



Published in final edited form as:

BMJ Qual Saf. 2017 December ; 26(12): 1015–1021. doi:10.1136/bmjqs-2016-006407.

Framework for direct observation of performance and safety in healthcare

Ken Catchpole¹, David M Neyens², James Abernathy³, David Allison⁴, Anjali Joseph⁴, and Scott T Reeves⁵

¹Department of Anesthesia and Perioperative Medicine, Medical University of South Carolina (MUSC), Charleston, South Carolina, USA

²Industrial Engineering, Clemson University, Clemson, South Carolina, USA

³Department of Anesthesiology and Critical Care Medicine, Johns Hopkins Medicine, Baltimore, Maryland, USA

⁴Graduate Program in Architecture + Health, Clemson University, Clemson, South Carolina, USA

⁵Department of Anesthesia and Perioperative Medicine, Medical University of South Carolina, Charleston, South Carolina, USA

INTRODUCTION

Non-participant direct observation of healthcare processes offers a rich method for understanding safety and performance improvement. As a prospective method for error prediction and modelling, observation can capture a broad range of performance issues that can be related to higher aspects of the system.^{1–5} It can help identify underlying and recurrent problems⁶ that may be antecedents to more serious situations.⁷ It is also a way to understand the complexity of healthcare work that might otherwise be poorly understood or ignored,⁸⁹ how workarounds influence work practices and safety,¹⁰ and is of fundamental importance to practitioners wishing to understand resilience in the face of conflicting workplace pressures.¹¹¹² In some cases it will lead to the direct observation of near-misses or precursor events that might otherwise not be reported,¹³¹⁴ while in others the observation process may lead to, or be a specific part of, improvement methodologies.¹⁵

Observation allows us to move from ‘work as imagined’ (ie, what should happen, what we think happens or what we are told happens) to ‘work as done’ (what really happens).^{16–18} This also creates a set of unique technical challenges, from the initial question of what should be observed, the role of the observer, supporting the observer in data collection and protecting human subjects, to the non-linear relationships between outcomes, accidents and

No commercial use is permitted unless otherwise expressly granted.

Correspondence to Professor Ken Catchpole, Department of Anesthesia and Perioperative Medicine, Medical University of South Carolina (MUSC), Charleston, South Carolina 29407, USA; catchpol@musc.edu.

Contributors KC conceived the paper. All authors discussed content and participated in writing and editing.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

their deeper systemic causes. The design of observation studies within a clinical context requires a range of trade-offs that need to be carefully considered, yet little has been formalised about how those decisions are made.

This viewpoint paper considers those design parameters and their impact on reliability, results and outcomes, and is specifically focused on researchers and quality improvement specialists seeking to design and conduct their own quantitative observational work, particularly, but not exclusively, in acute settings. Carthey's¹⁹ considerations of observer skills, derived from influential work on the quantification of process¹⁴ and behaviours²⁰ in relation to outcomes in surgery, served as the starting point for our own work. Here, we attempt to consider wider interactions between the study design, observation methodologies, and the relationship between observer and observed, overall seeking to demonstrate the necessarily adaptive, and unavoidably qualitative, nature of this type of research as it develops.

STUDY DESIGN

General methodology

Despite the strength and appropriateness of qualitative methods, there are many reasons to seek quantification of behaviours, processes and other system qualities. Quantitative data are amenable to prioritisation; outcomes, measurements and comparisons; cost/benefit considerations; statistical modelling; or simply to publish in journals and to reach audiences that traditionally hold quantification in higher scientific regard than qualitative results. Although the application of a unidimensional measure to multidimensional phenomena can be simplistic and potentially misleading, quantification through the systematization of measurements seems to allow purer objective evaluation of theories, engineering of systems, assessment of interventions, balancing of limited resources, and accessible, influential, falsifiable results. It is not our aim to rehearse further the advantages of qualitative or quantitative designs, or the codependence of observation and intervention, but note that it is understandable there is much motivation, especially at the outset of a project, to be focused on the quantification on a small number of dimensions or 'outcomes'. In the next sections, we illustrate why such designs are likely to drift to multiple dimensions, or, indeed, to qualities rather than quantities.

