Risk&Regulation

Magazine of the ESRC Centre for Analysis of Risk and Regulation

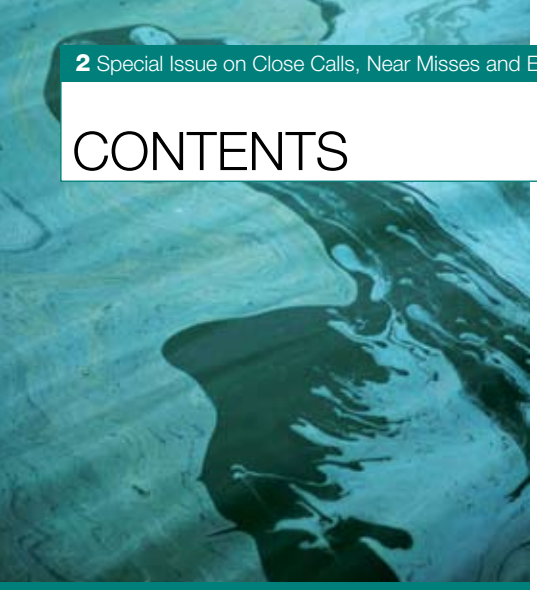
July 2010

Special Issue on Close Calls, Near Misses and Early Warnings

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

LSE

CONTENTS



RISK&REGULATION: CARR REVIEW SPECIAL ISSUE ON CLOSE CALLS, NEAR MISSES AND EARLY WARNINGS

EDITORS: John Downer and Michael Power

ENQUIRIES: Centre Administrator, ESRC Centre for Analysis of Risk and Regulation, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE United Kingdom

Tel: +44 (0)20 7955 6577
Fax: +44 (0)20 7955 6578
Website: www.lse.ac.uk/collections/CARR/
Email: risk@lse.ac.uk

PUBLISHED BY:
ESRC Centre for Analysis of Risk and Regulation, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE

DESIGN AND ART DIRECTION: LSE Design Unit
PHOTOGRAPHY: iStockphoto (p4), Dreamstime.com (p3, p5, p7, p9, p10, p12, p15, p16, p19)
PRINTED BY: Aquatint BSC
COVER IMAGE: Dreamstime.com

ISSN 1473-6004
Online ISSN: 1473-6012

carr
centre for analysis of risk and regulation

Risk&Regulation is also published on CARR's website.
www.lse.ac.uk/collections/CARR/publications/magazine.htm

www.lse.ac.uk/collections/CARR/

EDITORIAL

Close Calls

At a time when a number of high impact risks have crystallized, for example within the financial system and in the field of deep-sea drilling for oil, it might seem that the theme of close calls and early warnings somewhat misses the point. Yet nothing could be further from the truth. Analyses of the financial crisis are beginning to reveal the manner in which warning signs were ignored or misinterpreted. And it is likely that a future investigation into the BP oil rig disaster in the gulf of Mexico will suggest something similar. In short, how individuals, organizations and states deal with often weak and complex signals of possible disaster is perhaps the core problem of risk management practice and theory.

The collection of essays in this special issue of *Risk&Regulation* emerged from the Close Calls conference held at CARR in March 2009. All the contributions clearly show that, while there is an aspiration for an 'early warning' culture in a variety of fields, their effective realization depends on many variables, not least human psychology and the well-documented problem of optimism bias. Yet the collective tendency (so-called groupthink) to ignore warnings, to underestimate the significance of incidents and to under-report, is offset by the presence of blame and accountability environments which encourage frequent but often less meaningful responses. How organizations process and learn from incidents is therefore an important question in many of the contributions.

Another theme running through the papers is the double role of technology, both as a source of risk and also as a basis for mediating information and developing 'intelligent tools' relevant to human responses to risk. Information mediating technologies pose fundamental epistemological questions about our capacity to know the risks of failure in live, as compared with experimental, settings. They also raise operational issues about how to encode sensible trigger points for attention and intervention. As is well-known in the case of air traffic control, information technologies embody definitions of 'error' and 'near miss' but this is also common in other environments where conservative levels of operational compliance are required. More generally, material risk management infrastructures necessarily embody choices and values which may be not well understood.

In addition to technology, institutional factors shape where and how critical event or near miss information is assembled and used. Emerging problems may be invisible because critical data does not exist or is collated in the 'wrong' place. As one of the following essays suggests, undertakers can have a more timely awareness of unusual patterns in mortality than government departments. So the world of near misses and warnings is often a very local, bottom-up and specific one, whereas the policy domain charged with remediation after the event is highly public.

