

# The role of structured observational research in health care

J Carthey

Structured observational research involves monitoring of healthcare domains by experts to collect data on errors, adverse events, near misses, team performance, and organisational culture. This paper describes some of the results of structured observational studies carried out in health care. It evaluates the strengths, weaknesses, and future challenges facing observational researchers by drawing lessons from the human factors and neonatal arterial switch operation (ASO) study in which two human factors specialists observed paediatric cardiac surgical procedures in 16 UK centres. Lessons learned from the ASO study are germane to other research teams embarking on studies that involve observational data collection. Future research needs robust observer training, clear measurable criteria to assess each researcher's domain knowledge, and observational competence. Measures of inter-rater reliability are needed where two or more observers participate in data collection. While it is important to understand the factors that lead to error and excellence among healthcare teams, it is also necessary to understand the characteristics of a good observer and the key types of error that can occur during structured observational studies like the human factors and ASO project.

**E**thnography is the observation and systematic recording of human cultures and the descriptive work produced from such research.<sup>1</sup> It usually involves an individual or team of researchers who "live" alongside a given workforce, tribe or team, observing how they behave and systematically recording these observations. In health care, ethnographic approaches—originally developed by social scientists—have been adapted and extended; as well as collecting qualitative data, a quantitative (statistical) analysis is also carried out. This is called "structured observational research".

## STRUCTURED OBSERVATIONAL RESEARCH

To date, structured observational research in health care has identified the types, frequency, and severity of adverse events that occur in different domains, including drug administration,<sup>2-5</sup> emergency departments,<sup>6,7</sup> operating theatres,<sup>8,9</sup> anaesthesia,<sup>10</sup> obstetrics,<sup>11</sup> and intensive care units.<sup>12-14</sup> Structured observational studies have also identified the individual, team, and organisational precursors of adverse

*Qual Saf Health Care* 2003;12(Suppl II):ii13-ii16

events.<sup>6,7,15,16</sup> Some examples are given below to illustrate its potential value.

In one study<sup>17</sup> trained observers recorded all adverse events discussed during day shifts, medical ward rounds, nursing shift changes, case conferences, and departmental meetings in three wards. 185/1047 patients had at least one serious adverse event. The likelihood of experiencing an adverse event increased by about 6% for each day of hospital stay.

Structured observational research has also identified the type, frequency, and severity of drug administration errors. Barker *et al*<sup>3</sup> undertook a study of 36 US hospitals in which observations of drug rounds were carried out on nursing units which used high volumes of drugs. The results showed an overall drug administration error rate of 19% (605/3216). The most common error types were wrong time (43%), omission (30%), and wrong dose (17%); 7% of errors (40 per day per 300 patient unit) were judged by an expert panel of physicians to be harmful.<sup>3</sup>

A recent UK study carried out on 10 wards in two hospitals showed that 212/430 observed intravenous drug preparations and administrations had at least one error. An additional 37 had more than one error per drug dose. Giving boluses and preparing drugs that required multiple steps were identified as the drug administration tasks most susceptible to error.<sup>4</sup>

Research has also been carried out in the intensive care unit (ICU). One observational study showed an error rate of 1.7 errors per patient per day.<sup>15,16</sup> Another study based in a multidisciplinary ICU identified 777 adverse events among 1024 consecutive patients admitted in 1 year.<sup>14</sup> There were 241 human errors (31%) in 161 patients. These errors were classified as errors in planning ( $n = 75$ ), execution ( $n = 88$ ), and surveillance ( $n = 78$ ). The most serious errors were caused by planning failures. It was calculated that human errors prolonged ICU stay by 425 patient days or the equivalent of 15% of ICU time.<sup>14</sup>

Observational studies have also been carried out in emergency departments. In one US study a trained emergency room nurse and a physician observed cases and rated teamwork behaviours among various teams working in nine US hospitals.<sup>6</sup> Participating teams were then trained how to improve their performance using a framework called MedTeams which trains key communication and coordination skills. The results showed improved teamwork and a significant reduction in errors among trained teams.

Quite often the structure and culture of healthcare organisations acts as a barrier to effective teamwork; structured observational research is good at identifying these problems.

