theatre.

A common criticism of observational research is the bias created as a result of human subjects altering their behaviour when aware of being observed, i.e. the Hawthorne effect[24], although some doubt the importance of this phenomenon[25]. Whilst we cannot exclude bias of this type, the nature of several glitch categories excludes the possibility of mitigation by altered staff behaviour (e.g. the occurrence of a phone call or the dropping of an instrument). Secondly, due to the 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 staff behaviours did not appear to change when the observers were present. Throughout the large data sample, the same patterns of types and 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, whilst 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 observed in previous studies of this type using other direct observation methods[14, 16]. However, 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 observations from 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 methods developed independently by different groups[16-18] 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 substantial co-operative 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 both speciality and hospital environment and culture on glitch rates. It might be expected that different specialities would have different rates and types of glitches, but in fact types seemed to show a remarkably consistent pattern and speciality 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 speciality-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 to analysing and reducing 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 operating theatre teams. Our expansive data collection approach has been developed with two independent observers each collecting between 40-75% 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 specialities in the frequency of glitches. Through analysing the frequency and context around 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.

ETHICAL 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].

SOURCES OF FUNDING

This paper presents 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.

DATA SHARING

Access to full datasets with all anonymised data will be made available to other researchers after the Programme has been terminated and the principal papers from it have been published. We will maintain this dataset for 3 years and then assign it to the Oxford Research Depository.

CONTRIBUTORSHIP STATEMENT

The glitch count method was based on earlier work by our 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 and MH captured the glitch observations in theatre and 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 the final draft of the paper. All authors contributed to the writing process, including redrafting and editing of the text, and agreed the final version.

REFERENCES

1. 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**(4):290-298.
2. 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**(2):203-10.
3. Roese NJ, Vohs KD. Hindsight Bias. *Perspect Psychol Sci* 2012;**7**(5):411-426.
4. 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**(2):104-8.
5. Flin R, Patey R. Non-technical skills for anaesthetists: developing and applying ANTS. *Best Pract Res Clin Anaesthesiol* 2011;**25**(2):215-27.
6. Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care* 2004; **13**(Suppl 1): P.i33-40.
7. Carthey J. The role of structured observational research in health care. *Qual Saf Health Care* 2003;**12**(Suppl 2): P.ii13-6.
8. 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.
9. 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**(5): 330-4.
10. Reason JT. Managing the risks of organizational accidents. Aldershot, UK: Ashgate 1997.
11. Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;**320**(7237):781-5.

12. de Vries EN, Prins HA, Crolla RMPH, et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med* 2010;**363**(20):1928-1937.

13. Catchpole KR, Giddings AE, de Leval MR, et al. Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;**49**(5-6):567-88.

14. Catchpole, KR, Giddings AE, Wilkinson M, et al. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;**142**(1):102-10.

15. Parker WH. Understanding errors during laparoscopic surgery. *Obstet Gynecol Clin North Am* 2010;**37**(3):437-49.

16. Wiegmann DA, ElBardissi AW, Dearani JA, et al. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery* 2007;**142**(5):658-65.

17. Healey AN, Sevdalis N, Vincent CA. Measuring intra-operative interference from distraction and interruption observedin the operating theatre. *Ergonomics* 2006; **49**(5-6):589-604.

18. Catchpole K. Observing failures in successful orthopaedic surgery. In: Flin RH, Mitchell L, eds. Safer surgery : analysing behaviour in the operating theatre. Aldershot, UK: Ashgate 2009: p. xxvi

19. Morgan LJ, Pickering SP, Catchpole KC, et al. Observing and categorising process deviations in orthopaedic surgery. *Proc Hum Fact Ergon Soc Annu Meet* 2011;**55**(1):685-689.

20. 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**(7): 549.

21. Moss S. Human Factors Reference Group, Interim Report. Department of Health, UK 2012. p. 56.

22. Department of Health. An organisation with a memory : report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. Department of Health, UK: The Stationery Office 2000.

23. Kohn LT, Corrigan J, Donaldson MS. To err is human : building a safer health system. Washington, DC: National Academy Press 2000. p. xxi.

24. Adair JG. Hawthorne Effect. In: Lewis-Beck MS, Bryman A, Liao TF, eds. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: Sage Publications, Inc: 453.

25. Levitt SD, Lisa 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**(1):224-238.

Capturing intraoperative process deviations using a direct observational approach: The glitch method

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003519.R1
Article Type:	Research
Date Submitted by the Author:	09-Sep-2013

Complete List of Authors:	Morgan, Lauren; University of Oxford, Nuffield Department of Surgical Sciences Robertson, Eleanor; University of Oxford, Nuffield Department of Surgical Sciences Hadi, Mohammed; Warwick Medical School, Warwick Orthopaedics Catchpole, Ken; Cedars-Sinai Medical Center, Department of Surgery Pickering, Sharon; Warwick Medical School, Warwick Orthopaedics New, Steve; University of Oxford, Saïd Business School Collins, Gary; University of Oxford, Centre for Statistics in Medicine McCulloch, Peter; University of Oxford, Nuffield Department of Surgical Sciences

Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research, Surgery
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SURGERY, patient safety, quality improvement, process of care

1

TITLE PAGE

Capturing intraoperative process deviations using a direct observational approach: The glitch method

Lauren Morgan, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom OX3 9DU
lauren.morgan@nds.ox.ac.uk
Tel: 01865 740870; Fax: 01865 768876

Eleanor Robertson, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

Mohammed Hadi, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Ken Catchpole, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles 90048, United States of America

Sharon Pickering, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Steve New, Saïd Business School, University of Oxford, Oxford, United Kingdom
OX1 1HP

Gary Collins, Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom
OX3 7LD

Peter McCulloch, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

KEYWORDS

Quality in healthcare
Surgery
Patient safety
Quality improvement
Process of care

Word count: 4533~~3695~~

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 NHS 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 speciality.

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 sensitivity over a single observer. 429 operations were observed and 5742 glitches recorded (mean 14 per operation, range 0-83). Speciality-specific glitch rates varied from 6.9 to 8.3 per hour of operating (ns). The distribution of glitch categories was strikingly similar across specialities, with distractions the commonest type in all cases. The difference in glitch rate between speciality teams operating at different sites was larger than that between specialities (range 6.3 – 10.5 per hour, $p<0.001$). 40% 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 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 operative process more strongly than speciality.

4

ARTICLE SUMMARY

Article focus

- to study interventions to improve safety and reliability of theatre processes, accurate methods are needed to identify process glitches
- a direct observation method using pairs of observers with different skill sets should provide high sensitivity in identifying and categorising process glitches
- such a developed and validated method can be used to study the factors which affect the frequency and nature of imperfections in theatre process

Key messages

- the glitch method appears reliable in use and highlights the areas in which theatre process is vulnerable to deviations and errors
- glitch rates per hour of operating are significantly different in different hospitals, but the glitch rate and distribution of glitch types is remarkably consistent across different surgical specialities
- glitches are concentrated in the first quarter of operations, and are least likely to occur in the last quarter

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 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

INTRODUCTION

The delivery of safe surgical care is technically challenging. It requires considered patient selection, pre-operative assessment, and the coordination of technical expertise and resource. These requirements result in the creation of a challenging work environment which puts both healthcare workers and the enveloping healthcare system under considerable stress. The resultant iatrogenic patient harm has been recently estimated at 6.2%, with 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 observations[7] 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]. Both 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 'intra-operative 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 specialities, techniques and settings, do not make assumptions about causality and do not attempt to estimate 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 tackle the difficulties of learning, adapting and validating a new system, we decided to develop our own based

6

on our previous studies Since other published methods do not fulfil these criteria we present a method developed based on the approach used in previous studies[4, 18]. 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 speciality, 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 upon those developed by others [14-16]: the development of the glitch categories and testing of inter-observer reliability have been described previously[19]. The method involves two observers, one with a Human Factors (HF) 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 pre-designed procedure template as a guide. ABriefly, a sample set of 94 glitches from 10 elective orthopaedic operations were collected during the initial three month training phase, grouped in common themes, and assigned titles and definitions (Table 1). The reliability of the observers categorisation process was assessed using Cohen's kappa . Agreementand 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 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 was reviewed jointly by the observers (LM, MH, SP, ER). Glitches noted by both observers were categorised by consensus where there was a difference, and an overall glitch score was assigned comprising the sum of all unique glitches seen (i.e. those unique to observer A + those unique to observer B + those in common). 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 re-categorised during this consensus process.

Observer background, training and context

The clinical observers had a clinical qualification (Surgical Trainees and Operating Department Practitioners) and more than one years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HF. The HF observers were orientated to the

Field Code Changed

operating theatre environment and learnt technical aspects of the operative process through observation and through coaching from clinical observers. The clinical observers were introduced to HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from HF observers whilst 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 entire operations from the time the patient entered the operating theatre to the time they left, by pairs of six observers comprising of one clinical and one human factors (HF) researcher. Four of the six observers (MH, SP, ER and LM) were involved in the creation of the method, the remainder (LB and JM) were introduced to the categorisation at a later date. Any process disruption which occurred in the pre- or post- operative 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 intra-operative 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 de-duplicated and summed to provide a total glitch count for an operation. We recorded the detail of the glitch (e.g. '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 one month in training and orientation to the data collection methods before any real-time data was collected. Alongside the collection of glitches, the observers also assessed the teams' non-technical skills, WHO surgical safety checklist adherence and recorded 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 given 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. The study was approved by Oxford A Ethics Committee (REC:09/H0604/39).