The core policy question is whether organizations and individuals can improve their capacity to see crises and disasters coming. Can corporations and states get smarter at reading and acting upon the warning signs? What are the dangers of reacting to every 'weak signal' and every whistleblower? It is clear that there is no 'technological fix' available to answer these questions since they also speak to the values implicit in specific risk tolerances and appetites. In theory, tolerance values are specified *ex ante*, designed into technological infrastructures, and provide a basis for investigations when they are breached. Yet, in practice, many close calls and warnings may fall below or outside formal tolerance levels. Indeed, organizations which regard themselves as being committed to high reliability are often surprised in the aftermath of failure. Beyond the mechanics of key risk indicators, it is clear that there must be a moral climate in organizations which values curiosity about anomalous events.

In 2010 the risk management agenda is rather bruised. It feels as if a range of existing tools and techniques provided only an illusion of control and did little to assist in the intelligent anticipation of risks based on a capacity, both moral and technical, to listen to weak signals. The problem is one of designing practices in the space between a known past and a necessarily uncertain future. In addition, we are always likely to be disappointed in our foresight capacity and there will be a tendency to blame in retrospect when it appears that warnings were unheeded. But the contributions which follow also suggest that organizations can do a better job of finding smart ways of attending to, and acting upon, the incidents, anomalies and errors of the present.

Michael Power is Professor of Accounting, LSE, and Research Theme Director of CARR.

Why BP Ignored Close Calls at Texas City

Andrew Hopkins

In March 2005 a massive explosion at a BP-owned oil refinery in Texas killed 15 people and injured nearly 200. The explosion cost BP dearly, both in reputation and financially. It is one of the landmark accidents of the petroleum industry in recent decades, and the reports which it generated have been much studied by companies in many industries around the world. The explosion was caused by the escape of the equivalent of a road tanker load of petrol from a particular process unit, that is, from a piece of equipment used to process or refine a petroleum product. Such failures are described in the industry as *process safety* incidents, to distinguish them from more conventional safety incidents that are often referred to, in short hand fashion, as 'slips, trips and falls'.

What happened on this occasion was that plant operators who were filling a column lost situational awareness. There were no automatic cutouts on the column and operators filled the column until it overflowed. The escaping petrol was released from the top of a tower and fell to ground where it accumulated as a large vapour cloud. The vapour cloud eventually came into contact with an ignition source and exploded.

There were various warnings leading up to this failure, including a series of near miss events – close calls. On six previous occasions in the preceding ten years there had been releases from the same tower, producing vapour clouds at or near ground level. In two cases, operators thought the event serious enough to shut down the unit. On four occasions the matter was serious enough to call fire fighting staff who responded in some cases by spraying the vapour cloud with fog, to reduce the risk of ignition.

Each of these events could have resulted in an explosion, had there been an ignition source in the area, and they were undoubtedly close calls. Each was investigated but none of these investigations questioned the basic design or system of operation that was leading to these events. Perhaps most tellingly, information about these close calls was not readily available to investigators after the 2005 explosion. They found it necessary to sift through a variety of sources in order to reconstruct this history. It was not a history that was recognized at Texas City; it was not a history from which anything was learnt.

Part of the reason for this failure to learn from these close calls was the way safety was managed at the Texas City Refinery. Safety efforts were driven by workforce injury statistics. There is a problem with this focus. Most workforce injuries are a result of slips, trips and falls. On the other hand, although process safety accidents may be catastrophic when they occur, they are nevertheless rare events, and therefore do not contribute to injury statistics on an annual basis. A focus on injury statistics therefore leads to a focus on the 'trip' hazards, and a tendency to become complacent with respect to process safety hazards. This is basically what had happened at Texas City, and this is why there was no systematic or effective response to the process safety close calls that had occurred at the site.

The failure of Texas City management in this respect in surprising, given that this particular lesson has emerged from several major accident inquiries in process industries in recent years. I shall argue here that the explanation for this apparent blindness lies in the structure and functioning of BP as an organization.

First, consider the remuneration systems in place for senior managers. They were subject to individually constructed performance agreements that emphasized financial performance of the business unit. Safety was included in these agreements but the indicators of safety were workforce injury statistics, with no indicators that related to how well process safety was being managed. The Texas City site experienced hundreds of unintended gas releases and fires each year and this information could easily have been used to construct process safety measures for inclusion in performance agreements. But this was not done, and the result was that the incentive system diverted management attention from process safety hazards.

It is particularly important that process safety incentives be built into the remuneration packages of finance and personnel managers. These people often take the view that they have no role to play in safety, yet major accident investigations routinely conclude that staffing and financial cutbacks were among the root causes. At BP there were no financial incentives provided to personnel and finance managers to pay attention to the safety implications of staffing and financial cutbacks.

The second aspect of BP's operations that contributed to the blindness to warning signs was the relentless cost cutting to which sites like Texas City were subject. The site was generating in the order of 100 million dollars profit annually, but this delivered a rate of return on capital of less than 10 per cent, which was generally seen as inadequate. The strategy adopted was to decrease

the expenditure on maintenance and the like, in order to increase the return on capital.