Correspondence to:  
Dr J Carthey, Assistant  
Director of Patient Safety,  
Interagency Working  
Directorate, National  
Patient Safety Agency,  
4-8 Maple Street, London  
W1T 5HD, UK;  
jane.carthey@npsa.nhs.uk

During several of the adverse events observed in the MedTeams study, one team member possessed a skill or had doubts about the course of action that was being taken which, if communicated to the rest of the team, would have prevented the adverse event. Emergency room staff did not speak up because there was a culture of not questioning one's superiors. This behaviour mirrors findings in the aviation field where co-pilots and cabin crew are often reluctant to question the captain's authority.<sup>18, 19</sup>

Similar organisational problems have been identified in the UK accident and emergency (A&E) departments. In one A&E department a rigid horizontal structure, vertical division of labour, and the strict authority dynamics meant that staff who knew important patient safety information could not influence decision making. It was concluded that there was a need to engender a workplace culture in which sapiential authority—that is, derived from experience or availability in an emergency or holding key information—is recognised in addition to authority derived from formal status.<sup>7</sup>

In addition to faulty organisational structures and team cultures, structured observational research has also identified interruptions as a problem for emergency department doctors.<sup>15, 16</sup> Emergency department doctors are “interrupt driven”. They are frequently interrupted and many interruptions result in breaks in task which cause them to switch attention to a different task altogether.<sup>16</sup>

These (and other) studies have made an important contribution to our understanding of factors that influence patient safety. They illustrate what types of errors and adverse events occur in various healthcare settings and, in doing so, they make it clear why structured observational research should be prioritised on patient safety research agendas. However, research teams usually publish their findings and do not critique their methods. This prevents the broader health care and research communities from learning about the types of methodological problems and pitfalls experienced by observational researchers. The objective of this paper is to critique one structured observational study—the human factors and the arterial switch operation (ASO) study<sup>20–22</sup>—to learn about the methodological problems typically experienced when carrying out this type of research. These include measuring interobserver reliability, proving the validity of checklists and other data collection tools, the selection and training of observers, and the characteristics of a good observer.

## HUMAN FACTORS AND THE ARTERIAL SWITCH OPERATION (ASO) STUDY

Human factors is an umbrella term for multidisciplinary studies which analyse people at work to identify how equipment design, organisational, environmental, personal, and social factors influence their performance.

The ASO is carried out on babies who are born with the great vessels of the heart connected to the wrong ventricle. Hence, the aorta is connected to the right ventricle and the pulmonary artery to the left ventricle. The surgical procedure involves putting the patient onto a heart lung bypass machine and freezing the heart with a potassium based solution called cardioplegia. The heart lung bypass machine provides mechanical support to the heart and lungs while the surgical procedure is being carried out. Once the heart is frozen, the cardiac surgeon then transects the native aorta (connected to the right ventricle) and excises the coronary arteries (the tiny vessels that carry oxygenated blood to the heart) from the native aorta and re-implants them into a neo-aorta. A neo-pulmonary artery is also reconstructed using the tissue from the trunk of the native aorta and a piece of tissue called pericardium. Given that the coronary arteries are only

millimetres wide and are extremely fragile, the surgeon is working at the edge of the safety envelope.

The human factors and ASO study involved 16 UK paediatric cardiac centres. Observational case study data were collected by two trained human factors specialists on 173 ASOs. The observers watched each case from the point of induction of anaesthesia to handover of the patient from the operating theatre to the ICU. Throughout each case the observers noted down any errors, problems and notable aspects of good performance. The observer's interpretation was checked with the operating theatre team after each case and a summary report was written.

The case reports differentiated between major and minor events. Minor events are errors that disrupt the surgical flow of the procedure and would not be expected in isolation to have serious consequences to the patient. Major events are errors that are likely to have serious consequences for patient safety. In the statistical analysis, researchers generated two baseline regression models for negative outcomes (the first for death, the second for death and/or near miss) based on patient variables. In the first stage of the analysis the total number of major and minor events per case was added separately to the baseline regression models. The total number of major events per case influenced the probability of death ( $p < 0.001$ ) and death and/or near miss ( $p < 0.001$ ) after adjustment for patient variables. Whether or not major events were compensated for—that is, recognised and recovered from—was also a strong predictor of death ( $p < 0.003$ ). Even for the most serious types of errors (major events), appropriate compensation produces a good outcome.<sup>20</sup>

In contrast, when minor events were added separately to the baseline regression models it was the overall number per case that had a significant effect on death and death and/or near miss ( $p < 0.001$  in both models). For minor events there was no effect of compensation.