Purpose of study

The purpose of the study defines early goals and constraints that may need to be reconsidered as the study progresses. Studies that seek to quantify the frequency of specific events can quickly reveal unexpected complexities. Broader systems analysis studies—exploring what we 'don't know that we don't know'—require less specificity and an adaptive data collection method that, from the outset, is likely to be at least partly qualitative in nature. Intervention evaluation studies require sufficient rigour to be repeatable, which in turn requires specific quantification, tight definition and more focus, which ultimately means that qualities will be lost from the rest of the data.²¹ For studies that use the same observational measures as part of the improvement,¹⁵²² observations (and observers) become codependent with outcomes, which can create challenges for repeatability, and

spread. It is not unusual to set out to explore one phenomenon (eg, training reduces errors) and through observation find that another provides more insight (eg, errors are unchanged but efficiency improves). Thus, it is often necessary to adapt study goals and methods to address observational discoveries, while avoiding adaptations that undermine the original purpose. The next sections explore in more detail why these adaptations are likely.

Focus of study

Clinical processes can be opaque and socially situated, with uncertain goals that may be conflicting and sometimes impossible to achieve, and may not have been engineered with specific tolerances.^{23,24} It is tempting to think that books, guidelines, regulations or best practices might define a reasonable observational template. However, study designs that precisely define the ideal process from this ‘work as imagined’ frame of reference may not reflect the complexity of the work that is done and fail to represent critical performance mediators. What might be seen as ‘poor’ or highly variable performance may have no relationship with outcome, while a process that is measured as being highly reliable may not always reflect safe and appropriate conduct of care. Medical practice varies with country, site, unit, specialty and profession. Often there are disagreeing policies, differing processes, heated debates among professional groups about what is ‘right’ and equivocal evidence of effects on patient outcomes. Mask wearing in surgery (nationality, specialty and professionally dependent) or ‘no sleeves, watches and ties’ policies (nationality dependent) for infection control are examples of geographically and socially situated practices. Any basis for a structured data collection may simply be socially constructed, with deviations from this template seen as socially undesirable, even though they may not necessarily be clinically undesirable (and possibly, the opposite). Measures designed only with local knowledge may reflect deviations from national guidelines or accepted practice (and often deviate from the best evidence), while other variations may simply reflect reasonable and necessary adaptations.

Rules, protocols and best practices cannot reasonably account for all eventualities, nor are all unequivocally evidence-based, and there is often no direct relationship with harm. There are often multiple ways to complete the same tasks, no recognised ‘best way’ to do so, with legitimate reasons for variation between providers. Attempts to apply deterministic models of measurement to stochastic healthcare processes may not capture the variability required to deliver patient care. Any attempt to measure only a small number of items is unlikely to address the full complexity of work, and thus may misrepresent reliability, causation or behaviours. In some situations, it may encourage the shaping of behaviour towards a norm artificially constructed by the research.²⁵ Ideally, the relationships between observable events, safety-critical situations and outcomes would be empirically established. In practice, however, there are few observable events that are clearly identifiable, clearly measurable, have a clear effect on an outcome and that occur sufficiently frequently with enough variability. Observations must rely on defining and categorising surrogate measures of safety and performance, the identification and validation of which is a complete area of research in and of itself.^{26–30}

Study drift

One consequence of this uncertainty over what to measure and how to measure it (and the consequent fear of finding nothing) is that studies drift from ‘thin’ designs, which focus on one measure with a clearly defined collection method, towards ‘thick’ designs, which attempt to measure a collection of variables or concepts that initial observations reveal to be important. The focus on one specific, well-defined measure, with the exclusion of others, runs a risk of generating data with little variation or of little meaning. Thicker, qualitative or semiquantitative approaches provide an opportunity to explore the deeper meaning of the data collected, allow further study, classification and subcategorisation, and complement numerical results. They also ensure that researchers will eventually derive something that will enhance understanding of systems design and intervention, rather than simply a numerical value representing a dimension of interest.