One way of preventing this accident would have been to install automatic cutout mechanisms on the column, in accordance with recognized best practice, but this would have required capital expenditure that was not available at Texas City. Another way would have been to convert the release tower into a flare that would have ignited the escaping material at the point of release, thereby preventing it from accumulating as a vapour cloud that could subsequently explode. But again, this was an expenditure that Texas City consciously chose not to make. In these circumstances, there was not a lot of point in attending to close calls because the company was not willing to make the capital investments that were necessary to respond effectively to them.

To top it off, BP ordered a 25 per cent cut in operating cost six years before the accident. This is a massive cut, as a moment's reflection makes clear. It could only be achieved by cutting maintenance, cutting training, and cutting back on safety staff, all cuts Texas City felt compelled to make. The result of these cuts was that Texas City suffered from a kind of organizational paralysis that prevented it from dealing effectively with close calls.

A third factor that contributed to the organizational blindness to warning signs was the very structure of the BP organization, in particular the organizational location of its safety experts. There were safety experts at very senior positions in the corporation located in London. A vice president for health, safety and environment and a vice president for technology answered to a chief executive for 'functions' who in turn answered directly to the CEO. These people in 'functional' positions were responsible, among other things, for creating the safety standards according to which the corporation was supposed to operate, but they were not responsible for enforcing the standards; compliance with standards was the responsibility of line management and more specifically site managers. This was a clearly articulated company philosophy. What it meant was that the safety experts at head office had no authority over site managers when it came to safety. For instance, they were not in a position to demand that money be spent on safety improvements or to veto operations that did not comply with safety requirements. Had the Texas City site manager been accountable to these senior safety officers for compliance with process safety standards, as well as being accountable to company line management for financial performance, then it is likely that expenditure in relation to process safety would have received a higher priority and close calls would have been responded to more effectively.

It is interesting that the inquiry into the space shuttle Columbia accident came to a very similar conclusion. It argued for the creation of a technical regulatory authority within NASA that would have the capacity to override line managers and to veto launch decisions if there were significant safety breaches or irregularities.

There were safety specialists located in other parts of the BP organizational structure as well as at the top. At Texas City there was a dedicated process safety manager. He was well aware of the need to give process safety a higher priority and well aware of the significance of close calls. Indeed this man had co-authored an article on a previous BP accident at Grangemouth, in Scotland, which drew lessons about the need for a special focus on process safety. But he had very little organizational clout at Texas City. He did not report to the site manager and had relatively little access to that individual. Before the accident he had urged that process safety should have a 'seat at the management table', but his urgings had been to no avail. Furthermore, his process safety unit at Texas City was poorly resourced and quite incapable of carrying out all the tasks for which it was responsible.

It is clear then that safety specialists within the BP organization were systematically disempowered by their position and the function. BP's very organizational structures undermined the capacity of the organization to deal properly with process safety issues and, in particular, prevented it from recognizing and responding effectively to close calls.

A final factor in this story is the failure of leadership at the very highest level. The CEO was perceived by those around as unreceptive to bad news about safety. Consequently he was never informed about the deleterious impact of cost cutting at the Texas City site. Moreover, he did not understand the distinction between process safety and other forms of safety and he assumed that injury statistics provided an adequate measure of how well process safety was being managed. The fact that the injury rate at Texas City was low would have given him an unwarranted sense of confidence.

What this analysis makes clear is that in seeking to understand the adequacy of an organization's response to close calls or warning signs of any sort, it is useful to examine the structure and functioning of the organization as a whole. It is here that we are likely to find the root causes of failures to recognize and respond appropriately. In the case of the BP Texas City accident these organizational root causes include the following: an inappropriately focused remuneration system; cost cutting without regard to safety consequences; an organizational structure that disempowered safety experts; and a senior leadership that discouraged bad news and failed to understand the distinctive nature of process safety.

Andrew Hopkins is Professor of Sociology, Australian National University. For more information see Andrew Hopkins, *Failure to Learn: The BP Texas City Refinery Disaster.* (CCH, Sydney, 2008).

On Unforeseeable Failure

John Downer

Why do technological disasters happen? Until the late 1970s, this question belonged exclusively to engineers. By the 1980s, however, social scientists had begun to recognize that such accidents had social and organizational dimensions. The British sociologist Barry Turner investigated eighty-five 'man-made disasters' and noted that they were invariably preceded by a legacy of unheeded warnings that, if acted on, would have averted a misadventure. Turner's realization, that experts were overlooking clear and intelligible danger signals, created a space for a sociology of technological disaster, because it allowed sociologists to 'black-box' the engineering-level causes of disasters and recast them as 'social' rather than 'engineering' failings.

Black-boxing engineering-level explanations in this way has been incredibly fruitful, but, as we will see, there are good reasons to believe that social scientists are missing something important by disregarding the expert knowledge behind accidents. A few sociologists, most notably Charles Perrow, have argued that by returning to engineering-level explanations we can derive important insights about the technological world and our collective relationship to it.