Development and validation of methodology

The observational methods were based upon those developed by others [14-16]: the development of the glitch categories and testing of inter-observer reliability have been described previously[19]. Briefly, a sample set of 94 glitches were collected during the initial training phase, grouped in common themes, and assigned titles and definitions (Table 1). The reliability of the categorisation process was assessed using Cohen's kappa . Agreement was good between the four observers (0.70, 95% CI: 0.66 to 0.75). 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 was reviewed jointly by the observers (LM, MH, SP, ER). Glitches noted by both observers were categorised by consensus where there was a difference, and an overall glitch score was assigned comprising the sum of all unique glitches seen (i.e. those unique to observer A + those unique to observer B + those in common). 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 re-categorised during this consensus process.

Observer background, training and context

The clinical observers had a clinical qualification (Surgical Trainees and Operating Department Practitioners) and more than one years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HF. The HF observers were orientated to the operating theatre environment and learnt technical aspects of the operative process through observation and through coaching from clinical observers. The clinical observers were introduced to HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from HF observers whilst in theatre.

Large scale evaluation of glitch method

Intra-operative observations were made across five UK NHS sites and four specialities: elective orthopaedics, trauma orthopaedics, vascular surgery, 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 inter-hospital comparator speciality due to the homogeneity of operation technique and length. 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 speciality. On these occasions, other surgical specialities were recruited to the study, which in turn enabled the evaluation of the glitch and other intra-operative observational techniques across surgical specialities. The surgical specialities were chosen to provide homogeneity in operation type across sites, and also for their differing complexity and operative durations. We sought to capture the glitches that occur in the operating theatres across different specialities and sites toThis permitted the allow comparison of volume and profile of glitches across both sites and specialities.

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

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; 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 & Safety	Any observed physical risk to personnel	Mask violations; food/drink in theatre
Planning & Preparation	Instances that may otherwise been avoided with appropriate prior planning and preparation	Insufficient equipment resources; staffing levels; training

Patient Related	Issues relating to the physiological status of the patient	Difficulty in extracting previous implants; unexpected anatomically-related surgical difficulty; anaphylaxis
Process Deviation	Incomplete or re-ordered 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	De-sterilising of equipment/scrubbed staff on environment

Data analysis

Differences in mean glitch rates per operation between the sites and specialities 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 theatre (Table 2). Of the 5742 glitches, 64% were observed by the HF observers, 76% by the clinical observers (p=<0.001). The clinical observers consistently noted more glitches per operation than the HF (Table 1) but the difference varied markedly between glitch categories. Clinical observers noted a much larger proportion of Environment, Training, Health & Safety and Patient Related glitches, whilst there was minimal difference between the observers for Absence, Slips and Equipment Maintenance.

Table 2 Difference in observed glitches between observer specialities

Glitch category	Total observed n (% of total)	Observed by both HF and clinical n (% of	HF observed n (% of category)	Clinical observed n (% of category)	Difference % (95% CI)	P

		category)				
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 & 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 & 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

Observed glitches by site and speciality

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

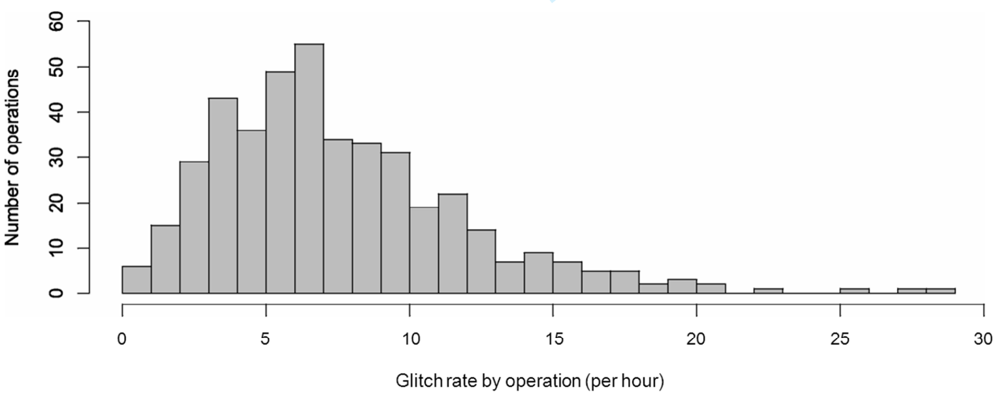
Table 3 Sample characteristics of observations

	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 to 13:32)	63 (2:02, 0:27 to 4:45)	96 (1:47, 0:22 to 3:58)	72 (1:46, 0:30 to 3:55)	24 (3:11, 0:43 to 7:28)
Elective orthopaedics	130 (1:49, 0:27 to 4:25)	63 (2:02, 0:27 to 4:45)	54 (1:49, 0:27 to 3:41)	51 (1:54, 1:01 to 3:55)	0
Trauma orthopaedics	0	0	42 (1:44, 0:22 to 3:58)	0	0
Plastic surgery	45 (2:17, 0:30 to 13:32)	0	0	0	0
Vascular surgery	0	0	0	21 (1:25,	24 (3:11,

				0:30 to 2.24)	0:43 to 7:28)

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~~1~~-83. The number of glitches per operation by speciality ranged from 1-63 in elective orthopaedic surgery, 1-35 in trauma orthopaedics, 2-49 in elective vascular surgery and 1-83 in elective plastic surgery. Due to the range of operation duration, both within and between specialities, a glitch rate is required to facilitate 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.

Figure 1 Distribution of glitch rate by operation

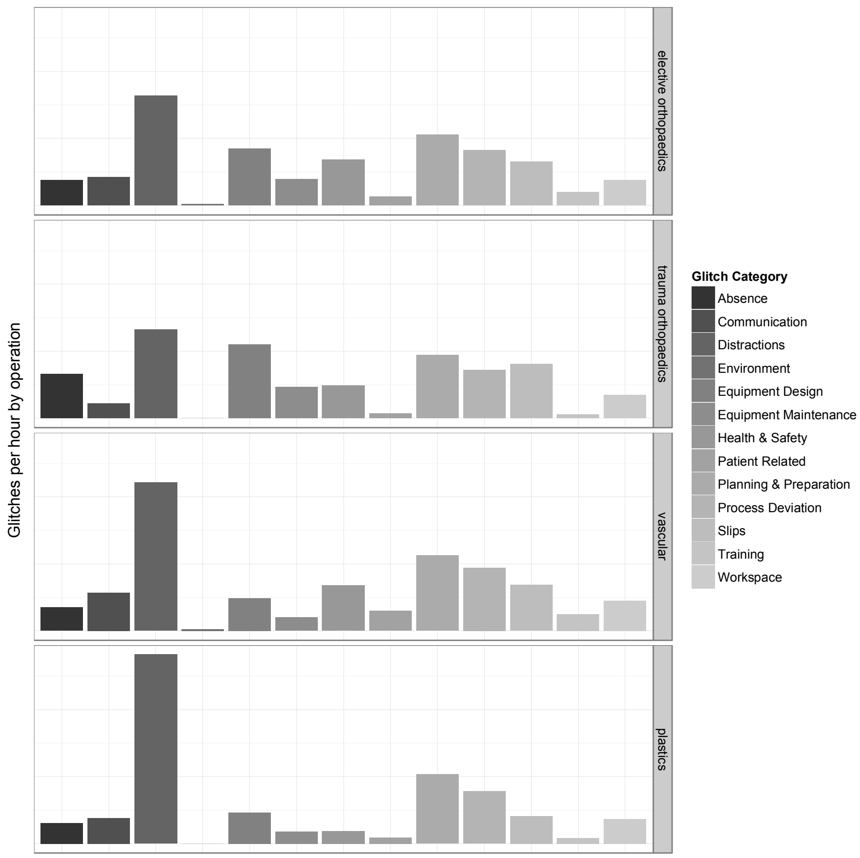


Although there is a strong clustering around the mean, the data are skewed, with a number of operations with high glitch rates at >20 per hour. The mean glitch rate for orthopaedics is 7.6 (range 0.4 to 28.4), trauma orthopaedics is 6.9 (range 1.3 to 15.3), vascular is 8.3 (range 1.5 to 20.6), and plastics is 7.1 (range 0.7 to 28). There was no statistically significant difference in average glitch rate across the four specialities (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 specialities (Figure 2).

Figure 2 Mean glitch rate by operation for each speciality



The profile of the glitch categories between the surgical specialities is strikingly similar. It can be seen that the most common glitches across all specialities 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 the greatest in plastic surgerys which relates anecdotally to the discursive and fluid nature of the teams involved. There is a higher rate of maintenance and absence glitches for trauma orthopaedics and a low level of slips in plastics in comparison to the other specialities.

Relationship between glitch rate and hospital site

E~~E~~lective orthopaedic and vascular surgical procedures were observed in multiple sites (four and two sites respectively), providing an opportunity for inter-site glitch rate comparison amongst teams performing the same types of surgery.

