System resilience and patient safety during a bed crisis in an NHS hospital in England

Mike D Williams
Andi Smart

m.d.a.williams@exeter.ac.uk
University of Exeter Business School
Rennes Drive
Exeter
EX4 4PU

Abstract

A resilience based conceptual model of a four boundary safe working envelop was developed and used in a case study in a NHS hospital to develop theoretical explanations about the actions and decisions of staff in terms of how they deal with competing requirements. A mixed method of data collection was used during a bed crisis. Evidence of a process of normalization relating to capacity problems was found; a focus on what was easily measured in terms of finance, waiting time targets, and staffing predominantly drove operational behaviour. Patient safety may be compromised by the process of normalisation through the theoretical concept of 'drift to danger'.

Keywords: resilience, normalisation, patient safety

Introduction

Hospital managers and clinicians face many competing demands. These include the need to meet patient expectations in terms of effective, safe and timely treatment whilst meeting the organisational expectations, including remaining within budget. The publically funded healthcare system in the UK, through the government, sets hospitals a number of standards and waiting time targets that hospitals have to meet. A conceptual model, derived from resilience engineering, was developed to help explain how staff actions keep hospital patients safe during periods of high levels of patient demand and competing pressures (Williams, 2008). This model was applied within a case study of a UK hospital to develop a systems resilience theory to explain the actions and decisions of managers and clinicians in relation to patient safety during a period of bed capacity constraint and high patient demand. Patient safety in hospitals throughout the world is a major issue with adverse events occurring to over 9% admitted patients (De Vries et al, 2008). Developing theory that will assist in reducing harm to patients during busy periods in hospitals has potential practical impact.

Literature

Resilience engineering is a developing field within the literature (Hollnagel et al, 2006). It seeks to take a systems and proactive perspective on safety. The definition of 'system resilience' in this paper relating to healthcare is the 'work system's ability to buffer, adapt to, absorb and prevent adverse patient outcomes in the face of disruption' (Woods, 2006). The resilience literature includes the idea of

a ‘safe working envelop’ within which a work system seeks to remain to avoid failure. Rasmussen (1997) describes three interacting boundaries to the safe working envelope; the boundary of economic failure, the boundary of unacceptable workload and the safety boundary of unacceptable performance. The model includes a marginal boundary or zone inside those boundaries which if breached would create the conditions for failure.

Cook and Rasmussen (2005), use Rasmussen’s safe working envelope model to examine safety problems for hospitals that are overfull. They suggest that it is normal for healthcare systems to work at the limit of their capacity and for Hirschhorn’s ‘law of stretched systems’ to apply: ‘every system always operates at its capacity. As soon as there is some improvement, some new technology, we stretch it…’. When that capacity is reached, failure or restructuring has to occur (Woods and Wreathrall, 2008). When systems are under resource and or performance pressures the benefits of change are taken in increased productivity or efficiency. Studies have shown an association, but not proven causation, between hospitals that have high occupancy rates and an increase in mortality and adverse events (Richardson, 2006; Sprivulis, et al, 2006; Weissman et al, 2007).

Building on the work by Cook and Rasmussen (2005) and Miller and Xiao (2007), we developed a conceptual model to explain system resilience of NHS hospital systems in the wider political and administrative context by proposing four interacting boundaries of a ‘safe working envelope’ (Williams, 2008). These are the boundaries of financial failure, target failure, unacceptable working conditions, and failure of safety (Figure 1). This model seeks to explain the dynamic pressures that apply to the concept of a safe working envelope for NHS hospitals.