This article will add to these arguments. It will argue that if we are looking at engineering-level explanations of technological accidents, and view them through the prism of the sociology of scientific knowledge, then new and important insights are revealed.

Engineering explanations

What are the social scientists ignoring when they 'black-box' the engineering-level causes of technological accidents? In many cases it is nothing. The black-box is literally empty, in that some technological accidents require no engineering explanation. This can be for different reasons.

Sometimes accidents require no engineering explanation because they result from human errors. 'Human error' is a constructed and contested category, which many sociologists argue has less explanatory power than observers often accord it. Yet such errors certainly exist, and some are unambiguous. When *Aeroflot Flight 593*, a passenger-laden Airbus A-310, abruptly nose-dived into the Siberian tundra on March 22, 1994, the final words on the black-box voice-recorder were those of an alarmed pilot extricating his 15-year-old son, Eldar, from the pilot seat. The accident report said little that tarnished Airbus's design.

Similarly, some accidents require no engineering explanation because they result from procedural or organizational errors. On July 1, 2002, for instance, *Bashkirian Airlines Flight 2937*, a Russian airliner laden with schoolchildren en route to Barcelona, collided with DHL Flight 611, over Überlingen in Germany, and, although there were various dimensions to this disaster, it was at least partly attributed to a procedural conflict. Both aircraft were given conflicting orders to 'climb' or 'descend' by the air traffic controller, on one side, and their automatic Traffic Collision Avoidance Systems (TCAS), on the other. Fatally, the Russians had been trained to prioritize the controller's instructions over those issued by TCAS, whilst the Americans had been trained to do the reverse. The result was that both aircraft descended into each other.

Perrovian Accidents

Technologies undoubtedly do malfunction, however, and therefore clearly many accidents require engineering-level explanations. The sociologist who has most systematically studied accidents on this level is Charles Perrow. Looking at the technological anatomy of technological failures, Perrow divided them into what he called 'Component-Failure Accidents' and 'Normal Accidents' (or 'System Accidents').

Normal Accidents, by Perrow's definition, are the product of random confluences of otherwise non-critical events. No complex technological system can operate in a vacuum, he suggests, since engineers cannot control every environmental factor or anticipate every relationship between elements. From this premise, he argues that seemingly unrelated and trivial events can sometimes interact in unexpected ways that thwart the very best engineering predictions and cause complete system failures, or Normal Accidents. Perrow suggests that the 1979 near-catastrophe at Three Mile Island is exemplary. By his account, the incident began when leaking moisture from a blocked water filter inadvertently tripped valves controlling the flow of cold water into the plant's cooling system. Redundant backup valves, that should have intervened, were also inexplicably closed, which should have been clear from a (rarely-needed) indicator-light in the control room, if it had not been obscured by a tag hanging from a switch above. The final line of technological defence was a tertiary system: a relief valve which should have opened but did not, while a malfunctioning indicator-light erroneously indicated that it did. Complexity, here, colluded with cruel coincidence, and the reactor's controllers understandably struggled to comprehend its condition in time to prevent a catastrophic meltdown.

The second element of Perrow's taxonomy of engineering explanations, 'Component-Failure Accidents', are essentially all technological failures that are not 'Normal' i.e. those that result from linear and predictable relationships between components within a system and from points of failure that are known to be critical. There are no one-in-a-billion coincidences here, no unanticipated interactions. Perrow argues that the majority of accidents caused by engineering failures fit into this category.

The distinction between 'Normal' and 'Component-Failure' accidents is important to the sociology of disaster (and, hence, 'engineering-level' explanations are important) because the former, critically, are fundamentally unavoidable. There are no Turnerian foresight failures in Normal Accidents. They are one-in-a-billion coincidences: perfect storms of small events. This 'one-in-a-billion' quality makes them impossible to predict. The factors that lead to them have no predictive value because none are particularly noteworthy in themselves. Investigators can note significant factors in retrospect, as Turner notes, but through Perrow's lens, their significance is an artifact of hindsight. There is literally no way of knowing, in advance, that a blocked valve will be any more critical than the millions of other minor deviances that inevitably characterize any technical system. By foregrounding the technology itself, therefore, Perrow is able to delimit a group of accidents that cannot be solved with organizational insights: a category of unforeseeable, unpreventable failures.

But what of the 'Component-Failure Accidents'? Perrow's analysis offers no argument that these are unpredictable and unpreventable, so the sociological view of them as 'failures of foresight' reigns. For the most part they are construed as accidents for which actors, or social structures, can be held accountable: accidents that might, in principle, have been avoided. This is often an error.