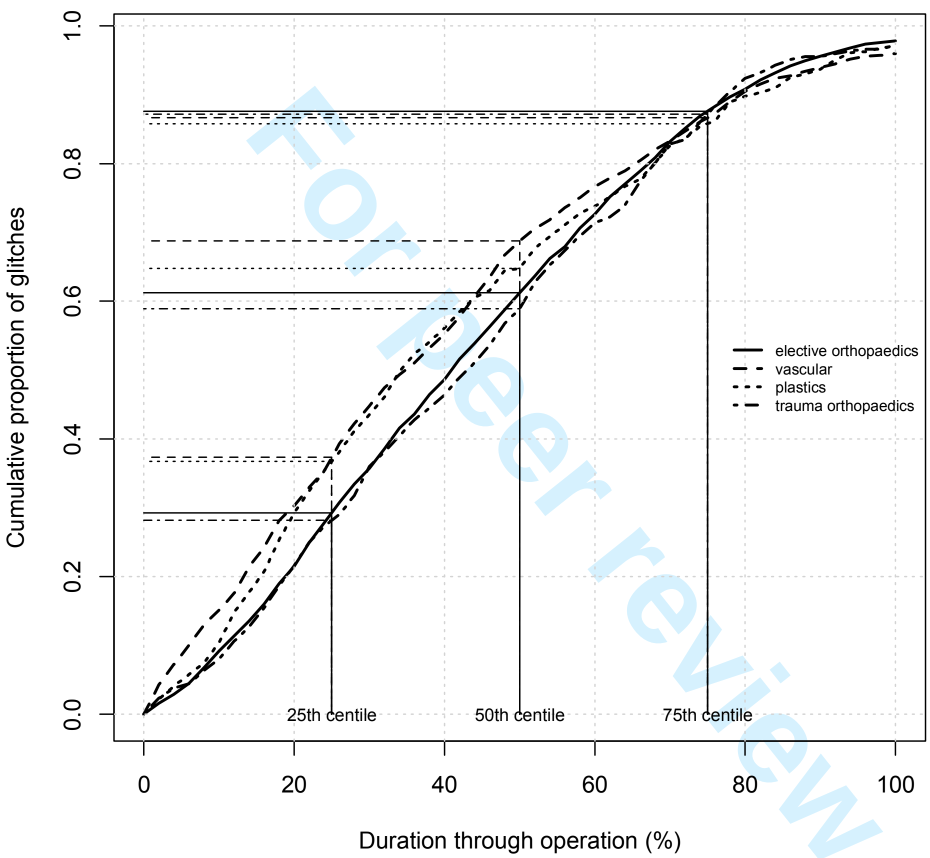
In elective orthopaedics, the mean operating duration varies by 14 minutes between the sites, with glitch rate varying between 6 and 8 glitches per hour. There was a statistically significant ~~hetereogeneity~~heterogeneity ~~differences~~ in mean glitch rates per operation between the four sites (p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

In vascular surgery, the difference in glitch rate between Site E (mean of 6.3 glitches per hour) and Site D (mean of 10.5 glitches per 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 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 glitches and duration of operation, suggesting a reduction in glitch occurrence in the last 20% of operation duration (Figure 3).

Figure 3 Distribution of glitch occurrence across operative duration



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 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 intra-operative events so that the system and the operative technique can be evaluated for quality and risk. The observed events include different sub-classes such as distractions, process deviations, equipment design problems, and slips (Table 1). These glitches are not necessarily associated with immediate consequences or due to any failure in 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 patient safety, as it may preserve the team's capacity for dealing with 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 potential for serious patient harm[14]. We did not seek to prove causality of glitches as we felt that it would be unwise to attempt to link a glitch with what could be multiple up-stream 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 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 we have called glitches. There may be circumstances where there is blending of both system and human factors. 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 pre-list 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 specialities in emergency and elective settings. The glitch rate can be used to detect similarities and differences in the volume and distribution of process imperfections amongst operating sites and specialities. This novel method builds upon 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 intra-operative 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 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 inter-speciality/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 40-75% of total glitch events. There were clear differences in event detection profiles between clinical and HF observers, which might have been expected, but clinical observers consistently collected significantly more glitches (Table 2). This finding is at odds with previous research where 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 clinical observers, but this did not appear to be the case. 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 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 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 inter-operator 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]. 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 amongst 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 the variety of glitches in the operative process, it would be unlikely that an intervention focussing on only one category of glitch (e.g. distractions) would have as significant an effect as one which focussed on the wider range of issues that our study identifies in the operating theatre. The methodology allows many layers of analysis, from basic arithmetic evaluation to richer contextual analysis. Analysis of the categories of glitches enables the consideration of a system-targeted intervention, with the focus on preventing the creation and propagation of additional work in the operating theatre.

A common criticism of observational research is the bias created as a result of human subjects altering their behaviour when aware of being observed, i.e. the Hawthorne effect[27], although some doubt the importance of this phenomenon[28]. Whilst we cannot exclude bias of this type, the nature of several glitch categories excludes the possibility of mitigation by altered staff behaviour (e.g. the occurrence of a phone call or the dropping of an instrument). Secondly, due to the 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 staff behaviours did not appear to change when the observers were present. Throughout the large data sample, the same patterns of types and 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, whilst 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 observed in previous studies of this type using other direct observation methods[14, 16]. However, 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 observations from 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 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 substantial co-operative 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 both speciality and hospital environment and culture on glitch rates. It might be expected that different specialities would have different rates and types of glitches, but in fact types seemed to show a remarkably consistent pattern and speciality 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 speciality-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 to analysing and reducing 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 operating theatre teams. Our expansive data collection approach has been developed with two independent observers each collecting between 40-75% 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 specialities in the frequency of glitches. Through analysing the frequency and context around 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.

ETHICAL 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].

AKNOWLEDGEMENTS

We would like to thank Julia Matthews and Laura Bleakley for their assistance in intra-operative data collection for the study.

SOURCES OF FUNDING

This paper presents 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.

DATA SHARING

Access to full datasets with all anonymised data will be made available to other researchers after the Programme has been terminated and the principal papers from it have been published. We will maintain this dataset for 3 years and then assign it to the Oxford Research Depository.

REFERENCES

1. Baines, R.J., et al., *Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study.* BMJ Quality & Safety, 2013. **22**(4): p. 9.
2. Hu, Y.Y., et al., *Protecting patients from an unsafe system: the etiology and recovery of intraoperative deviations in care.* Ann Surg, 2012. **256**(2): p. 203-10.
3. Roese, N.J. and K.D. Vohs, *Hindsight Bias.* Perspectives on Psychological Science, 2012. **7**(5): p. 411-426.
4. Mishra, A., K. Catchpole, and P. McCulloch, *The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre.* Qual Saf Health Care, 2009. **18**(2): p. 104-8.
5. Flin, R. and R. Patey, *Non-technical skills for anaesthetists: developing and applying ANTS.* Best Pract Res Clin Anaesthesiol, 2011. **25**(2): p. 215-27.
6. Healey, A.N., S. Undre, and C.A. Vincent, *Developing observational measures of performance in surgical teams.* Qual Saf Health Care, 2004. **13 Suppl 1**: p. i33-40.
7. Carthey, J., *The role of structured observational research in health care.* Qual Saf Health Care, 2003. **12 Suppl 2**: p. ii13-6.
8. de Leval, M.R., et al., *Human factors and cardiac surgery: a multicenter study.* J Thorac Cardiovasc Surg, 2000. **119**(4 Pt 1): p. 661-72.
9. Lingard, L., et al., *Communication failures in the operating room: an observational classification of recurrent types and effects.* Qual Saf Health Care, 2004. **13**(5): p. 330-4.
10. Reason, J.T., *Managing the risks of organizational accidents*. 1997, Aldershot: Ashgate.
11. Helmreich, R.L., *On error management: lessons from aviation.* BMJ, 2000. **320**(7237): p. 781-5.
12. de Vries, E.N., et al., *Effect of a Comprehensive Surgical Safety System on Patient Outcomes.* New England Journal of Medicine, 2010. **363**(20): p. 1928-1937.

13. Catchpole, K.R., et al., *Identification of systems failures in successful paediatric cardiac surgery*. Ergonomics, 2006. **49**(5-6): p. 567-88.
14. Catchpole, K.R., et al., *Improving patient safety by identifying latent failures in successful operations.* Surgery, 2007. **142**(1): p. 102-10.
15. Parker, W.H., *Understanding errors during laparoscopic surgery.* Obstet Gynecol Clin North Am, 2010. **37**(3): p. 437-49.
16. Wiegmann, D.A., et al., *Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation.* Surgery, 2007. **142**(5): p. 658-65.
17. Healey, A.N., N. Sevdalis, and C.A. Vincent, *Measuring intra-operative interference from distraction and interruption observedin the operating theatre.* Ergonomics, 2006. **49**(5-6): p. 589-604.
18. Schraagen, J.M., et al., *Assessing and improving teamwork in cardiac surgery.* Quality and Safety in Health Care, 2010. **19**(6): p. e29.
19. Sevdalis, N., et al., *Annoyances, disruptions, and interruptions in surgery: the Disruptions in Surgery Index (DiSI).* World J Surg, 2008. **32**(8): p. 1643-50.
20. Morgan, L., et al., *Observing and Categorising Process Deviations in Orthopaedic Surgery.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2011. **55**(1): p. 685-689.
21. Robertson, E., Hadi, M., Morgan, L., Pickering, S., Catchpole, K., Griffin, D., McCulloch, P., *The development of process maps in the training of surgical and human factors observers in orthopaedic surgery.* Int J Surg, 2011. **9**(7): p. 1.
22. Moss, S., *Human Factors Reference Group, Interim Report*, 2012, Department of Health. p. 56.
23. 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*. 2000: The Stationery Office.
24. Kohn, L.T., J. Corrigan, and M.S. Donaldson, *To err is human : building a safer health system*. 2000, Washington, D.C.: National Academy Press. xxi, 287 p.
25. Catchpole, K., *Observing Failures in Successful Orthopaedic Surgery*, in *Safer surgery : analysing behaviour in the operating theatre*, R.H. Flin and L. Mitchell, Editors. 2009, Ashgate: Aldershot. p. xxvi, 456 p.
26. Catchpole, K. and D. Wiegmann, *Understanding safety and performance in the cardiac operating room: from 'sharp end'to 'blunt end'.* BMJ Quality & Safety, 2012. **21**(10): p. 807-809.
27. Sage Publications, I., *Hawthorne Effect. The SAGE Encyclopedia of Social Science Research Methods. Sage Publications, Inc*. Thousand Oaks, USA: Sage Publications, Inc.
28. Levitt, S.D. and J.A. List, *Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments.* American Economic Journal-Applied Economics, 2011. **3**(1): p. 224-238.