The study design and measurement paradoxes therefore are that when we seek to measure something specific in a complex healthcare system, we initially base our measure on ‘work as imagined’, and it is only in performing the study that we recognise that ‘work as done’ might differ in important ways. An insufficiently sophisticated model of the work system using simplistic measurement methods and ideal system states may appear to detect negative deviations in quality, reliability or successful system function that instead reflect positive deviations necessary for individualised care. Moving towards a ‘thicker’ design that encompasses a more realistic view of the real work environment might substantially change the course of a study, and invalidate previous data, and possibly the entire research question. This necessarily iterative process of developing the design, measurement and analysis of a direct observation study is usually omitted from research discussions, plans or manuscripts. Thus, researchers are not able to learn from others’ experiences and must experience adaptability and drift in their own research questions.

OBSERVATIONAL METHODOLOGY

Since most healthcare work is not wholly predictable, we rely on observers to make necessary evaluative judgements, record their findings and participate in the analysis. In the process of developing the methodology, the observer becomes the instrument of detection, adapting their observations based on the perceived purpose of the study, the observability and frequency of events, and their experience.^{1931–34} The following are the general processes of observation: (1) an event or events of interest need to occur in the presence of the observer, (2) it needs to be detected by the observer, (3) it needs to be recorded, and (4) it will usually need to be classified either immediately or post-hoc, (5) then analysed in order to reach a higher level of understanding (figure 1). The specificity of the metric, the method of data recording, the classification scheme, the skill, background and training of the observer, and the nature and tempo of work environment will all affect the reliability of the data collected.

Observer background and skills

Observers are not mute, asocial, disengaged data collection instruments functioning independently of clinical context and social interactions, but respond to social and

situational factors. Indeed, they frequently use a broad range of contextual cues to make judgements that arrive at observations that can be analysed. Carthey¹⁹ offers a detailed exploration of the selection, development and training, which will not be covered here. The background, training and experience of the observer in the clinical context being studied, the observational methodology being deployed, and theoretical perspective being explored affect the quality of the observations. Local or professional biases can be addressed with pairs of observers (a context expert and a theoretician), by employing post-hoc critiques to observational data, or by asking participants to comment on their observed behaviours (which video-reflexive methodology takes further by using participant feedback as the intervention). At the very least, a period of obtaining and demonstrating a sufficient level of contextual, theoretical and observational expertise should be part of the design of the observational methodology.

Sampling

Events of interest need to occur that have some aspect of observability. Some concept of the natural incidence of the event needs to be already understood in order for the observer to be present to capture it. This may mean being present for a significant time prior to the events to ensure reliable capture and to establish some context for observations, however limited. This event then needs to be perceived by the observer above the 'noise' of otherwise normal system function. The clearer the ability of the observer to detect signal from noise, the more reliable capture will be, while over-specificity may exclude data of interest. Frequency, predictability and repetition of the event of interest affect the response bias, and thus the likelihood of perception and consequent recording.³⁵

Recording

A 'check box' system³⁶ can be quick and easy to employ, but has all the risks associated with a 'thin' design, which might miss key details, without the possibility for further post-hoc exploration. Employing a note taking system is a 'thicker' approach that still allows quantification of prespecified events¹³ but might also allow post-hoc analysis and retention of some of the complex contextual richness of the observations. During high-tempo events, the observer will need to manage the requirements for data recording with their ability to observe unfolding events, resulting in reliability and validity variations between high-tempo and low-tempo periods. It may also be those high-tempo periods that provide the most insight into unstable or unsafe system function, and that present deviations from a norm that a 'thin' methodology would not capture.