The Epistemology of Failure

Perrow's Component-Failure Accidents often reveal flaws in engineering assumptions, models, or data. (A bridge collapses because it experiences forces beyond those its designers anticipated, for instance.) Such accidents are often construed as errors because engineers are thought to work in an empirical realm of measurable facts. Facts are knowable. Facts are binary. True or false, ontologically distinct. So when the facts are wrong, this wrongness can be viewed as a methodological, organizational – even moral – failing; one that proper engineering discipline should have avoided and one that sociologists might one day prevent.

This picture of facts is rooted in what the sociologist Harry Collins has called the 'canonical rational-philosophical model' of expert knowledge, which is intuitive, convincing, and entirely rejected by modern epistemologists. The idea that erroneous knowledge-claims are fundamentally and recognizably distinct from those that are true (such that perfect tests and calculations should be able to eliminate errors) was a cornerstone of Western philosophy for centuries. Logical philosophers – from Francis Bacon, through William Whewell to Karl Popper and beyond – worked continuously to hone the 'scientific method' in an effort to ensure it led ineluctably towards truth: fiercely debating the nature of proof, the foundations of evidence, and the essence of facts.

Their pursuit ultimately proved futile, however. The several hundred years separating Bacon from Popper speak to the elusiveness of the 'ideal experiment', 'perfect proof' or 'indubitable fact'. Indeed, beginning with Wittgenstein's later work, logical philosophers began to reject the entire enterprise. Together with some historians of science, they started invoking intractable logical paradoxes to argue that no facts – even scientific or technological – are, or could be, completely and unambiguously determined by logic or experiment. Today, few epistemologists doubt that the canonical rational-philosophical model of scientific knowledge is inadequate for explaining science in action and the 'facts' it produces.

It is a very timid sociology, therefore, that merely accepts the rational-philosophical vision of technological knowledge. In fact, there is no need for it to do so, because in the mid 1970s, philosophers, notably David Bloor, outlined the case for a sociology of scientific knowledge (as opposed to practice) built on the idea that – from the perspective of the actors who define them – 'true' beliefs can be ontologically indistinguishable from those that are 'false'.

Bloor's argument was influential and, over roughly the same period as some sociologists were encroaching disaster investigations, others embraced the new epistemologists and began to explore the properties of scientific and technological knowledge. Through a series of epistemologically-conscious ethnographies (or 'epistemographies') they demonstrated that the seemingly abstract concerns of philosophers have very tangible consequences in practice: illustrating the surprising degree to which credible and informed experts often disagree over seemingly objective 'facts' and the frequency with which expert communities reverse their opinion on well-established and apparently inviolable 'truths'.

Some of these sociologists studied engineering knowledge directly, exploring the epistemology of bench tests, much as sociologists and philosophers of science examine laboratory experiments: subverting the orderly public image of engineering by unveiling the messy reality of technological practice. Their studies highlight the practical manifestations of epistemological dilemmas, such as the 'problem of relevance,' to illustrate why engineers cannot definitively interrogate a technology or know the truth of its functioning.

This insight has far-reaching ramifications for the sociology of disaster. It means that there need not be anything ontologically distinct about failure: nothing identifiable that actors 'miss', 'ignore', or 'normalize' in the lead-up to an accident. That is to say: nothing that sociologists can fix. Put differently, if some accidents result from engineering beliefs that prove to be wrong, and it is impossible to be certain of one's beliefs, then some accidents are unavoidable. If *facts* are problematic, then so is the notion of *error*.

There is a 'truth' of technological functioning, as becomes obvious in retrospect, but epistemologically speaking, actors have no objective and unfiltered access to it. There can be no perspective, process, or procedure that will *infallibly* distinguish errors from 'non-errors', and so there can be no way of definitively knowing that errors exist until they manifest in a failure. Even sociologically 'perfect' systems will occasionally fail, and not always for the reasons that are suggested by Perrow.

John Downer is an ESRC Research Officer at CARR.

Detecting the Dodgy Doctor

Mary Dixon-Woods, Justin Waring and **Charles Bosk**

A history of 'bad' physicians stretches back through many different countries. The French doctor Marcel Petiot was a mass-murderer during the Second World War, and the US doctor Herman Webster Mudgett was a serial killer in the 1880s and 1890s. The UK has not been immune. Harold Shipman notoriously murdered over 200 of his patients. Other doctors – including Clifford Ayling, Michael Haslam, and William Kerr – were found to have committed sexual assaults on their patients. Less dramatically, though with devastating implications for the patients involved, have been multiple examples of serious failures in meeting the necessary standards of clinical performance.

Many of these cases have led to official Inquiries. A common theme of these Inquires is that 'warning signs' were visible, often many years before effective action was taken. For example, Shipman's drug addiction had been formally documented while his administration of a fatal dose of an opioid to a patient with asthma did not trigger an investigation. Allegations about Ayling's sexualized conduct went back over more than 20 years – including, in 1980, an allegation that he was found masturbating while conducting a vaginal examination on a patient, but was allowed to continue practising. Others – including Rodney Ledward and Richard Neale – displayed substandard clinical performance for long periods. Ledward, for example, left many of his patients with surgical complications, including permanent incontinence.