Capturing intraoperative process deviations using a direct observational approach: The glitch method

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003519.R2
Article Type:	Research
Date Submitted by the Author:	16-Oct-2013
Complete List of Authors:	Morgan, Lauren; University of Oxford, Nuffield Department of Surgical Sciences Robertson, Eleanor; University of Oxford, Nuffield Department of Surgical Sciences Hadi, Mohammed; Warwick Medical School, Warwick Orthopaedics Catchpole, Ken; Cedars-Sinai Medical Center, Department of Surgery Pickering, Sharon; Warwick Medical School, Warwick Orthopaedics New, Steve; University of Oxford, Saïd Business School Collins, Gary; University of Oxford, Centre for Statistics in Medicine McCulloch, Peter; University of Oxford, Nuffield Department of Surgical Sciences
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research, Surgery
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SURGERY, patient safety, quality improvement, process of care

TITLE PAGE

Capturing intraoperative process deviations using a direct observational approach: The glitch method

Lauren Morgan, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom OX3 9DU
lauren.morgan@nds.ox.ac.uk
Tel: 01865 740870; Fax: 01865 768876

Eleanor Robertson, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

Mohammed Hadi, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Ken Catchpole, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles 90048, United States of America

Sharon Pickering, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Steve New, Saïd Business School, University of Oxford, Oxford, United Kingdom
OX1 1HP

Gary Collins, Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom
OX3 7LD

Peter McCulloch, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

KEYWORDS

Quality in healthcare
Surgery
Patient safety
Quality improvement
Process of care

Word count: 4533

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 NHS 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 speciality.

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 sensitivity over a single observer. 429 operations were observed and 5742 glitches recorded (mean 14 per operation, range 0-83). Speciality-specific glitch rates varied from 6.9 to 8.3 per hour of operating (ns). The distribution of glitch categories was strikingly similar across specialities, with distractions the commonest type in all cases. The difference in glitch rate between speciality teams operating at different sites was larger than that between specialities (range 6.3 – 10.5 per hour, $p<0.001$). 40% 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 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 operative process more strongly than speciality.

ARTICLE SUMMARY

Article focus

- to study interventions to improve safety and reliability of theatre processes, accurate methods are needed to identify process glitches
- a direct observation method using pairs of observers with different skill sets should provide high sensitivity in identifying and categorising process glitches
- such a developed and validated method can be used to study the factors which affect the frequency and nature of imperfections in theatre process

Key messages

- the glitch method appears reliable in use and highlights the areas in which theatre process is vulnerable to deviations and errors
- glitch rates per hour of operating are significantly different in different hospitals, but the glitch rate and distribution of glitch types is remarkably consistent across different surgical specialities
- glitches are concentrated in the first quarter of operations, and are least likely to occur in the last quarter

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 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

INTRODUCTION

The delivery of safe surgical care is technically challenging. It requires considered patient selection, pre-operative assessment, and the coordination of technical expertise and resource. These requirements result in the creation of a challenging work environment which puts both healthcare workers and the enveloping healthcare system under considerable stress. The resultant iatrogenic patient harm has been recently estimated at 6.2%, with 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 observations[7] 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]. Both 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 'intra-operative 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 specialities, techniques and settings, do not make assumptions about causality and do not attempt to estimate 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 tackle 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 speciality, 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 upon those developed by others [14-16]: the development of the glitch categories and testing of inter-observer reliability have been described previously[19]. The method involves two observers, one with a Human Factors (HF) 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 pre-designed procedure template as a guide. A sample set of 94 glitches from 10 elective orthopaedic operations were collected during the initial three 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 kappa 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 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 was reviewed jointly by the observers (LM, MH, SP, ER). Glitches noted by both observers were categorised by consensus where there was a difference, and an overall glitch score was assigned comprising the sum of all unique glitches seen (i.e. those unique to observer A + those unique to observer B + those in common). 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 re-categorised during this consensus process.

Observer background, training and context

The clinical observers had a clinical qualification (Surgical Trainees and Operating Department Practitioners) and more than one years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HF. The HF observers were orientated to the operating theatre environment and learnt technical aspects of the operative process through observation and through coaching from clinical observers. The clinical observers were introduced to HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from HF observers whilst 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 activity (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 human factors (HF) researcher. Four of the six observers (MH, SP, ER and LM) were involved in the creation of the method, the remainder (LB and JM) were introduced to the categorisation at a later date. Any process disruption which occurred in the pre- 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 intra-operative 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 de-duplicated and summed to provide a total glitch count for an operation. We recorded the detail of the glitch (e.g. '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 one month in training and orientation to the data collection methods before any real-time data was collected. Alongside the collection of glitches, the observers also assessed the teams' non-technical skills, WHO surgical safety checklist adherence and recorded 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 given 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. The study was approved by Oxford A Ethics Committee (REC:09/H0604/39).

Large scale evaluation of glitch method

Intra-operative observations were made across five UK NHS sites and four specialities: 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 inter-hospital comparator speciality due to the homogeneity of operation technique and length. 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 speciality. On these occasions, other surgical specialities were recruited to the study, which in turn enabled the evaluation of the glitch and other intra-operative observational techniques across surgical specialities. This permitted the comparison of volume and profile of glitches across both sites and specialities.

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

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; 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 & Safety	Any observed physical risk to	Mask violations; food/drink in

	personnel	theatre
Planning & Preparation	Instances that may otherwise been avoided with appropriate prior planning and preparation	Insufficient equipment resources; staffing levels; training
Patient Related	Issues relating to the physiological status of the patient	Difficulty in extracting previous implants; unexpected anatomically-related surgical difficulty; anaphylaxis
Process Deviation	Incomplete or re-ordered 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	De-sterilising of equipment/scrubbed staff on environment

Data analysis

Differences in mean glitch rates per operation between the sites and specialities 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 theatre (Table 2). Of the 5742 glitches, 64% were observed by the HF observers, 76% by the clinical observers (p=<0.001). The clinical observers consistently noted more glitches per operation than the HF (Table 1) but the difference varied markedly between glitch categories. Clinical observers noted a much larger proportion of Environment, Training, Health & Safety and Patient Related glitches, whilst there was minimal difference between the observers for Absence, Slips and Equipment Maintenance.

Table 2 Difference in observed glitches between observer specialities

Glitch category	Total observed n (% of total)	Observed by both HF and clinical n (% of category)	HF observed n (% of category)	Clinical observed n (% of category)	Difference % (95% CI)	P
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 & 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 & 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

Observed glitches by site and speciality

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

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 to 13:32)	63 (2:02, 0:27 to 4:45)	96 (1:47, 0:22 to 3:58)	72 (1:46, 0:30 to 3:55)	24 (3:11, 0:43 to 7:28)
Elective orthopaedics	130 (1:49, 0:27 to 4:25)	63 (2:02, 0:27 to 4:45)	54 (1:49, 0:27 to 3:41)	51 (1:54, 1:01 to 3:55)	0
Trauma orthopaedics	0	0	42 (1:44, 0:22 to 3:58)	0	0
Plastic surgery	45	0	0	0	0

	(2:17, 0:30 to 13:32)				
Vascular surgery	0	0	0	21 (1:25, 0:30 to 2.24)	24 (3:11, 0:43 to 7:28)

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 speciality ranged from 1-63 in elective orthopaedic surgery, 1-35 in trauma orthopaedics, 2-49 in elective vascular surgery and 1-83 in elective plastic surgery. Due to the range of operation duration, both within and between specialities, a glitch rate is required to facilitate 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.

Figure 1 Distribution of glitch rate by operation

Although there is a strong clustering around the mean, the data are skewed, with a number of operations with high glitch rates at >20 per hour. The mean glitch rate for orthopaedics is 7.6 (range 0.4 to 28.4), trauma orthopaedics is 6.9 (range 1.3 to 15.3), vascular is 8.3 (range 1.5 to 20.6), and plastics is 7.1 (range 0.7 to 28). There was no statistically significant difference in average glitch rate across the four specialities (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 specialities (Figure 2).

Figure 2 Mean glitch rate by operation for each speciality

The profile of the glitch categories between the surgical specialities is strikingly similar. It can be seen that the most common glitches across all specialities 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 the greatest 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 glitches for trauma orthopaedics and a low level of slips in plastics in comparison to the other specialities.

Relationship between glitch rate and hospital site

Elective orthopaedic and vascular surgical procedures were observed in multiple sites (four and two sites respectively), providing an opportunity for inter-site glitch rate comparison amongst teams performing the same types of surgery.

In elective orthopaedics, the mean operating duration varies by 14 minutes between the sites, with glitch rate varying between 6 and 8 glitches per 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 per hour) and B (mean of 6.0 glitches per hour) (difference = 2.1; 95%CI 0.9 to 3.3; $p=0.001$), sites D (mean of 8.7 glitches per hour) and B (mean of 6.0 glitches per hour) (difference = 2.7; 95%CI 1.4 to 4; $p<0.001$), and sites C (mean of 7.3 glitches per hour) and B (mean of 6.0 glitches per 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 glitch rate between Site E (mean of 6.3 glitches per hour) and Site D (mean of 10.5 glitches per 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 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 glitches and duration of operation, suggesting a reduction in glitch occurrence in the last 20% of operation duration (Figure 3).