Recording results straight to an electronic device such as a tablet alleviates the need to input data later, but is limited by an a priori specific data collection scheme. Tablets may not be ideal for making freehand notes or diagrams or descriptions of observed event, are limited by battery life, and can be uncomfortable to use for long periods while standing. This often makes paper the recording medium of choice because it is reliable, robust, compact, discreet, fast and highly adaptable, despite the necessary collation of data after the observation session.

Coding

Ideally, a coding scheme should be explicitly defined, independent of context, exhaustive, mutually exclusive and easy to record.³⁷ While a greater number of categories can provide more detail or specificity, the learning curve for the observer will be steeper the greater the number of categories and the more overlap they have. This can make classification schemes particularly prone to iteration during methodological development. Post-hoc classifications may allow more complex coding schemes, but need sufficient descriptive notes at the observation stage to ensure appropriate representation at the classification stage. Indeed, since context is not independent of observations (rather, it is a necessary part), the ideal of a coding scheme independent of context is arguably impossible to achieve, while post-hoc interpretations may sometimes miss the context in which the original observation was made. Once again, high specificity ('Thin') or numerical focus alone may lose important context information, whereas lower specificity and multiple dimensions of interest ('Thick') may place greater demands on the observer. Consequently, this should be carefully considered during pilot studies and observer training.

Analysis

When approaching analysis, it is important to consider the following: (1) how the observational method may have changed over the course of the study, (2) how the observer(s) may have changed, (3) how the research question may have changed, and (4) the role of the observer and their experience on the observations. A drift from single to multivariable data collection requires a more complex analytical model of system function than may have originally been planned. This model will be directly informed by the experiences of the observer (eg, 'in measuring the effect of our intervention on outcome measure X, observers found that Y, and in some instances, Z might have confounded our results and thus needed to be taken into account'). The immersion of the observer in the work context will position them to speculate on the meaning and analysis of the data in a way that pure numbers may not reflect. Variability is a necessity of healthcare delivery, so interventional studies in particular may confuse genuine improvements with increased adherence to a process; or may simply reflect the response of the study population to what is being measured, rather than improved quality or safety. Understanding the qualities of measured data is fundamental to establishing their meaning, impact and underlying causes, and may also lead to stories or clinical examples that can be more powerfully convincing than data alone. In multiple studies in surgery, surprising near-miss safety events were observed that were not the focus of the study but were important findings with the broader context of surgical safety, and thus formed the subsequent analytical approach and mechanistic hypotheses.¹³ Consequently, utilising the experience of the observer in the interpretation of data yields a richer, more representative analysis, which may not be purely empirical, and yet is rarely reported.

Video-reflexive techniques—where participants view and respond to their own videos—can also be powerful.¹⁵³⁸³⁹ Ostensibly a qualitative sociological methodology where video recordings are used as part of the intervention to help clinical teams to reflect on their own work, the active involvement of the 'observer' in video-reflexive studies is in the focus of the video (which is not fixed), in the subsequent feedback to staff, and in assisting staff with the

selection and development of their own interventions. The observer thus works alongside the observed, and the video is not ‘data’ but the catalyst which enables change. This particular method overcomes many of the contextual, quantification, power-related and other challenges of observational research, although may not lend itself easily to replication, falsification, comparative or cost/benefit analyses, or the identification of deeper systems solutions.

OBSERVER AND OBSERVED

It can be difficult to prepare for the visceral, emotive and existentially challenging nature of healthcare. Collecting observational data in clinical settings can be cognitively, physically, philosophically and emotionally challenging, especially since impartiality is required. Events can be traumatic even for staff. Working alone, in this complex experiential, moral, technical and emotional milieu, can be isolating. Any direct observation study needs to ensure that the observer has the appropriate support to address the range of personal and ethical challenges that they will face. Encouragingly, a growing cadre of publications and experienced observers are available for support that was not available a decade ago.⁴⁰⁴¹

Interdependence

It may be expedient to keep the details of the study design opaque to staff, or it may be detrimental to transparency, participation and goodwill. The observer may be unwelcome, mistrusted or belittled. Alternatively, they may become close to one or more members of staff or patients whom they are observing. Units with ongoing conflicts will try to court observers to be partisan, requiring considerable diplomacy so as not to appear preferential or exacerbate tensions. Observers may be asked to help in peripheral clinical work—for example, running small errands or answering phones—which can affect the observations but is necessary to maintain the goodwill of the people they are observing.