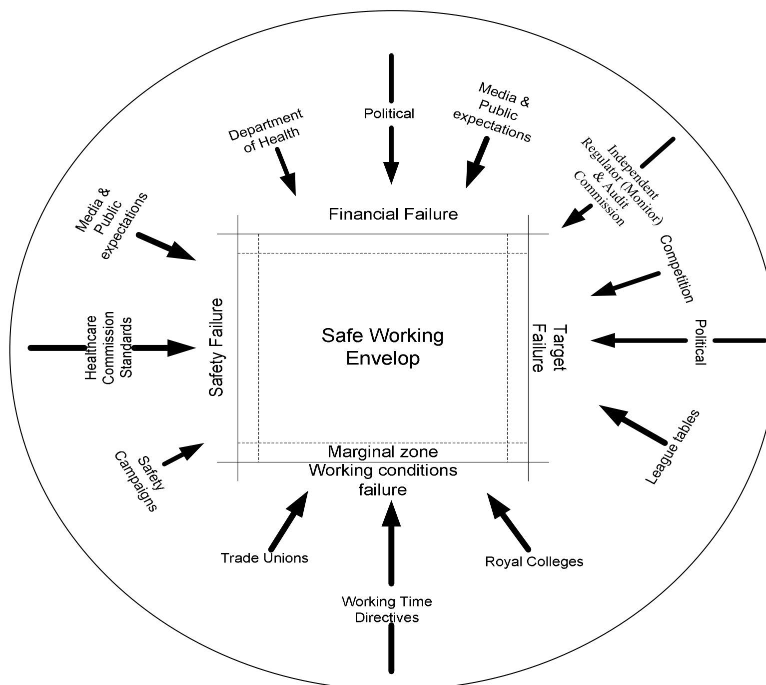

Figure 1: A conceptual model of a ‘safe working envelope’ for an NHS hospital

Vaughan (1996) shows that normalisation of deviance is key issue for safety. Her study of the Challenger Space Shuttle disaster illustrated that the system within which people work can produce a culture where, through small incremental steps, new situations are seen as ‘normal’. This is described as the ‘native view’. Outsiders looking at the situation are more likely to see the situation as deviant, not acceptable and therefore potentially dangerous. Waring et al (2007) point to types of behaviour by health care workers that illustrate ‘taken for granted assumptions about clinical risk…’ Such behaviours ‘normalize risk’ and as such mean that risk of harm to patients is not addressed. Waring (2005) suggests that medical staff regard error as ‘inevitable’ which can lead to errors being seen as ‘normal’. These errors may not be reported as incidents and therefore do not provide the basis for improvements in safety. Normalization in the context of the safe working envelop can be regarded as a ‘drift to danger’ of the operating point. Equally, normalization can be conceived of as shifting the safety failure boundary to a new but more dangerous position.

Method

A case study approach was used in an NHS hospital during an outbreak of a sickness virus. The case study approach was chosen because it is suitable to investigate phenomenon within their context Yin (2003) and where there is complexity in the subject matter (Stuart et al, 2002). It is also used where the boundaries between the phenomenon are unclear, where multiple sources of evidence are needed to converge, and where there is a theoretical background to guide the data collection and analysis (Yin, 2003). Case studies do suffer from the weakness of being context specific and therefore transferability of findings to other situations is problematic. However, when case studies are used to draw out key points of an explanation they can be a powerful contribution (Nightingale et al, 2003).

The case was chosen as an example of a high performing teaching hospital with a stable leadership team where the externally validated track record would suggest an ability to manage both the external and internal competing pressures. A period of intense pressure on bed capacity (a sickness virus that closed wards and increased staff absence) was studied as a means of magnifying the competing pressures, the staff actions in response to those pressures and the implications for patient safety. Full NHS ethical approval was gained.

Within the case study the unit of analysis is the whole hospital as a system. There is an embedded approach (Yin, 2003), using a stratified purposeful approach (Miles and Huberman, 1994) to examine three within case units of analysis to build a picture of the wider hospital. The reason for choosing an embedded approach is to overcome the weakness of a holistic case study design. Yin (2003) argues that a holistic approach may mean the study is conducted at an abstract level as the researcher may not get into the operational detail. One of the reasons behind selecting three embedded areas is to cover both the formal and informal actions that occur, to understand the macro context and then the managers and individual team members (micro) response. The three units were:

Organisational:	Trust Board and hospital wide operational processes
Sub unit:	Division of Medicine
Team:	Ward / Consultant Team / Department members

A mixed method approach using the simultaneous collection of qualitative and quantitative data was used (Creswell, 2003; Bergman, 2008). The underlying pragmatic philosophy is that to obtain a

better understanding of the phenomenon, a range of epistemological perspective were used. Multiple sources of data collection were utilized. This assists in achieving triangulation in data collection, the interpretation and subsequent theory development. A concurrent nested approach to gathering data through quantitative and qualitative means was employed to allow different questions to be addressed during the study. Such an approach allowed different perspectives to be obtained about the hospital being studied and by bringing the data together during the analysis phase. This helped to increase the validity of the findings (Creswell, 2003). The collection of quantitative data was nested within and informed the wider qualitative study (Creswell and Plano Clark, 2007).