Figure 3 Distribution of glitch occurrence across operative duration

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 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 intra-operative events so that the system and the operative technique can be evaluated for quality and risk. The observed events include different sub-classes such as distractions, process deviations, equipment design problems, and slips (Table 1). These glitches are not necessarily associated with immediate consequences or due to any failure in 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 patient safety, as it may preserve the team's capacity for dealing with 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 potential for serious patient harm[14]. We did not seek to prove causality of glitches as we felt that it would be unwise to attempt to link a glitch with what could be multiple up-stream 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 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 we have called glitches. There may be circumstances where there is blending of both system and human factors. 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 pre-list 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 specialities in emergency and elective settings. The glitch rate can be used to detect similarities and differences in the volume and distribution of process imperfections amongst operating sites and specialities. This novel method builds upon 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 intra-operative 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 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 inter-speciality/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 40-75% of total glitch events. There were clear differences in event detection profiles between clinical and HF observers, which might have been expected, but clinical observers consistently collected significantly more glitches (Table 2). This finding is at odds with previous research where 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 clinical observers, but this did not appear to be the case. 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 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 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 inter-operator 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 amongst 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 the variety of glitches in the operative process, it would be unlikely that an intervention focussing on only one category of glitch (e.g. distractions) would have as significant an effect as one which focussed on the wider range of issues that our study identifies in the operating theatre. The methodology allows many layers of analysis, from basic arithmetic evaluation to richer contextual analysis. Analysis of the categories of glitches enables the consideration of a system-targeted intervention, with the focus on preventing the creation and propagation of additional work in the operating theatre.

A common criticism of observational research is the bias created as a result of human subjects altering their behaviour when aware of being observed, i.e. the Hawthorne effect[27], although some doubt the importance of this phenomenon[28]. Whilst we cannot exclude bias of this type, the nature of several glitch categories excludes the possibility of mitigation by altered staff behaviour (e.g. the occurrence of a phone call or the dropping of an instrument). Secondly, due to the 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 staff behaviours did not appear to change when the observers were present. Throughout the large data sample, the same patterns of types and 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, whilst 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 observed in previous studies of this type using other direct observation methods[14, 16]. However, 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 observations from 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 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 substantial co-operative 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 both speciality and hospital environment and culture on glitch rates. It might be expected that different specialities would have different rates and types of glitches, but in fact types seemed to show a remarkably consistent pattern and speciality 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 speciality-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 to analysing and reducing 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 operating theatre teams. Our expansive data collection approach has been developed with two independent observers each collecting between 40-75% 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 specialities in the frequency of glitches. Through analysing the frequency and context around 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.

Figure Legends:

Figure 1. Distribution of glitch rate by operation

Figure 2. Mean glitch rate by operation for each speciality

Figure 3. Distribution of glitch occurrence across operative duration

ETHICAL 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]

CONTRIBUTORSHIP STATEMENT

The glitch count method was based on earlier work by our 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, LB and JM 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 the final draft of the paper. All authors contributed to the writing process, including redrafting and editing of the text, and agreed the final version.

AKNOWLEDGEMENTS

We would like to thank Julia Matthews and Laura Bleakley for their assistance in intra-operative data collection for the study.

SOURCES OF FUNDING

This paper presents 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.

DATA SHARING

Access to full datasets with all anonymised data will be made available to other researchers after the Programme has been terminated and the principal papers from it have been published. We will maintain this dataset for 3 years and then assign it to the Oxford Research Depository.

REFERENCES

1. Baines, R.J., et al., *Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study.* BMJ Quality & Safety, 2013. **22**(4): p. 9.
2. Hu, Y.Y., et al., *Protecting patients from an unsafe system: the etiology and recovery of intraoperative deviations in care.* Ann Surg, 2012. **256**(2): p. 203-10.
3. Roese, N.J. and K.D. Vohs, *Hindsight Bias.* Perspectives on Psychological Science, 2012. **7**(5): p. 411-426.
4. Mishra, A., K. Catchpole, and P. McCulloch, *The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre.* Qual Saf Health Care, 2009. **18**(2): p. 104-8.
5. Flin, R. and R. Patey, *Non-technical skills for anaesthetists: developing and applying ANTS.* Best Pract Res Clin Anaesthesiol, 2011. **25**(2): p. 215-27.
6. Healey, A.N., S. Undre, and C.A. Vincent, *Developing observational measures of performance in surgical teams.* Qual Saf Health Care, 2004. **13 Suppl 1**: p. i33-40.
7. Carthey, J., *The role of structured observational research in health care.* Qual Saf Health Care, 2003. **12 Suppl 2**: p. ii13-6.
8. de Leval, M.R., et al., *Human factors and cardiac surgery: a multicenter study.* J Thorac Cardiovasc Surg, 2000. **119**(4 Pt 1): p. 661-72.
9. Lingard, L., et al., *Communication failures in the operating room: an observational classification of recurrent types and effects.* Qual Saf Health Care, 2004. **13**(5): p. 330-4.
10. Reason, J.T., *Managing the risks of organizational accidents*. 1997, Aldershot: Ashgate.
11. Helmreich, R.L., *On error management: lessons from aviation.* BMJ, 2000. **320**(7237): p. 781-5.
12. de Vries, E.N., et al., *Effect of a Comprehensive Surgical Safety System on Patient Outcomes.* New England Journal of Medicine, 2010. **363**(20): p. 1928-1937.
13. Catchpole, K.R., et al., *Identification of systems failures in successful paediatric cardiac surgery.* Ergonomics, 2006. **49**(5-6): p. 567-88.
14. Catchpole, K.R., et al., *Improving patient safety by identifying latent failures in successful operations.* Surgery, 2007. **142**(1): p. 102-10.
15. Parker, W.H., *Understanding errors during laparoscopic surgery.* Obstet Gynecol Clin North Am, 2010. **37**(3): p. 437-49.
16. Wiegmann, D.A., et al., *Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation.* Surgery, 2007. **142**(5): p. 658-65.
17. Healey, A.N., N. Sevdalis, and C.A. Vincent, *Measuring intra-operative interference from distraction and interruption observedin the operating theatre.* Ergonomics, 2006. **49**(5-6): p. 589-604.
18. Schraagen, J.M., et al., *Assessing and improving teamwork in cardiac surgery.* Quality and Safety in Health Care, 2010. **19**(6): p. e29.
19. Sevdalis, N., et al., *Annoyances, disruptions, and interruptions in surgery: the Disruptions in Surgery Index (DiSI).* World J Surg, 2008. **32**(8): p. 1643-50.
20. Morgan, L., et al., *Observing and Categorising Process Deviations in Orthopaedic Surgery.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2011. **55**(1): p. 685-689.

21. 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**(7): p. 1.
22. Moss, S., *Human Factors Reference Group, Interim Report*, 2012, Department of Health. p. 56.
23. 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*. 2000: The Stationery Office.
24. Kohn, L.T., J. Corrigan, and M.S. Donaldson, *To err is human : building a safer health system*. 2000, Washington, D.C.: National Academy Press. xxi, 287 p.
25. Catchpole, K., *Observing Failures in Successful Orthopaedic Surgery*, in *Safer surgery : analysing behaviour in the operating theatre*, R.H. Flin and L. Mitchell, Editors. 2009, Ashgate: Aldershot. p. xxvi, 456 p.
26. Catchpole, K. and D. Wiegmann, *Understanding safety and performance in the cardiac operating room: from 'sharp end'to 'blunt end'.* BMJ Quality & Safety, 2012. **21**(10): p. 807-809.
27. Sage Publications, I., *Hawthorne Effect. The SAGE Encyclopedia of Social Science Research Methods. Sage Publications, Inc*. Thousand Oaks, USA: Sage Publications, Inc.
28. Levitt, S.D. and J.A. List, *Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments.* American Economic Journal-Applied Economics, 2011. **3**(1): p. 224-238.

1

TITLE PAGE

Capturing intraoperative process deviations using a direct observational approach: The glitch method

Lauren Morgan, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom OX3 9DU
lauren.morgan@nds.ox.ac.uk
Tel: 01865 740870; Fax: 01865 768876

Eleanor Robertson, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

Mohammed Hadi, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Ken Catchpole, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles 90048, United States of America

Sharon Pickering, Warwick Orthopaedics, Warwick Medical School, Coventry, United Kingdom
CV2 2DX

Steve New, Saïd Business School, University of Oxford, Oxford, United Kingdom
OX1 1HP

Gary Collins, Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom
OX3 7LD

Peter McCulloch, Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom
OX3 9DU

KEYWORDS

Quality in healthcare
Surgery
Patient safety
Quality improvement
Process of care

Word count: 4533

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 NHS 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 speciality.

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 sensitivity over a single observer. 429 operations were observed and 5742 glitches recorded (mean 14 per operation, range 0-83). Speciality-specific glitch rates varied from 6.9 to 8.3 per hour of operating (ns). The distribution of glitch categories was strikingly similar across specialities, with distractions the commonest type in all cases. The difference in glitch rate between speciality teams operating at different sites was larger than that between specialities (range 6.3 – 10.5 per hour, $p<0.001$). 40% 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 changes following interventions to improve process safety, and appears reliable and sensitive. Glitches occur more

3

frequently in the early stages of an operation. Hospital environment, culture and work systems may influence operative process more strongly than speciality.