Carroll⁴¹ characterises these interdependencies as working ‘outside’ (ie, observing the team as an outsider with minimal interference, treating participants as objects of scrutiny), ‘inside’ (observing the team as one of their own, similar to participant observation or traditional ethnography) and ‘alongside’ (where the researchers and the researched, and the observations and the observed, are considered simultaneously). Planning of research strategies to support both the observed and the observer can help avoid many political, interpersonal and power-related challenges in observational designs.

Power dynamics

Observers can alter, or be altered by, power dynamics within a unit. Especially when observing safety-related events, being unable to help staff or patients can create feelings of powerlessness and frustration, and observers can feel implicated in accidents simply through this powerlessness. They can face criticism from their own colleagues for not reporting incidents or not taking a more active role in events, despite a range of practical barriers to doing so, the ethics of which are far from clear. Unwelcoming participants may be a reflection of a perceived shift or challenge to power relations within a team or organisation than a specific rejection of the observer or their research. Observers may also find

themselves deliberately placed into a clinical situation to challenge a perceived power dynamic, which may be an entirely inappropriate perception (eg, ‘go look at these people who are the problem’), requiring a difficult conversation with the study leader. While somewhat unpredictable, this needs to be carefully considered and managed, through design, observer training and in understanding the responses of staff to the presence of the observer.

Protection of human subjects

Meaningful results need to be balanced with appropriate voluntary, informed and consensual participation, confidentiality, and medicolegal protection. Given how data collection changes over the course of a study, it may not be possible to fully inform staff or patients of what will be collected—or at least it may be uncertain at the outset of the project. Furthermore, observational data may expose units or whole hospitals to criticism, with studies withheld by organisations because results appear to be unfavourable. If there have not yet been instances where an observational research project is used to inform medicolegal or disciplinary procedures, it seems likely this will eventually happen. The approach to these studies may need to be cocreated between scientists, those being observed, ethicists and patients. Establishing this dialogue early in the study design process and accepting there will be iterations will bring benefits.

CONCLUSIONS

The progress from the initial ‘work as imagined’ state to understanding ‘work as done’ is a key part of developing an observational approach but is often overlooked or under-reported. In the pursuit of objectivity, it is not unusual to find a priori assumptions have been misleadingly simplistic; measures are not as definitive as initially hoped; necessary variation that invalidates the measurement; processes that are not as unreliable as suspected; measurement methods that do not translate from one unit to the next; or hypotheses that do not sufficiently represent an observed mechanism of effect. Studies often drift from ‘Thin’ designs, which can focus on a small number of specific metrics but may misrepresent complexity, towards more complex ‘Thick’ approaches, which might include a broader range of less well-defined quantitative and qualitative data, but may not be as statistically or methodologically definitive. Immersion of the observer within the workspace informs the hypothesis generation, measurement, interpretation and subsequent analysis. Consequently, observational approaches will often require the simultaneous iterative development of hypotheses, system models, metrics, methods, observer expertise and ethical protections. In table 1, we have summarised these considerations into a set of dimensions that broaden considerations for observational designs. Although this iterative process may be among the most substantive, labour-intensive and content-rich parts of a study—and regularly continues throughout—it is rarely reported systematically. We hope this will encourage improved designs, more detailed reporting and extended methodological considerations in the future.

Acknowledgments

Funding This paper was supported by grant number P30HS0024380 from the Agency for Healthcare Research and Quality.