In the second stage of the analysis the total number of major and minor events per case was added jointly to the baseline models. Minor events still influenced surgical outcomes even after adjustment for the total number of major events per case and whether or not they were compensated for ( $p < 0.03$  for death;  $p < 0.001$  for death and/or near miss). It was concluded that seemingly trivial problems can accumulate to have negative effects.

### Understanding surgical excellence

The study also identified the behavioural markers which may lead to surgical excellence.<sup>22</sup> Excellence was defined as achieving a standard beyond the norm expected for a particular type of case.

Procedural excellence scores were developed for 16 surgeons. These scores were based on the difference between observed and expected risk measures derived from baseline multivariable logistic regression models. A negative mean difference equalled fewer major and minor events and fewer uncompensated events than expected after adjusting for patient factors. Differences between procedural excellence scores for the best and worst performing surgeons were explained using a framework of behavioural markers of surgical excellence which comprised individual, team, and organisational factors.<sup>22</sup>

### Learning methodological lessons from the human factors and ASO study

There are several limitations with this study. These relate to (1) observer training and competency assessment, (2) lack of inter-rater reliability measures, (3) the sole reliance on observational data collection, and (4) transfer of learning from an expert to a novice observer late in the data collection period.

### Observer training and competency assessment

Observer 1, a PhD level human factors specialist, was trained by watching three cardiac surgeons performing the surgical procedure in two centres, reading surgical textbooks, and shadowing operating theatre staff. Learning was iterative; it also depended on her confidence to ask questions and her ability to assimilate complex medical information. There was no formal competency assessment (in terms of a written or verbal test of her knowledge). Rather, the senior surgeon who carried out the study reviewed the case reports she produced following each ASO and made a judgement about her readiness to collect data from the operating theatre.

Observer 2, a psychology graduate, never developed sufficient understanding of the ASO to make meaningful observations and his data were not used in the final analysis.

Observer 3, a human factors specialist educated to MSc level, was trained by observer 1. She produced a booklet summarising the ASO in layman's terms. This document also described key errors and problems that occurred during each stage of the surgical procedure and operating theatre etiquette (where to stand, when to ask the team questions, etc). Observer 3 was trained by shadowing observer 1 for 2 months. During this time, both observer 1 and observer 3 took notes for the same case and compared them afterwards. This proved a valuable way to train observer 3 as he could validate his observations against those of a more experienced colleague. Observer 3 proved to be a competent observer but, as with observer 1, there was no formal examination of his knowledge. Rather, observer 1 and a senior surgeon decided when he was ready to observe cases independently.

### Lack of inter-rater reliability measurement

Only one observer was present during each case (with the exception of the training period for observer 3 described above). This precluded inter-rater reliability measures to check the consistency of observations between researchers. This is a serious methodological flaw. However, the research team was undertaking a massive logistical task by collecting data from 16 centres; often there was more than one case per day and obtaining a good sample size took precedence over measuring inter-rater reliability. Furthermore, the operating theatre does not easily lend itself to having multiple observers without compromising the theatre team's access to the patient.

### Sole reliance on observational data collection

The research team did not video record ASOs and relied solely on case reports. Other studies have shown that video recording produces reliable information on team performance.<sup>23-24</sup> Video recordings could have validated and supplemented the information collected by the two observers.

### Transfer of learning from an expert to a novice observer

Observer 1 had acquired expertise in observing ASOs by the time observer 3 joined the research team. There were some difficulties in transferring her knowledge to a novice as she had internalised a surgeon's instinct for the ASO. Difficulties were experienced in verbalising implicit knowledge of "what happens next" and why certain recovery strategies were appropriate in particular circumstances.

## DISCUSSION

The preceding discussion has focused exclusively on why we should carry out structured observational research, citing examples from published literature to illustrate what such studies have contributed to our knowledge about adverse events in health care. The review of the human factors and ASO study has identified the types of problems likely to be faced by ethnographic research teams. This discussion

focuses on answering questions pertaining to the "where, how and who" of structured observational research in health care.