Data was collected from a number of sources. Initially there was analysis of both internal and external context setting documents to provide the contextual overview. Hospital administrative data provided the descriptive statistics relating to demand and capacity before and during the sickness virus. During and after the period of high pressure semi-structured interviews of eight doctors, nine nurses, ten managers were conducted, recorded and transcribed. Non-participative observation of staff actions and meetings were undertaken over a four month period. A grounded theory approach was used in the analysis to develop categories and concepts from the different data sources using NVivo (Strauss and Corbin, 1990). Thematic analysis of the concepts was used to interpret the data to identify phenomena about the competing priorities and the failure boundaries (Miles and Huberman, 1994).

Case study hospital

The hospital studied is a 760 bed NHS teaching hospital serving a largely rural population in England. The nearest acute hospital is over twenty miles away. The main targets that impact on hospital bed capacity are (DoH 2007):

- A maximum of waiting time of one month from diagnosis to treatment for all cancers.
- Waiting time in Emergency Department (ED)– maximum of 4 hours for 98% of patients
- Waiting time from Referral To Treatment (RTT) – maximum of 18 weeks with a push to reduce to 15 then 13 weeks (This target relates to non-emergency patients)

Patient routes into the hospital

Where cancer is suspected then the referring General Practitioner (GP) uses a rapid access route to ensure that the patient has an outpatient appointment within 14 days and then, where necessary, treatment commences within 31 days.

Emergency patients enter the hospital through the ED, or if they have been seen by a GP prior to attending hospital, they will be sent directly to the surgical assessment unit (SAU) or emergency medical unit (EMU). Ninety eight percent of those patients arriving at the ED must be treated and admitted or discharged from ED within four hours. There are no targets relating to patients sent directly to the SAU or EMU.

Non-emergency and non cancer patients are referred by their GP to a consultant in the hospital. The usual pathway is for a consultation in an outpatient clinic either before or after some diagnostic tests. Once a decision on treatment is made, the patient may be admitted as a day case or inpatient for treatment. Ninety percent of such patients being admitted have to have their treatment started within the RTT target.

The three target areas can create competing dynamics within a hospital service. For the ED to succeed on admitting patients within 4 hours depends on other parts of the hospital having the capacity at the time required. However, at the same time the SAU/EMU receives patients directly from GPs. Once these patients have been assessed, they need to be accommodated in suitable specialty wards (Proudlove *et al*, 2003). Those patients on the RTT pathway also have to be admitted within the target of 18 weeks and also consume bed capacity.

The high pressure event

The case study focused on before, during and after a specific incident in the hospital. An airborne sickness virus in one ward that then spread to other patients and staff. Within a short time the virus was present in eight wards (around 30% of total bed capacity). The Control of Infection Team procedures for such an outbreak means the affected wards were closed to new admissions. The reduction in bed capacity continued for the following two weeks with ward-based capacity being reintroduced following sterilization and confirmation of the elimination of the virus.

Findings

Identifying the four boundaries of the safe working envelop

Analysis of the performance management documentation and interviews showed that both the financial and target boundaries were easily measured, monitored and strictly performance managed. The financial performance was reviewed at least monthly in substantial detail. The four hour target was monitored continuously at busy times and other targets at least weekly. Staff numbers were linked to budgets and monitored monthly for turnover and sickness rate. Schedules of expected numbers of doctors and nurses per ward per shift were in place and staffing was monitored informally each day. The measurement of patient infections was undertaken daily and regarded as the primary patient safety issue. ‘Patient falls’ and incident reports were formally monitored retrospectively each month. Within certain areas senior nurses reviewed incidents of patient falls within a week. Apart from infections and number of serious incidents no other patient safety measures where reported to the Hospital Board on a routine basis. Financial and target performance were reported to the Board in detail each month. Staff turnover and sickness rates were reported every quarter.

Pressure on the ‘operating point’

Theoretically ‘operating point’ is the location of the system within the safe working envelop. Dynamic pressures influence the location of the operating point and can push the system towards a boundary of failure. The case study hospital was in a strong financial position with a projected large surplus for the year. The interviews confirmed that staff did not feel pressure to make decisions that might compromise targets, staffing or patient safety due to budget considerations.

The combination of the four hour ED, Cancer and RTT targets did create substantial pressure. Any NHS hospital has to be able to manage peaks in emergency demand whilst at the same time maintain the pattern of non-emergency (elective) admissions from both the cancer and RTT pathways. Bagust et al (1999) showed that for a hospital to be in a position to manage peaks in demand the bed capacity should not exceed 85% occupied.