REFERENCES

1. Gurses AP, Kim G, Martinez EA, et al. Identifying and categorising patient safety hazards in cardiovascular operating rooms using an interdisciplinary approach: a multisite study. *BMJ Qual Saf* 2012;21:810–8.
2. Cohen TN, Cabrera JS, Sisk OD, et al. Identifying workflow disruptions in the cardiovascular operating room. *Anaesthesia* 2016;71:948–54. [PubMed: 27396248]
3. Palmer G, Abernathy JH, Swinton G, et al. Realizing improved patient care through human-centered operating room design: a human factors methodology for observing flow disruptions in the cardiothoracic operating room. *Anesthesiology* 2013;119:1066–77. [PubMed: 23811697]
4. Wiegmann DA, ElBardissi AW, Dearani JA, et al. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery* 2007;142:658–65. [PubMed: 17981185]
5. Catchpole KR, Giddings AEB, Hirst G, et al. A method for measuring threats and errors in surgery. *Cogn Technol Work* 2008;10:295–304.
6. Tucker AL, Spear SJ. Operational failures and interruptions in hospital nursing. *Health Serv Res* 2006;41:643–62. [PubMed: 16704505]
7. Catchpole KR, Giddings AE, Wilkinson M, et al. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;142:102–10. [PubMed: 17630006]
8. Christian CK, Gustafson ML, Roth EM, et al. A prospective study of patient safety in the operating room. *Surgery* 2006;139:159–73. [PubMed: 16455323]
9. Morgan L, Robertson E, Hadi M, et al. Capturing intraoperative process deviations using a direct observational approach: the glitch method. *BMJ Open* 2013;3:e003519.
10. Rivera-Rodriguez AJ, Karsh BT. Interruptions and distractions in healthcare: review and reappraisal. *Qual Saf Health Care* 2010;19:304–12. [PubMed: 20378621]
11. France DJ, Levin S, Hemphill R, et al. Emergency physicians' behaviors and workload in the presence of an electronic whiteboard. *Int J Med Inform* 2005;74:827–37. [PubMed: 16043391]
12. Edwards A, Fitzpatrick LA, Augustine S, et al. Synchronous communication facilitates interruptive workflow for attending physicians and nurses in clinical settings. *Int J Med Inform* 2009;78:629–37. [PubMed: 19482544]
13. Catchpole KR, Giddings AE, de Leval MR, et al. Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;49:567–88. [PubMed: 16717010]
14. de Leval MR, Carthey J, Wright DJ, et al. Human factors and cardiac surgery: A multicenter study. *J Thorac Cardiovasc Surg* 2000;119:661–72. [PubMed: 10733754]
15. Carroll K, Iedema R, Kerridge R. Reshaping ICU ward round practices using video-reflexive ethnography. *Qual Health Res* 2008;18:380–90. [PubMed: 18235161]
16. Clay-Williams R, Hounsgaard J, Hollnagel E. Where the rubber meets the road: using FRAM to align work-as-imagined with work-as-done when implementing clinical guidelines. *Implement Sci* 2015;10:125. [PubMed: 26319404]
17. McNab D, Bowie P, Morrison J, et al. Understanding patient safety performance and educational needs using the 'Safety-II' approach for complex systems. *Educ Prim Care* 2016;27:443–50. [PubMed: 27800711]
18. Blandford A, Furniss D, Vincent C. Patient safety and interactive medical devices: Realigning work as imagined and work as done. *Clin Risk* 2014;20:107–10. [PubMed: 25866466]
19. Carthey J The role of structured observational research in health care. *Qual Saf Health Care* 2003;12:13ii–16. [PubMed: 12571339]
20. Carthey J, de Leval MR, Wright DJ, et al. Behavioural markers of surgical excellence. *Saf Sci* 2003;41:409–25.
21. Dixon-Woods M, Leslie M, Tarrant C, et al. Explaining Matching Michigan: an ethnographic study of a patient safety program. *Implement Sci* 2013;8:13. [PubMed: 23347357]
22. Dixon-Woods M, Bosk C. Learning through observation: the role of ethnography in improving critical care. *Curr Opin Crit Care* 2010;16:639–42. [PubMed: 20808219]
23. Kapur N, Parand A, Soukup T, et al. Aviation and healthcare: a comparative review with implications for patient safety. *JRSM Open* 2016;7:205427041561654.