Why were these doctors allowed to behave or perform so badly, and why was so little done to respond to the apparent warning signs? The Inquiries tended towards the view that professional and organizational systems and cultures were, in effect, pathogenic: either facilitating or complicit in the conduct of problem doctors. Weak management structures in the NHS, including a lack of clarity about who was in charge of policing problem doctors, and about what authority would legitimate disciplinary action, were argued to have made a major contribution.

The Inquiries also saw the professional ethics of medicine as part of the problem rather than part of the solution to protecting patients. Claims of trustworthiness and virtue were seen to provide a cloak under which nefarious activities could be conducted. The specific form that professional ethics took was also seen to be at fault in promoting the wrong values, especially in encouraging professional solidarity and a culture of doctors not raising official concerns about colleagues.

The period since 2000 has seen a major programme of institutional change in the NHS and the medical profession. Largely shaped, and certainly legitimized, by the findings and recommendations of the Inquiries, many of the new regulatory measures represent significant ruptures in the 150 year regulatory traditions of the medical profession. They include an emphasis on attentiveness to 'warning signs' as a means of detecting and dealing with 'dodgy doctors'. New measures include licensing, certification and revalidation for doctors (where they will have to demonstrate periodically that they are fit to practise) and a system of 'recorded concerns', where concerns about a doctor may be formally collected to identify patterns. But can warning signs or close calls of 'dodgy' performance or behaviour be agreed upon, shared and put to use by policy-makers, professionals and the public alike – especially before patients come to harm?

One important challenge is definitional. It is easy to secure agreement on the broad principle (eg, 'doctors should be good'). But, beyond extreme examples, any attempt to determine whether a particular instance of behaviour counts as a warning sign that an individual doctor or a particular action might not be 'good' is fraught with ambiguities and uncertainties, including questions about who should own the definition of the situation. It can be anticipated, for example, that professional, public and political expectations of medical performance and conduct will differ, especially at the margins of 'good' and 'bad' practice.

Producing definitive rules to govern the classification of actions and practices as deviant, particularly in professional practice, is far from straightforward and can produce unwanted effects. Though such effects are not inevitable, they include the risk that systems of accreditation will descend into ritualized paperwork that simply provokes displays of compliance without capturing problem behaviour or performance, normalizes risks in new ways, or stymies excellent practice.

Setting the standards of acceptable performance is just one area prone to contestation. The report into Rodney Ledward was one of several that advocated the collection and use of data to allow variations in practice and outcome to be monitored. However, the available methods frustrate the use of such data for purposes of detecting the dodgy doctor. One study, using indicators derived from hospital episode statistics, identified Ledward as an outlier in three of five consecutive years. But it also identified 8 other outlier consultants from a sample of 143, and the authors caution that being outlier is not by itself indicative of 'poor' performance. Similarly, a simulated statistical study of mortality monitoring in general practice suggested that it might take 30 excess deaths to detect a murderous trend, and that such a system could result in a high level of false alarms. If such alerts could be treated like smoke detectors that go off every time the toast is burned, that would be one thing. But each alarm is likely to trigger an obligation to investigate that is likely to be shattering for individuals under suspicion and consume resources that might be better spent elsewhere.

Nor is it easy to access less systematic evidence of warning signs. As the Inquiries reported, much of the knowledge about problem doctors was highly idiosyncratic, distributed, and serendipitous, which, as sociological work has earlier suggested, is how such knowledge builds. The clues to Shipman's 'odd' behaviour were widely dispersed and not amenable to formal accounting: they included the taxi-driver who noticed the high death rate among his patients; the undertaker who noticed that his patients often died fully dressed with no signs of serious illness; and the relatives who noticed that Shipman was often cold and aloof when a patient died, and once removed a recently deceased patient's sewing machine while the family watched. The distribution of often fragmented, partial, and difficult-to-codify knowledge amongst isolated or poorly connected individuals and groups represents a considerable challenge for efforts to use warning signs in a systematic and proactive way. On the other hand, a small number of doctors have been subject to campaigns of complaints and vilification by patients and members of the public, posing different and very difficult challenges.

Those working close to doctors are often those who are best placed to identify problems. But it is probably unfair to lay the blame for failure to raise concerns entirely on a collegial conspiracy to maintain solidarity among doctors. Though the exhortation not to speak ill of a brother physician dates back to Hippocrates, it is also true that

every schoolchild learns in the playground that snitching on one's peers is taboo: the etiquette of not 'telling tales' is a generalized one. Speaking up is also difficult when the matters of complaint are so horrifyingly unusual as to strain credulity; a problem, again, that is not UK-specific – it also occurred in the case of the killer doctor Michael Swango in the US.