The hospital managers at the outbreak of the sickness virus had taken the decision that it was not feasible to divert emergency patients. Therefore, the remaining potential action available was to cancel non-emergency admissions to help relieve pressure. From the interviews there appeared to be a strong view that non-emergency patients are equally deserving of admission.

> "The non-emergency patient has a problem that needs treatment. …From the patient's perspective a bed crisis is not their problem – they just want to be admitted." (Nurse)

As well as the patient perspective from observations and interviews it was clear that to cancel an elective patient on the RTT pathway created a capacity problem in trying to rebook the patient within the target period. Also there was a government target of the percentage (<1%) of patients that could be cancelled on the day of admission for non-medical reasons. The competing dynamics led to a process of tight coupling where operational problems in one part of the hospital impacted adversely on other areas (Cook and Rasmussen, 2005). This generated considerable pressure on staff to find innovative ways to maintain the flow of patients through the system by seeking to accelerate patient discharges.

Data analysis showed there was no significant change in the level of elective admissions during the bed capacity crisis (Figure 2) compared to the same period the previous year. Without a change the level of non-emergency admissions, other actions relating to capacity, particularly maintaining the flow of patients, became the focus for attention.

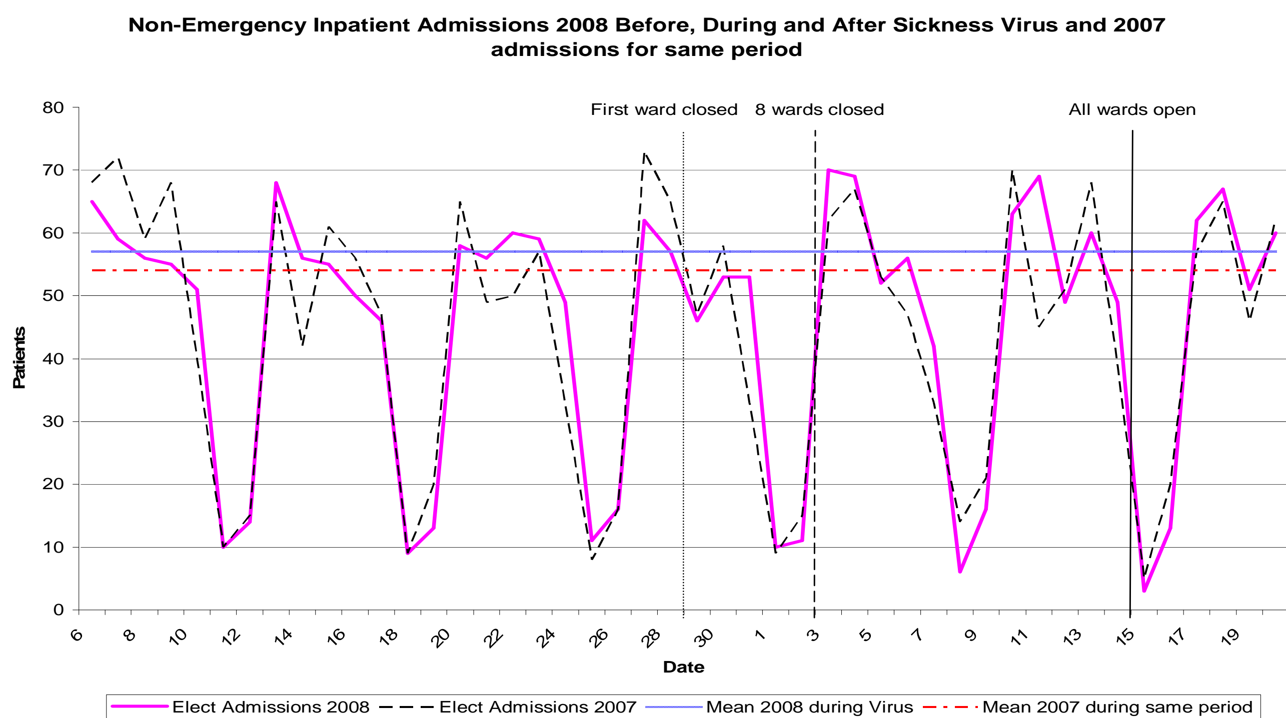

Fig 2 Elective Admissions 2008 Before, During and After Sickness Virus and 2007 over same period

A number of actions where taken to generate new staffed bed capacity (overflow capacity). This included opening an area normally used as a preadmission area (normally only open during the working day) in orthopedics. This small 13 bed ward is situated at the other end of the hospital from the medical wards. Patients already in medical beds, identified as being well enough, were then moved to this area. A further area of 6 beds was brought into use in an area close to the ED for medical patients. Areas within wards that were normally used as day case beds were brought into service as inpatient beds.