24. Catchpole K Spreading human factors expertise in healthcare: untangling the knots in people and systems. *BMJ Qual Saf* 2013;22:793–7.
25. Kovacs-Litman A, Wong K, Shojania KG, et al. Do physicians clean their hands? Insights from a covert observational study. *J Hosp Med* 2016;11:862–4. [PubMed: 27378510]
26. Howell AM, Burns EM, Bouras G, et al. Can Patient Safety Incident Reports Be Used to Compare Hospital Safety? Results from a Quantitative Analysis of the English National Reporting and Learning System Data. *PLoS One* 2015;10:e0144107. [PubMed: 26650823]
27. Schnoor J, Rogalski C, Frontini R, et al. Case report of a medication error by look-alike packaging: a classic surrogate marker of an unsafe system. *Patient Saf Surg* 2015;9:12. [PubMed: 25767564]
28. Laureshyn A, Goede M, Saunier N, et al. Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway. *Accid Anal Prev* 2017;105. [PubMed: 28399463]
29. Ghanipoor Machiani S, Abbas M. Safety surrogate histograms (SSH): A novel real-time safety assessment of dilemma zone related conflicts at signalized intersections. *Accid Anal Prev* 2016;96:361–70. [PubMed: 25951965]
30. Moorthy K, Munz Y, Forrest D, et al. Surgical crisis management skills training and assessment: a simulation[corrected]-based approach to enhancing operating room performance. *Ann Surg* 2006;244:139–47. [PubMed: 16794399]
31. Guerlain S, Adams RB, Turrentine FB, et al. Assessing team performance in the operating room: Development and use of a “black-box” recorder and other tools for the intraoperative environment. *J Am Coll Surg* 2005;200:29–37. [PubMed: 15631917]
32. Guerlain S, Calland JF. RATE: an ethnographic data collection and review system. *AMIA Annu Symp Proc* 2008:959. [PubMed: 18999151]
33. Morgan L, Pickering S, Catchpole K, et al. Observing and Categorising Process Deviations in Orthopaedic Surgery. *Proc Hum Factors Ergon Soc Annu Meet* 2011;55:685–9.
34. Parker SE, Laviana AA, Wadhera RK, et al. Development and evaluation of an observational tool for assessing surgical flow disruptions and their impact on surgical performance. *World J Surg* 2010;34:353–61. [PubMed: 20012288]
35. Swets JA, Green DM, Getty DJ, et al. Signal detection and identification at successive stages of observation. *Percept Psychophys* 1978;23:275–89. [PubMed: 748850]
36. Lingard L, Regehr G, Espin S, et al. A theory-based instrument to evaluate team communication in the operating room: balancing measurement authenticity and reliability. *Quality and Safety in Health Care* 2006;15:422–6. [PubMed: 17142591]
37. Robson C *Observational Methods Real World Resarch*. Chichester, UK: John Wiley & Sons Ltd, 2011:315–41.
38. Hor SY, Iedema R, Manias E. Creating spaces in intensive care for safe communication: a video-reflexive ethnographic study. *BMJ Qual Saf* 2014;23:1007–13.
39. Iedema R, Ball C, Daly B, et al. Design and trial of a new ambulance-to-emergency department handover protocol: ‘IMIST-AMBO’. *BMJ Qual Saf* 2012;21:627–33.
40. Blandford A, Berndt E, Catchpole K, et al. Strategies for conducting situated studies of technology use in hospitals. *Cogn Technol Work* 2015;17:489–502.
41. Carroll K Outsider, insider, alongsider: Examining reflexivity in hospital-based video research. *International Journal of Multiple Research Approaches* 2009;3:246–63.