4

ARTICLE SUMMARY

Article focus

- to study interventions to improve safety and reliability of theatre processes, accurate methods are needed to identify process glitches
- a direct observation method using pairs of observers with different skill sets should provide high sensitivity in identifying and categorising process glitches
- such a developed and validated method can be used to study the factors which affect the frequency and nature of imperfections in theatre process

Key messages

- the glitch method appears reliable in use and highlights the areas in which theatre process is vulnerable to deviations and errors
- glitch rates per hour of operating are significantly different in different hospitals, but the glitch rate and distribution of glitch types is remarkably consistent across different surgical specialities
- glitches are concentrated in the first quarter of operations, and are least likely to occur in the last quarter

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 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

INTRODUCTION

The delivery of safe surgical care is technically challenging. It requires considered patient selection, pre-operative assessment, and the coordination of technical expertise and resource. These requirements result in the creation of a challenging work environment which puts both healthcare workers and the enveloping healthcare system under considerable stress. The resultant iatrogenic patient harm has been recently estimated at 6.2%, with 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 observations[7] 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]. Both 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 'intra-operative 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 specialities, techniques and settings, do not make assumptions about causality and do not attempt to estimate 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 tackle the difficulties of learning, adapting and validating a new system, we decided to develop our own based

6

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 speciality, 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 upon those developed by others [14-16]: the development of the glitch categories and testing of inter-observer reliability have been described previously[19]. The method involves two observers, one with a Human Factors (HF) 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 pre-designed procedure template as a guide. A sample set of 94 glitches from 10 elective orthopaedic operations were collected during the initial three 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 kappa 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 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 was reviewed jointly by the observers (LM, MH, SP, ER). Glitches noted by both observers were categorised by consensus where there was a difference, and an overall glitch score was assigned comprising the sum of all unique glitches seen (i.e. those unique to observer A + those unique to observer B + those in common). 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 re-categorised during this consensus process.

Observer background, training and context

The clinical observers had a clinical qualification (Surgical Trainees and Operating Department Practitioners) and more than one years' operating theatre experience. The HF observers had an undergraduate and/or postgraduate qualification in HF. The HF observers were orientated to the operating theatre environment and learnt technical aspects of the operative process through observation and through coaching from clinical observers. The clinical observers were introduced to

HF system principles and observational methodology through classroom-based teaching and introduction to the HF literature, and through coaching from HF observers whilst 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 activity (surgical, nursing and anaesthetic) in the operating theatre~~entire operations~~ from the time the patient entered ~~the operating theatre~~ to the time they left, by pairs of six observers comprising of one clinical and one human factors (HF) researcher. Four of the six observers (MH, SP, ER and LM) were involved in the creation of the method, the remainder (LB and JM) were introduced to the categorisation at a later date. Any process disruption which occurred in the pre- or post- ~~operative~~ 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 intra-operative 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 de-duplicated and summed to provide a total glitch count for an operation. We recorded the detail of the glitch (e.g. '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 one month in training and orientation to the data collection methods before any real-time data was collected. Alongside the collection of glitches, the observers also assessed the teams' non-technical skills, WHO surgical safety checklist adherence and recorded 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.

Comment [LM1]: Response to reviewer 1

Patients were informed of the possibility of observations taking place and given 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. The study was approved by Oxford A Ethics Committee (REC:09/H0604/39).

Large scale evaluation of glitch method

8

Intra-operative observations were made across five UK NHS sites and four specialities: 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 inter-hospital comparator speciality due to the homogeneity of operation technique and length. 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 speciality. On these occasions, other surgical specialities were recruited to the study, which in turn enabled the evaluation of the glitch and other intra-operative observational techniques across surgical specialities. This permitted the comparison of volume and profile of glitches across both sites and specialities.

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

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; 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 & Safety	Any observed physical risk to personnel	Mask violations; food/drink in theatre

Planning & Preparation	Instances that may otherwise been avoided with appropriate prior planning and preparation	Insufficient equipment resources; staffing levels; training
Patient Related	Issues relating to the physiological status of the patient	Difficulty in extracting previous implants; unexpected anatomically-related surgical difficulty; anaphylaxis
Process Deviation	Incomplete or re-ordered 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	De-sterilising of equipment/scrubbed staff on environment

Data analysis

Differences in mean glitch rates per operation between the sites and specialities 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 theatre (Table 2). Of the 5742 glitches, 64% were observed by the HF observers, 76% by the clinical observers (p=<0.001). The clinical observers consistently noted more glitches per operation than the HF (Table 1) but the difference varied markedly between glitch categories. Clinical observers noted a much larger proportion of Environment, Training, Health & Safety and Patient Related glitches, whilst there was minimal difference between the observers for Absence, Slips and Equipment Maintenance.

Table 2 Difference in observed glitches between observer specialities

Glitch category	Total observed	Observed by both HF	HF observed	Clinical observed	Difference %	P

10

	n (% of total)	and clinical n (% of category)	n (% of category)	n (% of category)	(95% CI)	
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 & 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 & 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

Observed glitches by site and speciality

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

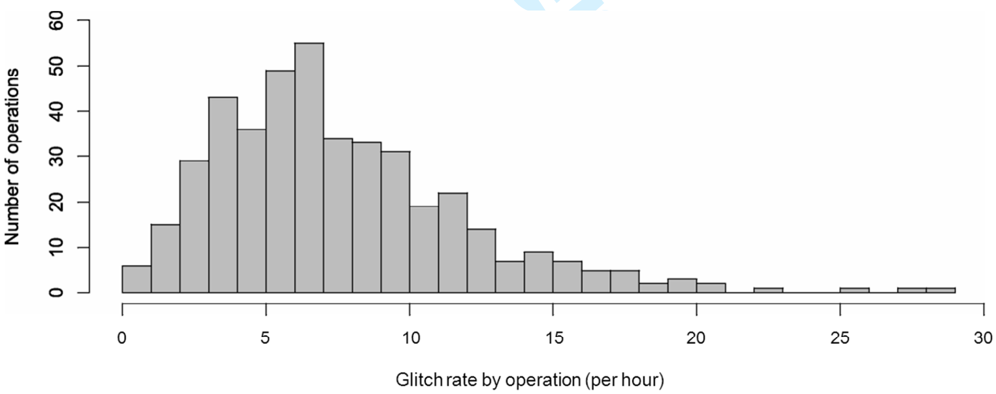
Table 3 Sample characteristics of observations

	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 to 13:32)	63 (2:02, 0:27 to 4:45)	96 (1:47, 0:22 to 3:58)	72 (1:46, 0:30 to 3:55)	24 (3:11, 0:43 to 7:28)
Elective orthopaedics	130 (1:49, 0:27 to 4:25)	63 (2:02, 0:27 to 4:45)	54 (1:49, 0:27 to 3:41)	51 (1:54, 1:01 to 3:55)	0
Trauma orthopaedics	0	0	42 (1:44, 0:22 to 3:58)	0	0
Plastic surgery	45 (2:17, 0:30 to 13:32)	0	0	0	0
Vascular surgery	0	0	0	21 (1:25,	24 (3:11,

				0:30 to 2.24)	0:43 to 7:28)

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 speciality ranged from 1-63 in elective orthopaedic surgery, 1-35 in trauma orthopaedics, 2-49 in elective vascular surgery and 1-83 in elective plastic surgery. Due to the range of operation duration, both within and between specialities, a glitch rate is required to facilitate 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.

Figure 1 Distribution of glitch rate by operation



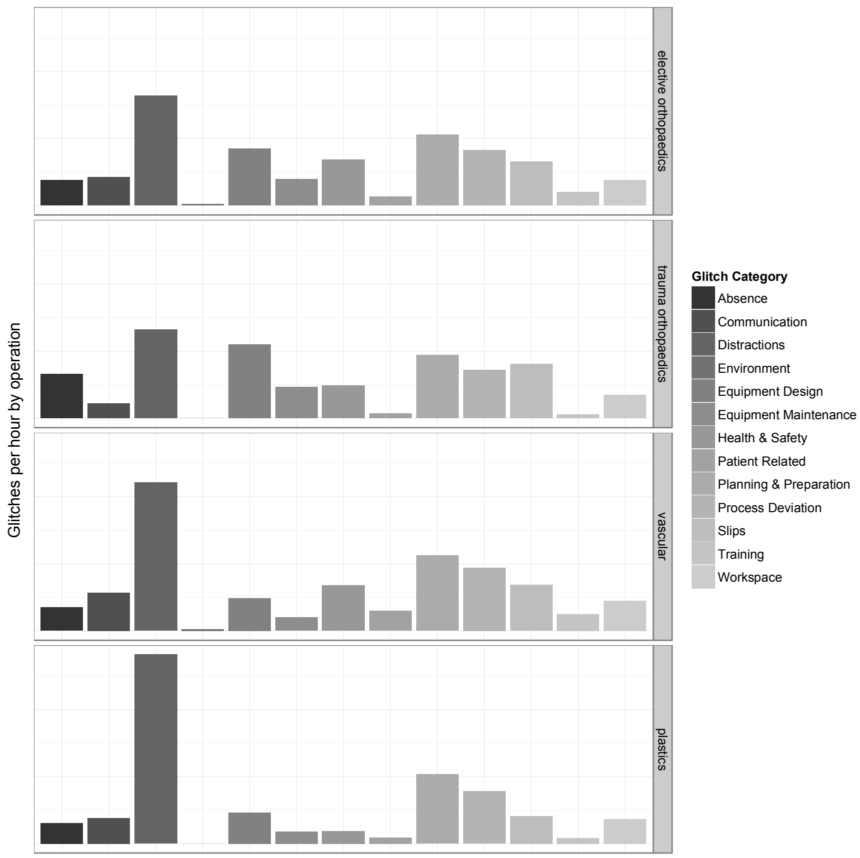
Although there is a strong clustering around the mean, the data are skewed, with a number of operations with high glitch rates at >20 per hour. The mean glitch rate for orthopaedics is 7.6 (range 0.4 to 28.4), trauma orthopaedics is 6.9 (range 1.3 to 15.3), vascular is 8.3 (range 1.5 to 20.6), and plastics is 7.1 (range 0.7 to 28). There was no statistically significant difference in average glitch rate across the four specialities (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 specialities (Figure 2).