Staffing the additional areas combined with the short-term sickness of doctors and nurses due to the virus pushed the operating point towards the staffing boundary and an inadequate level of staff in ward areas. Concerns where expressed about staff having to move to cover sickness and open additional beds. Staff where unfamiliar with certain types of patients and ward areas which lacked key equipment.

Infection control measures were applied rigorously. Wards with the sickness virus remained closed until all the patients and staff were symptom free for forty eight hours and then the areas thoroughly cleaned. Emergency patients admitted with an infection where placed in a side room. These actions maintained the patient safety boundary in respect of infection but then created pressure on the remaining bed and staff capacity. That pressure on the remaining bed capacity meant that medical patients were transferred onto surgical wards (medical outliers). By moving patients into surgical beds it provided the capacity within the medical wards to accept transfers from EMU. In turn then EMU can accept admissions from the ED. With the 4 hour target there has to be a constant supply of beds being found in the hospital. During the sickness virus period both the EMU and sometimes the SAU would not have any empty beds. Emergency patients requiring admission after seeing their GP, were then diverted to the ED. This created a reinforcing loop (Sterman, 2000) by creating more patients to be processed within the four hour target which meant the need to create more empty beds in the hospital within a short time frame. As the majority of emergency admissions were medical patients, current medical patients were moved to surgical wards (medical outliers) in increasing numbers to create the bed capacity in the 'right place'.

> "Sometimes you have two or three bed moves to create a bed on EMU to get an ED patient in. …The medical patients that move are not always suitable." (Nurse)

As suggested by Wheeler (2006), a moderate but sustained effect on the number of medical outliers was observed during the virus period (Figure 3).

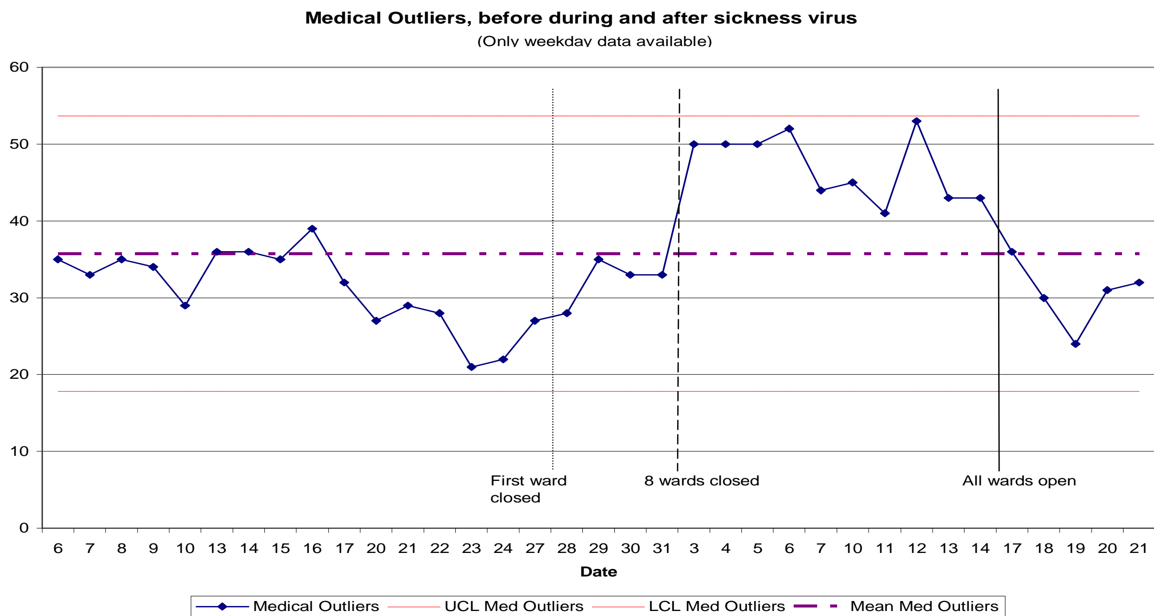

Fig 3 Statistical Process Control Chart - Medical outliers before, during and after sickness virus

Interviews and observation confirmed that bed capacity and flow problems were common and regarded as 'normal'. There was evidence of admitting and preparing non-emergency patients for procedures in corridors and waiting rooms. NHS hospitals count the bed occupancy at midnight. In 2008 the midnight occupancy for all wards in the case study hospital was 89%. The medical wards occupancy ranged from 90-98%. The midnight census underestimates the bed occupancy during peak period of activity during the day. Occupancy for many wards during the day was above 100%. Staff had learnt to accept the pressure on the system and had found innovative ways to manage the situation.