### Where is structured observational research viable?

Structured observational research may be more suited to some healthcare domains than others. Whereas the operating theatre provides an environment where clinical tasks have a clear start and end point, the type of elective surgical procedure is usually known beforehand, and there are consistent team roles, A&E departments and ICUs are more challenging. Their unpredictable diverse case mix, larger size, and the greater movement of staff around a wider area while treating the patient can create difficulties for observers.<sup>25</sup>

Some types of healthcare tasks are easier to observe than others—for example, staff using bar coding devices,<sup>26</sup> drug dispensing, drug administration,<sup>2-5</sup> and hand offs between teams. In general, tasks that involve verbal communication or a potentially high frequency of omission and commission errors provide good observational settings.

### How should observers be trained?

Healthcare studies should learn from research previously carried out in anthropology, human factors, and organisational psychology, all of which have developed assessment instruments in an effort to make observational data collection systematic and to standardise analysis.

Observers might benefit from training that involves videotapes of procedures with simultaneous explanations from healthcare professionals. Future studies should also assess observer competence from two perspectives: (1) their domain knowledge (that is, their technical know how about the specialty) and (2) their observational ability (that is, whether they have acquired the key skills to make meaningful observations). The key skills or attributes of an excellent observer remains an unanswered question and is still open to debate.

### Who is the most appropriate observer?

There is an ongoing debate about who is the most appropriate observer—a medical or a non-medical professional? The operating theatre simulator literature presents two viewpoints on observer qualifications. Some studies advocate medical experts as observers, but others support trained non-medical observers. The literature shows that there is not much difference in the assessments of both types of observers except that medical experts are better at assessing content specific attributes,<sup>27-28</sup> while non-medical observers are better at assessing interpersonal factors.<sup>29-30</sup> Research in other industries has shown that researchers who develop good domain knowledge can make consistent and meaningful observations.

In the ASO study, observers 1 and 3 were better at identifying minor events than the operating theatre team. For example, interruptions by colleagues who asked the surgeon questions while he was operating were classed as minor events by the researchers. These were accepted as common practice by the operating theatre team who did not appreciate the increased risk to the patient of frequent task distractions. Medical professionals sometimes do not recognise an event as an error or problem and may also be reluctant to report errors if it makes the team being observed look bad. However, medical professionals are more likely to be accepted by those under observation than non-medical professionals who may be perceived as "outsiders". This was certainly the experience in the human factors and ASO study where it took time to win the trust and confidence of participating operating theatre teams.

The ASO study has shown that creating a good observer requires consideration of factors other than a person's

## Key messages

- Structured observational research has a key role in identifying the types, frequency, and severity of errors and adverse events in health care.
- Important methodological lessons can be learned from structured observational studies that have already been carried out. It is essential that these lessons inform the development of future research methodologies.
- Decisions about who is the most appropriate observer to collect data from healthcare domains should be based on an appraisal of each candidate's personal skills including the ability to win people's trust, to maintain attention for long time periods, as well as their domain knowledge and observational experience.

professional background. It may be equally relevant to consider their interpersonal skills, their ability to reassure healthcare staff afraid of the medicolegal and punitive consequences of the data, to maintain concentration for long periods of time, to keep to the stated objectives, and to cope with the psychological aftermath of witnessing adverse events. We should work towards creating multidisciplinary observational teams which capitalise on the different perspectives contributed by people from diverse specialities.

## CONCLUSIONS

Patient safety experts should value learning methodological lessons from successful ethnographic studies as these can inform the design of future research. While it is important to understand the factors that lead to error and excellence among healthcare teams, it is also necessary to understand the characteristics of an excellent observer and the methodological flaws in studies like the human factors and arterial switch operation project.

## ACKNOWLEDGEMENTS

The human factors and neonatal arterial switch operation study was supported by a research grant from the British Heart Foundation (PG94166).