Further, it was clear from the Inquiries that raising concerns was costly and risky. Reducing the risks for concerned colleagues will require more than chiding those who remain silent. It will need strong systems of organizational support – including effective and well-trained human resources departments – to deal effectively with the kind of tyrannical and threatening behaviour displayed by the Rodney Ledwards of this world. Even with such systems, disruptive doctors may have effective means of neutralizing formal authority, including recourse to employment law. Individual physicians intent on malice may be especially skilful at evading detection. And there are real risks that cooperation and trust among colleagues will be eroded because of the heightened sense of suspicion that may accompany unexpected failures.

It is important not to forget that informal peer-based sanctioning may have highly productive and functional effects that are achieved at low cost. These may be disrupted by efforts to codify and formalize surveillance of problem doctors. Informal interventions, especially if staged by well-trained colleagues, may sometimes be very useful, though clearly they need to be done well, and if they fail, the appropriate formal response needs to be available. The much-maligned 'terribly quiet chat' may, for example, be enough to avert future poor behaviour or conduct in many cases, and may avoid the deviance amplification or labelling effects associated with formal recording of problems.

A final important issue concerns the expressive function of the regulatory system for doctors. There may be good reasons to be cautious about what assumptions and values are expressed by systems that try to use 'warning signs' as a predictor of future misconduct. Does it matter whether we ask doctors to behave well because they know that there are systems of oversight, detection, trapping and control that will compel them to behave in the way required, or is there something important lost or damaged by this? Therein lies the rub.

Mary Dixon-Woods is Professor of Medical Sociology, Department of Health Sciences, University of Leicester. **Justin Waring** is Associate Professor in Public Services Management, Nottingham University Business School. **Charles Bosk** is Professor of Sociology and Medical Ethics, School of Arts and Sciences, University of Pennsylvania. Professor Bosk's research is supported by a Health Investigator Award from the Robert Wood Johnson Foundation.

Constructing Near-Misses:
Proximity, Distance and the Space Between

Carl Macrae

'Near-miss.' The term is redolent of the lucky escape, the close brush with disaster, the narrowly avoided catastrophe. Near-misses don't come much closer than on the morning of November 21st 1989, at London's Heathrow airport. Flying blind in thick cloud and fog, relying on instruments and struggling with bouts of food poisoning and the notoriously tricky old 'Sperry'-type autopilot, the crew of a Boeing 747 began their final approach. Breaking through heavy cloud just seconds before touch down, they had the gut-wrenching realization that they had drifted way off the runway centre-line, out over the airport's perimeter fence. Punching the engines to full go-around power to abort the landing, the aircraft lumbered away, clearing the luxury Penta Hotel with little more than 12 feet (3.65 metres) to spare, sending staff and guests screaming into the street. Near-misses don't come much closer. Nothing but providence and a few feet separated hundreds of people from a horrific catastrophe.

Thanks in part to striking examples such as this, the terminology of 'near-miss' has become firmly established in the risk management industry. Analyzing and learning from near-misses is a central component of many risk management systems. Employees are encouraged to identify and report near-miss events. Companies in industries from healthcare to banking operate near-miss reporting programmes. And regulators and safety agencies, like the UK Civil Aviation Authority and the NHS National Patient Safety Agency, are explicitly charged with analyzing and learning from safety incidents and near-miss events.

Expanding the boundaries of proximity

As near-miss management systems have become more widespread, the range of organizational events that are reported and analyzed within them has expanded. All manner of procedural mishaps, human errors and operational failures are now subject to near-miss analysis. In the majority of cases, it can be hard for the uninitiated observer to identify what, exactly, these near-miss events were 'near' to. In many industries, the stuff of near-miss management is no longer the dramatic and startling crisis in which lives hang in the balance. Instead, near-miss management increasingly focuses on what appear, at first blush, to be rather humdrum and mundane moments of organizational life. As risk management becomes more precautionary, the point in a causal chain that is labeled the 'near-miss' gets pushed further and further away from any actual adverse outcome. In airlines, for instance, typical incident reports highlight that 'flights AB2490, AB2940 and AB2840 all operate from the same station at the same or similar departure times, which causes call-sign confusion.' Or that, 'during pre-flight checks, the wrong departure route was entered into the flight computer and the error was only noticed and corrected after take-off.' These events are a long way indeed from the near-death connotations of the vernacular near-miss. The boundary that determines which organizational events are worthy of risk management attention has expanded dramatically. 'Distant-misses' might be a more appropriate term for many of the events that routinely exercise risk managers.

How can we explain this expanded boundary of the near-miss event, and what does it mean for risk management? One of the most fundamental implications is that near-miss reporting and analysis represents a process of active production rather than passive discovery of risk. Near-miss events are made, rather than thrust upon us. By recognizing and labelling some occurrence a near-miss, an otherwise unremarkable moment of organizational life is transformed into a risk event worthy of examination and analysis and may become the source of considerable organizational change. This active construction of near-miss events raises three issues that can help us better understand risk management practice.