Discussion

From the literature on safety theory and stretched systems (Cook and Rasmussen, 2005; Miller and Xiao 2007; Rasmussen, 1997; Reason, 1997; Sheridan, 2008; Woods and Cook, 2002, Woods, 2006), we can conceptualize that the operating point would be close to the safety boundary for a hospital experiencing capacity and production pressures. In the case study only three out twenty seven staff interviewed believed the operating point was within or near the marginal zone. Most regarded the hospital as operating in a way that prioritized the safety of patients. Previous (albeit limited) studies suggest that hospitals that have high occupancy rates, experience an increase in mortality and adverse events (Richardson, 2006; Sprivulis, *et al*, 2006; Weissman *et al*, 2007). It is therefore surprising that more staff did not regard the hospital as being unsafe during this period. It can be argued that the hospital system experienced a period of 'drift to danger' (Rasmussen, 1997).

There was evidence that the buffer capacity and flexibility required for a resilient response was present only on the financial failure boundary. The admission process in terms of beds and staff became brittle (no flexible capacity) and tightly coupled (Woods, 2006; Cook and Rasmussen, 2005).

Role of culture

Vaughan (1996) describes how the decision to launch the ill fated shuttle, Challenger, was as much about culture as it was about technical systems going wrong. Weick and Sutcliffe (2003) point to the importance of a safety culture, defining culture as 'what we expect around here'. There was considerable evidence from the observation and interview data of staff normalizing to the pattern of actions that maintained the level of admissions. They appeared to make sense of the situation by placing importance on not canceling patients who needed non-emergency admissions. We can theorise that the reality was more likely the system pressure on staff to achieve multiple targets. As Vaughan (1996) suggests: 'We reconstruct history every day, not to fool others but to fool ourselves, because it is integral to the process of going on.'

In the NHS waiting times and measurable infection control targets have been given a high priority backed by a strong performance management ethos. This contributes to a culture of finding all possible ways to meet the targets. The wider NHS culture also allowed a process of normalizing by staff to levels of risk previous thought to be unacceptable. Safety being a dynamic non-event (Reason, 1997), and not as clearly defined or measured as finance and waiting time targets has received far less management attention. The second aspect is the culture of production. Staff in the hospital had normalized to being extremely busy to compensate for the constant pressure of competing demands and the need to find beds for admissions. The focus on maintaining flow of patients through the and out of the hospital has become hard wired into the system in terms of staff attitudes, processes and structures.

Conclusion

The conceptual model of a safe working envelop for an NHS hospital has provided the framework through which to explore the compensating actions staff take in the face of competing pressures. The research has led to the articulation that the system pressures and failure boundaries influence the attitudes of staff and consequently the culture which drives the actions. This conclusion may relate to what Weick and Sutcliffe (2003) describe a process of cultural entrapment – '…the process by which people get locked into lines of action, subsequently justify those lines of action, and search for confirmation that they are doing what they should be doing.'

The need for the safe working envelop model to include the wider system is needed to conceptualize the impact of externally generated pressures both on the production of culture and the culture of production. To extend this approach, to explain the impact of conflicting pressures on patient safety, further research incorporating additional cases is needed. A single case study has limitations. The situation of an NHS hospital facing the pressures described is particular to England. With those limitations in mind the development of the conceptual model to a particular system wide context, does facilitate the articulation of compensating actions as a response to competing pressures.

From a practitioner perspective the model helps illustrate the dynamic nature of the competing pressures and the need to set clearer boundary measures in the area of staff workload and patient safety. Policy makers, performance managers and regulators may be able to appreciate better the impact their decisions have on the actions of front line staff in hospitals and the potential impact this, in turn, has on patient safety.

References

Bagust, A., Place, M., and Posnett J.W. (1999), Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. *BMJ,* Vol. 319**,** pp.155-8.