12

Figure 2 Mean glitch rate by operation for each speciality



The profile of the glitch categories between the surgical specialities is strikingly similar. It can be seen that the most common glitches across all specialities 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 the greatest 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 glitches for trauma orthopaedics and a low level of slips in plastics in comparison to the other specialities.

Relationship between glitch rate and hospital site

Elective orthopaedic and vascular surgical procedures were observed in multiple sites (four and two sites respectively), providing an opportunity for inter-site glitch rate comparison amongst teams performing the same types of surgery.

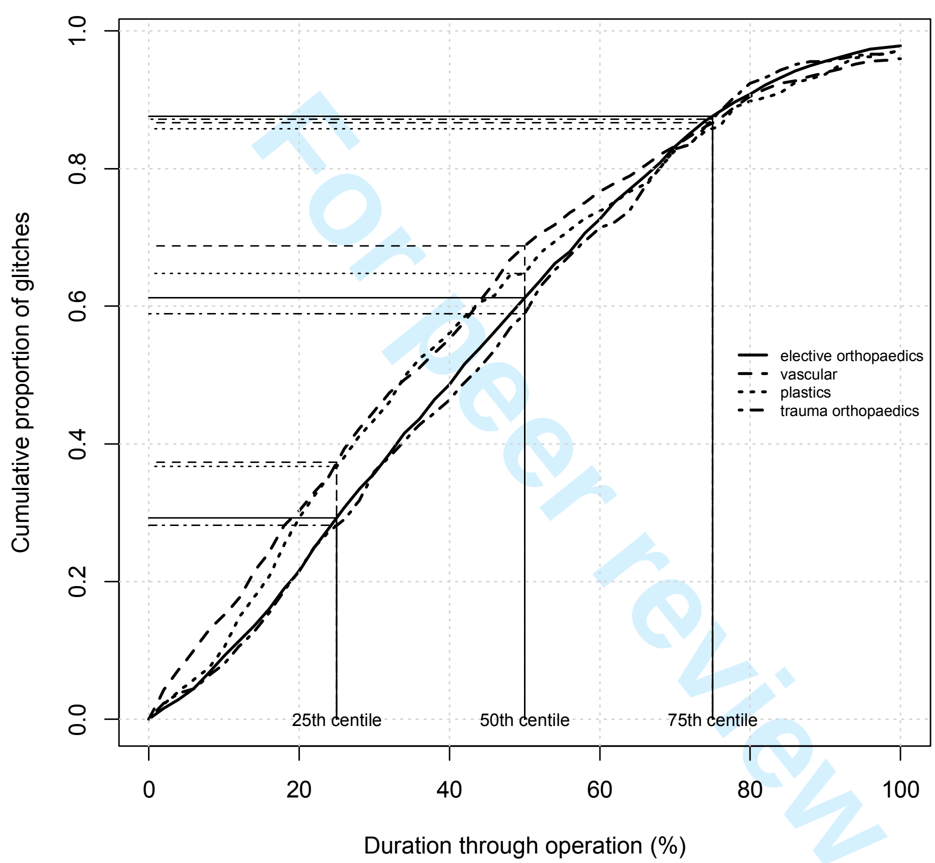
In elective orthopaedics, the mean operating duration varies by 14 minutes between the sites, with glitch rate varying between 6 and 8 glitches per hour. There was a statistically significant heterogeneity in mean glitch rates per operation between the four sites ($p<0.001$). This was explained in 1-1 comparisons by significant differences between individual sites A (mean of 8.1 glitches per hour) and B (mean of 6.0 glitches per hour) (difference = 2.1; 95%CI 0.9 to 3.3; p=0.001), sites D (mean of 8.7 glitches per hour) and B (mean of 6.0 glitches per hour) (difference = 2.7; 95%CI 1.4 to 4; $p<0.001$), and sites C (mean of 7.3 glitches per hour) and B (mean of 6.0 glitches per 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 glitch rate between Site E (mean of 6.3 glitches per hour) and Site D (mean of 10.5 glitches per 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 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 glitches and duration of operation, suggesting a reduction in glitch occurrence in the last 20% of operation duration (Figure 3).

Figure 3 Distribution of glitch occurrence across operative duration



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 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 intra-operative events so that the system and the operative technique can be evaluated for quality and risk. The observed events include different sub-classes such as distractions, process deviations, equipment design problems, and slips (Table 1). These glitches are not necessarily associated with immediate consequences or due to any failure in 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 patient safety, as it may preserve the team's capacity for dealing with 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 potential for serious patient harm[14]. We did not seek to prove causality of glitches as we felt that it would be unwise to attempt to link a glitch with what could be multiple up-stream 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 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 we have called glitches. There may be circumstances where there is blending of both system and human factors. 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 pre-list 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 specialities in emergency and elective settings. The glitch rate can be used to detect similarities and differences in the volume and distribution of process imperfections amongst operating sites and specialities. This novel method builds upon previous

16

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 intra-operative 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 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 inter-speciality/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 40-75% of total glitch events. There were clear differences in event detection profiles between clinical and HF observers, which might have been expected, but clinical observers consistently collected significantly more glitches (Table 2). This finding is at odds with previous research where 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 clinical observers, but this did not appear to be the case. 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 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 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 inter-operator 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 amongst them.

Comment [LM2]: Response to reviewer 2

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 the variety of glitches in the operative process, it would be unlikely that an intervention focussing on only one category of glitch (e.g. distractions) would have as significant an effect as one which focussed on the wider range of issues that our study identifies in the operating theatre. The methodology allows many layers of analysis, from basic arithmetic evaluation to richer contextual analysis. Analysis of the categories of glitches enables the consideration of a system-targeted intervention, with the focus on preventing the creation and propagation of additional work in the operating theatre.

A common criticism of observational research is the bias created as a result of human subjects altering their behaviour when aware of being observed, i.e. the Hawthorne effect[27], although some doubt the importance of this phenomenon[28]. Whilst we cannot exclude bias of this type, the nature of several glitch categories excludes the possibility of mitigation by altered staff behaviour (e.g. the occurrence of a phone call or the dropping of an instrument). Secondly, due to the 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 staff behaviours did not appear to change when the observers were present. Throughout the large data sample, the same patterns of types and 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, whilst 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 observed in previous studies of this type using other direct observation methods[14, 16]. However, 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 observations from 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 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 substantial co-operative 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 both speciality and hospital environment and culture on glitch rates. It might be expected that different specialities would have different rates and types of glitches, but in fact types seemed to show a remarkably consistent pattern and speciality 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 speciality-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 to analysing and reducing 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 operating theatre teams. Our expansive data collection approach has been developed with two independent observers each collecting between 40-75% 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 specialities in the frequency of glitches. Through analysing the frequency and context around 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.

ETHICAL 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].

AKNOWLEDGEMENTS

We would like to thank Julia Matthews and Laura Bleakley for their assistance in intra-operative data collection for the study.

SOURCES OF FUNDING

This paper presents 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.

DATA SHARING

Access to full datasets with all anonymised data will be made available to other researchers after the Programme has been terminated and the principal papers from it have been published. We will maintain this dataset for 3 years and then assign it to the Oxford Research Depository.

REFERENCES

1. Baines, R.J., et al., *Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study.* BMJ Quality & Safety, 2013. **22**(4): p. 9.
2. Hu, Y.Y., et al., *Protecting patients from an unsafe system: the etiology and recovery of intraoperative deviations in care.* Ann Surg, 2012. **256**(2): p. 203-10.
3. Roese, N.J. and K.D. Vohs, *Hindsight Bias.* Perspectives on Psychological Science, 2012. **7**(5): p. 411-426.
4. Mishra, A., K. Catchpole, and P. McCulloch, *The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre.* Qual Saf Health Care, 2009. **18**(2): p. 104-8.
5. Flin, R. and R. Patey, *Non-technical skills for anaesthetists: developing and applying ANTS.* Best Pract Res Clin Anaesthesiol, 2011. **25**(2): p. 215-27.
6. Healey, A.N., S. Undre, and C.A. Vincent, *Developing observational measures of performance in surgical teams.* Qual Saf Health Care, 2004. **13 Suppl 1**: p. i33-40.
7. Carthey, J., *The role of structured observational research in health care.* Qual Saf Health Care, 2003. **12 Suppl 2**: p. ii13-6.
8. de Leval, M.R., et al., *Human factors and cardiac surgery: a multicenter study.* J Thorac Cardiovasc Surg, 2000. **119**(4 Pt 1): p. 661-72.
9. Lingard, L., et al., *Communication failures in the operating room: an observational classification of recurrent types and effects.* Qual Saf Health Care, 2004. **13**(5): p. 330-4.
10. Reason, J.T., *Managing the risks of organizational accidents*. 1997, Aldershot: Ashgate.
11. Helmreich, R.L., *On error management: lessons from aviation.* BMJ, 2000. **320**(7237): p. 781-5.
12. de Vries, E.N., et al., *Effect of a Comprehensive Surgical Safety System on Patient Outcomes.* New England Journal of Medicine, 2010. **363**(20): p. 1928-1937.
13. Catchpole, K.R., et al., *Identification of systems failures in successful paediatric cardiac surgery*. Ergonomics, 2006. **49**(5-6): p. 567-88.