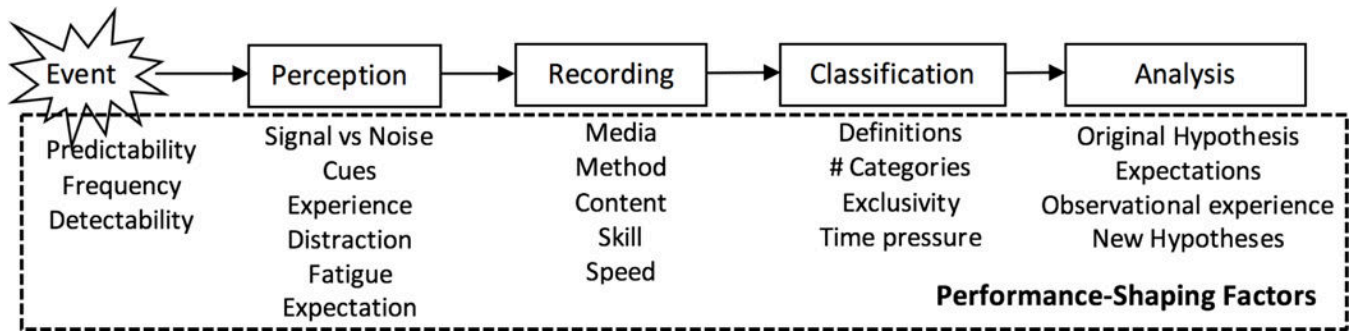


Figure 1.
The observation process.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Design framework for direct observation studies

	Design dimension	Feature
Study design	General methodology	Empiricism that withstands scrutiny is the basic construct for a scientific approach, but quantitative research will not address the full complexity of work. Qualitative observations may be necessary even when conducting quantitative studies.
	Purpose	Diagnostic studies obtain prospective evidence of systemic deficiencies to predict new problems or eliminate hindsight bias, which demands a more qualitative method. Evaluative studies require tighter definition, more focus and sufficient rigour to be repeatable.
	Focus of study	Process-oriented tasks are easier to observe and quantify, but processes may differ unexpectedly. Many safety-critical healthcare tasks have a weakly defined process, or have multiple ways to achieve a goal.
	Study drift	Initial study designs are often based on a simplistic understanding of work; failure to adapt may misrepresent the work, while drifting towards a 'thicker' more representative model may change the research questions or invalidate previous data.
Observational methodology	Observer background and skills	The skills, background and experience of observer(s) will reflect their ability to understand, record and interpret observations. Observer training and evaluation prior to the main data collection are essential.
	Sampling	Frequency, predictability and repetition of the tasks to be studied define how easily meaningful data will be collected. Highly variable processes will require a greater sample size. Rare events may need to be captured over long periods.
	Recording	Highly structured recording is quicker but note taking offers more flexibility in analysis and interpretation. The tempo of events and recording demands need to be balanced. Recording to an electronic device may help data management, but may not be as robust or discrete.
	Coding	Coding systems require a balance of breadth and depth. Coding can occur at the time of observation, or post-hoc if sufficient detail is retained.
	Analysis	The experience of the observer is an essential part of the analysis. The adaptations made to the design over the course of the study should also be carefully considered as part of the analysis. Intervention effects may be due to the observation.
Observer and observed	Interdependence	Observers need to be sensitive to operational needs, may not be welcomed, and events observed can be traumatic. The observer can work with, collaborate with, or be as far as possible separate from, those being observed.
	Power dynamics	Observation can affect and be affected by a variety of power dynamics within the study environment. Sensitivity and planning will help reduce these issues.
	Protection of human subjects	Observation falls outside traditional medical ethical guidance. How and when should an observer intervene, and what is the medicolegal status of observer and their data?