Bergman M.M. Ed (2008), *Advances in Mixed Methods Research*, Sage, London.

Cook, RI, Rasmussen J. (2005), "Going solid": a model of system dynamics and consequences for patient safety. *Qual Saf Health Care*, 14(2): 130-134.

Creswell J.W. (2003), *Research Design* (2nd ed) Sage, London

Creswell J.W, and Plano Clark V.L. (2007), *Designing and Conducting Mixed Methods Research*, Sage, London.

Department of Health. (2007), *Operating Framework 2008-09for the NHS in England* available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081094 (accessed 24 November 2008)

De Vries E., Ramrattan M., Smorenburg S., Gouma D., and Boermeester M. (2008), The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care,* 17: 216-223.

Hollnegal E, Woods D, Leveson N. (eds) (2006), *Resilience Engineering: concepts and precepts*, Aldershot, Ashgate.

Miles M.B, and Hubermann A.M. (1994), *Qualitative Data Analysis* (2nd ed), Sage, Thousand Oaks, CA.

Miller A, Xiao Y. (2007), Multi-level strategies to achieve resilience for an organisation operating at capacity: a case study at a trauma centre. *Cogn Tech Work,* 9: 51-66

Nightingale, P., Brady, T., Davies A. and Hall, J.(2003), Capacity utilization revisited: software, control and the growth of large technical systems. *Industrial and Corporate Change,* 12(3):477-517.

Proudlove NC, Gordon K, Boaden R. (2003), Can good bed management solve the overcrowding in accident and emergency departments? *Emerg Medicine Journal,* 20:149-155.

Rasmussen J. (1997), Risk Management in a dynamic society: a modelling problem, *Saf Sci,* 27(2/3) 183-213

Reason J. (1997), *Managing the Risks of Organizational Accidents,* Ashgate, Aldershot.

Richardson, D. B. (2006), Increase in patient mortality at 10 days associated with emergency department overcrowding. *Medical Journal of Australia,* 184(5):213-216

Sheridan T.B. (2008), Risk, Human Error, and System Resilience: Fundamental Ideas, *Human Factors,* 50**,** 418-426.

Sprivulis, P. C., Da Silva J-A, Jacobs IG. et al. (2006), The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Medical Journal of Australia,* 184(5): 208-212

Sterman J.D. (2000), *Business Dynamics*, Irwin McGraw-Hill, Boston.

Strauss A, and Corbin J.M. (1990), *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage, Newbury Park, California.

Stuart, I., Mccutcheon, D., Handfield, R., Mclachlin, R. and Samson, D. (2002), Effective case research in operations management - a process perspective. *Journal of Operations Management,* 20 (5): 419-33.

Vaughan D. (1996), *The Challenger Launch Decision*, Chicago University Press, Chicago.

Waring, J.J. (2005), Beyond blame: cultural barriers to medical incident reporting. *Social Science & Medicine,* 60**,** 1927-1935.

Waring J, Harrison S, McDonald R. (2007), A culture of safety or coping? Ritualistic behaviours in the operating theatre. *Journal of Health Services Research and Policy,* 12 (Suppl 1):3-9.

Weick K, Sutcliffe K. (2003), Hospitals as Cultures of Entrapment: A Re-Analysis of the Bristol Royal Infirmary. *California Management Review,* 48 (2):73-84.

Weissmann J, et al (2007), Hospital Workload and Adverse Events, *Medical Care,* 45(5) 448-455

Wheeler D.J. (2003), *Making Sense of Data*, SPC Press, Knoxville, Tennessee.

Woods DD (2006), Essential Characteristics of Resilience. In: Hollnegal E, Woods D, Leveson N. (eds) *Resilience Engineering: concepts and precepts*. Aldershot, Ashgate pp.21-34.

Woods D.D, Cook RI. (2002), Nine Steps to Move Forward from Error, *Cognition, Tech & Work,* 4:137-144.

Woods D.D, and Wreathrall J. (2008), Stress-strain Plots as a Basis for Assessing System Resilience. In Hollnegal E, Nemeth C, Dekker S. (eds) *Remaining Sensitive to the Possibility of Failure*. Aldershot, Ashgate, pp.143-158.

Williams M.D. (2008), Designing System Resilience in NHS Hospitals. *Proceedings of the 15th EUROMA Conference*, EUROMA, Groningen.

Yin R.K. (2003), *Case Study Research, Design and Methods*, Sage, London.