14. Catchpole, K.R., et al., *Improving patient safety by identifying latent failures in successful operations.* Surgery, 2007. **142**(1): p. 102-10.
15. Parker, W.H., *Understanding errors during laparoscopic surgery.* Obstet Gynecol Clin North Am, 2010. **37**(3): p. 437-49.
16. Wiegmann, D.A., et al., *Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation.* Surgery, 2007. **142**(5): p. 658-65.
17. Healey, A.N., N. Sevdalis, and C.A. Vincent, *Measuring intra-operative interference from distraction and interruption observedin the operating theatre.* Ergonomics, 2006. **49**(5-6): p. 589-604.
18. Schraagen, J.M., et al., *Assessing and improving teamwork in cardiac surgery.* Quality and Safety in Health Care, 2010. **19**(6): p. e29.
19. Sevdalis, N., et al., *Annoyances, disruptions, and interruptions in surgery: the Disruptions in Surgery Index (DiSI).* World J Surg, 2008. **32**(8): p. 1643-50.
20. Morgan, L., et al., *Observing and Categorising Process Deviations in Orthopaedic Surgery.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2011. **55**(1): p. 685-689.
21. Robertson, E., Hadi, M., Morgan, L., Pickering, S., Catchpole, K., Griffin, D., McCulloch, P., *The development of process maps in the training of surgical and human factors observers in orthopaedic surgery.* Int J Surg, 2011. **9**(7): p. 1.
22. Moss, S., *Human Factors Reference Group, Interim Report*, 2012, Department of Health. p. 56.
23. 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*. 2000: The Stationery Office.
24. Kohn, L.T., J. Corrigan, and M.S. Donaldson, *To err is human : building a safer health system*. 2000, Washington, D.C.: National Academy Press. xxi, 287 p.
25. Catchpole, K., *Observing Failures in Successful Orthopaedic Surgery*, in *Safer surgery : analysing behaviour in the operating theatre*, R.H. Flin and L. Mitchell, Editors. 2009, Ashgate: Aldershot. p. xxvi, 456 p.
26. Catchpole, K. and D. Wiegmann, *Understanding safety and performance in the cardiac operating room: from 'sharp end'to 'blunt end'.* BMJ Quality & Safety, 2012. **21**(10): p. 807-809.
27. Sage Publications, I., *Hawthorne Effect. The SAGE Encyclopedia of Social Science Research Methods. Sage Publications, Inc*. Thousand Oaks, USA: Sage Publications, Inc.
28. Levitt, S.D. and J.A. List, *Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments.* American Economic Journal-Applied Economics, 2011. **3**(1): p. 224-238.

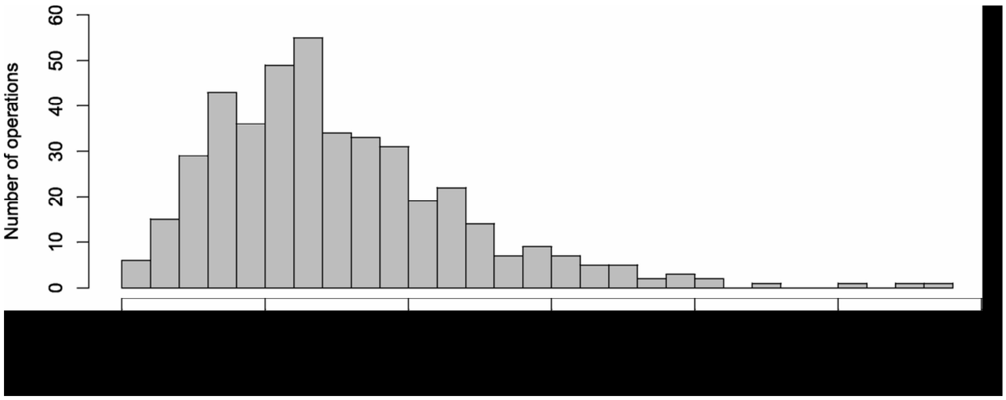


Figure 1 Distribution of glitch rate by operation

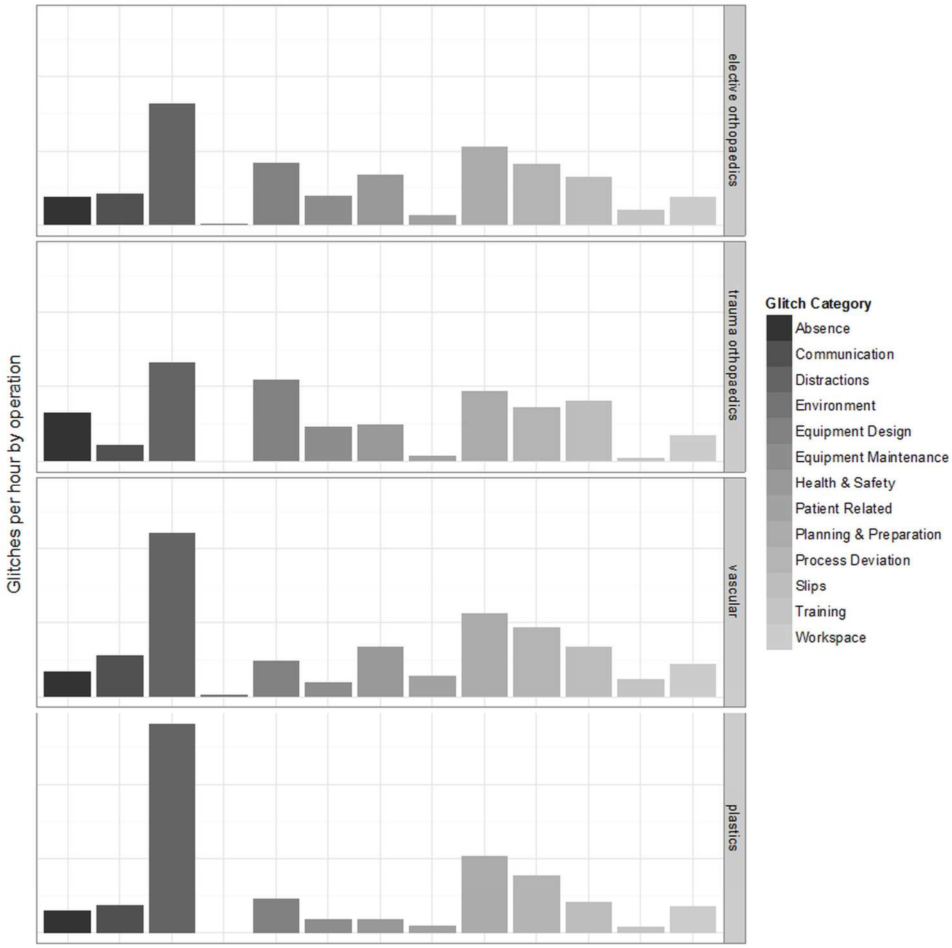


Figure 2 Mean glitch rate by operation for each speciality

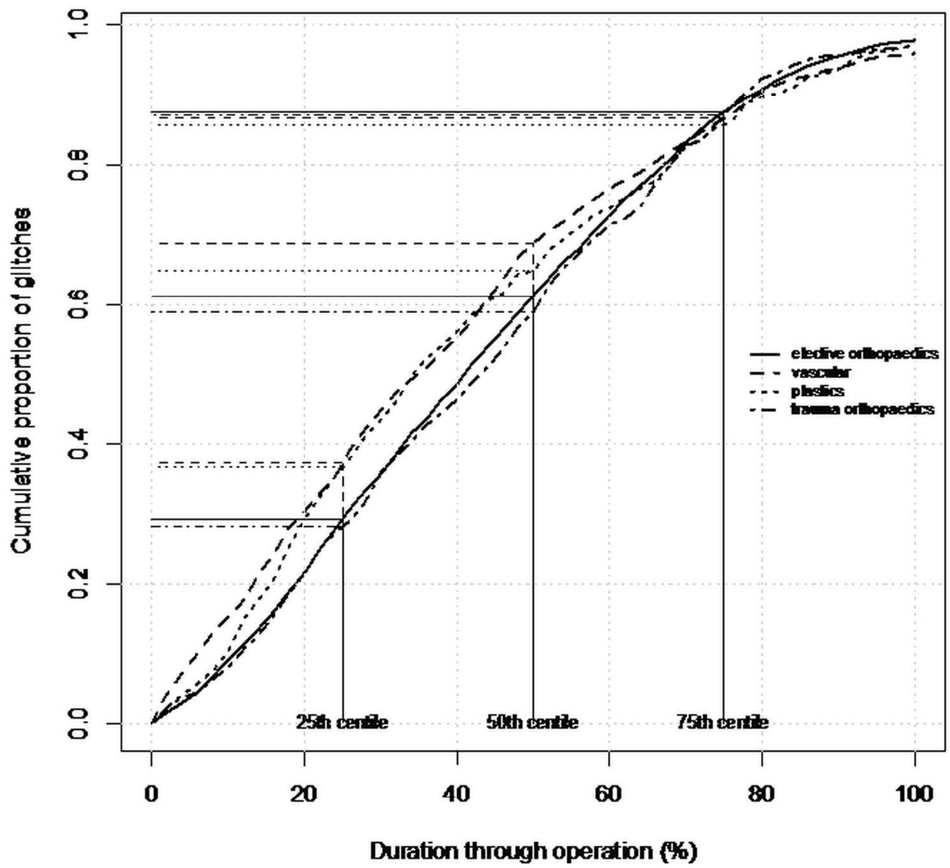


Figure 3 Distribution of glitch occurrence across operative duration