## REFERENCES

- 1 Anon. *Webster's Dictionary*. Harper Collins Publications, 1998.
- 2 Dean B, Barber N. Validity and reliability of observational methods for studying medication administration errors. *Am J Health Syst Pharm* 2001;**58**:54–9.
- 3 Barker KN, Flynn EA, Pepper GA, et al. Medication errors observed in 36 health care facilities. *Arch Intern Med* 2002;**162**:1897–903.
- 4 Taxis K, Barber N. Ethnographic study of incidence and severity of intravenous drug errors. *BMJ* 2003;**326**:684.
- 5 Wirtz V, Taxis K, Barber ND. An observational study of intravenous medication errors in the United Kingdom and in Germany. *Pharm World Sci* 2003;**25**:104–11.
- 6 Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002;**37**:1553–81.
- 7 Boreham NC, Shea CE, Mackway-Jones K. Clinical risk and collective competence in the hospital emergency department in the UK. *Soc Sci Med* 2000;**51**:83–91.
- 8 Schaefer HG, Helmreich RL, Scheidegger D. Safety in the operating theatre—part 1: interpersonal relationships and team performance. *Curr Anaesth Crit Care* 1995;**6**:48–53.
- 9 Schaefer HG, Helmreich RL. The importance of human factors in the operating room. *Anesthesiology* 1994;**80**:479.
- 10 Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. *Hum Factors* 1995;**37**:20–31.
- 11 Stanhope N, Vincent C, Taylor-Adams SE, et al. Applying human factors methods to clinical risk management in obstetrics. *Br J Obstet Gynaecol* 1997;**104**:1225–32.
- 12 Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. *Crit Care Med* 1995;**23**:294–300.
- 13 Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. *Qual Saf Health Care* 2003;**12**:143–7.
- 14 Bracco D, Favre JB, Bissonnette B, et al. Human errors in a multidisciplinary intensive care unit: a 1-year prospective study. *Intensive Care Med* 2001;**27**:137–45.
- 15 Coiera EW, Jayasuriya RA, Hardy J, et al. Communication loads on clinical staff in the emergency department. *Med J Aust* 2002;**176**:415–8.
- 16 Chisholm CD, Collison EK, Nelson DR, et al. Emergency department workplace interruptions: are emergency physicians “interrupt-driven” and “multitasking”? *Acad Emerg Med* 2000;**7**:1239–43.
- 17 Andrews LB, Stocking C, Krizek T, et al. An alternative strategy for studying adverse events in medical care. *Lancet* 1997;**349**:309–13.
- 18 Helmreich RL, Merritt AC. *Culture at work in aviation and medicine: national, cultural and professional influences*. Aldershot, UK: Ashgate Publishers, 1998.
- 19 Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;**320**:781–5.
- 20 de Leval MR, Carthey J, Wright DJ, et al. Human factors and cardiac surgery: a multi-center study. *J Thorac Cardiovasc Surg* 2000;**119**:661–72.
- 21 Carthey J, de Leval MR, Reason JT. The human factor in cardiac surgery: errors and near misses in a high technology medical domain. *Ann Thorac Surg* 2001;**72**:300–5.
- 22 Carthey J, de Leval MR, Wright DJ, et al. Behavioural markers of surgical excellence. *Safety Sci* 2003;**41**:409–25.
- 23 Xiao Y, MacKenzie C, Orasanu J, et al. Information acquisition from audio-video-data sources: an experimental study on remote diagnosis. The LOTAS Group. *Telemed J* 1999;**5**:139–55.
- 24 Mackenzie CF, Jefferies NJ, Hunter WA, et al. Comparison of self-reporting of deficiencies in airway management with video analyses of actual performance. Level One Trauma Anesthesia Simulation (LOTAS) Group. *Hum Factors* 1996;**38**:623–35.
- 25 Wears RL. Beyond error. *Acad Emerg Med* 2000;**7**:1175–6.
- 26 Patterson ES, Cook RI, Render ML. Improving patient safety by identifying side effects from introducing bar coding in medication administration. *J Am Med Assoc* 2002;**287**:540–53.
- 27 Slagle J, Weinger MB, Dinh MT, et al. Assessment of inter-rater reliability of an established clinical task analysis methodology. *Anesthesiology* 2002;**96**:1129–39.
- 28 Fraind DB, Slagle JM, Tubbessing VA, et al. Re-engineering intravenous drug and fluid administration processes in the operating room. Step one: task analysis of existing processes task analysis of existing processes. *Anesthesiology* 2002;**97**:139–47.
- 29 Schaefer HG, Helmreich RL, Scheidegger D. Safety in the operating theatre—part 1: interpersonal relationships and team performance. *Curr Anaesth Crit Care* 1995;**6**:48–53.
- 30 Schaefer HG, Helmreich RL, Scheidegger D. Human factors and safety in emergency medicine. *Resuscitation* 1994;**28**:221–5.