Near to what?

What exactly are near-miss events near to? This used to be simple: a harmful outcome. However, the majority of events reported to risk managers are, in most industries, far from any obvious outcome. Near-miss events are increasingly defined relatively, in terms of proximity to some predetermined level of acceptable safe performance, rather than proximity to a catastrophic event. Take, for instance, another example from aviation. A crew reported that they had landed with only a little more fuel than the required emergency reserves, because they were kept holding for longer than expected and then a snow storm delayed landing further. All aircraft are required to maintain a certain level of reserve fuel for use in an emergency, and the airline's risk managers were concerned that this limit was nearly breached. This event was a concern due to its proximity to a predetermined

safety margin – that the aircraft came close to dipping into its emergency fuel reserves. A similar way of thinking is found elsewhere. In air traffic control, the focus of safety management is on avoiding 'loss of separation' events, where aircraft breach regulated levels of vertical and horizontal separation (typically 1,000 feet and five nautical miles). The event to be avoided is breaching the predefined regulatory limit – even though at that point, the aircraft are still five miles apart. So in precautionary risk management systems the reference point for nearness and proximity becomes displaced, from actual adverse outcomes to predefined limits of safety. Near-misses come to be defined relative to predefined performance limits – rather than the adverse loss event that might lie somewhere beyond it. Coming close to that limit in itself represents a risk, and is perceived as a close brush with danger.

The space between

What makes up the 'space' between a near-miss event and what it is coming near to? At its simplest, the near-miss metaphor implies that proximity is equivalent to risk, and distance equals safety. So what determines distance? When 'near-miss' refers to dramatic brushes with catastrophe, the answer is often simple and literal – small distances in space and time. The few feet between plane and hotel, the disaster avoided with seconds to spare. But these simple metrics are of little use when the near-miss events in question are far removed in time and space from any potential disaster, as is the case in precautionary risk management systems. The answer to what makes this space, in most cases, is organizational.

'Distance' in precautionary risk management is provided by the human, social and technical controls that are able to protect precious assets from coming to harm. 'Space' is created by the defences that prevent small errors snowballing into major catastrophes, and is equivalent to the level of organizational control and resilience available to deal with any particular disruption. This can be illustrated by returning to our two aviation examples above. In the first case, emergency fuel reserves act as a safety margin against an unforeseen emergency. But that safety margin needs to be actively maintained. The flight crew must effectively plan their hold fuel requirements, monitor changing weather conditions and make early decisions to divert to an alternate airport if required; air traffic controllers must monitor and communicate any expected delays; and a vast array of social and technical infrastructure is required to support these activities. The risk managers worried that in this instance, these activities had failed.

The idea of organizational 'space' is equally important in understanding loss of separation events. The UK air traffic services provider, NATS, formally assesses the risk of near-miss events along two dimensions: any loss of separation in geographical terms, and any loss of system defences and controls in organizational terms. So risk managers often worry about events where there was no geographical loss of separation, but the controllers and crew in question did not appear to be fully aware and in control; that is, where the organizational space between aircraft had been reduced. In precautionary risk management systems, safety is not predicated on the mere existence of technical safety margins, but on the organizational work that goes into actively maintaining and protecting those safety margins. It is where this work breaks down that organizations risk breaching their predefined limits of safety and so experience near-miss events.

Bounds of precaution

At what point does a moment of organizational life become a near-miss? How near is near enough? The point at which an event becomes a near-miss, and therefore a risk event worthy of attention, can vary greatly across industries, organizations and situations. Broadly, some industries have well-established and highly precautionary approaches to risk management. In aviation, for example, the bounds of precaution that near-misses come close to appear to exist far from any potential adverse event – a missed routine check of a slats-drive system can provoke considerable concern. Other industries, such as healthcare, are at earlier stages of developing cultures of precautionary risk management. Patient safety, for instance, remains a relatively recent invention, and therefore the recognition of near-miss events routinely occurs at a point much closer to potential harm. The World Health Organization, for instance, defines a near-miss as 'an incident that did not reach the patient', such as where a unit of blood is erroneously connected to the wrong patient's intravenous line but is detected before infusion is begun. This definition places near-miss events close to the point of harm – potentially separated from a harmful outcome by mere inches and seconds.

Just as the general level of risk management maturity can determine these bounds of precaution, so too specific organizational history plays a key role. A recent accident or major incident tends to result in considerably expanded boundaries around similar types of events. One implication of this is that the general level of precaution in industries such as aviation may be the result of a long history of experiencing and learning from accidents – a costly history that healthcare has the opportunity to avoid.

The production of proximity

Near-miss events are not as simple as they once were. Dramatic encounters with risk are increasingly rare in many industries, forcing risk managers to develop more precautionary and expanded boundaries of safety, against which small moments of organizational life can be transformed into consequential near-miss events. As such, the role of risk managers involves something of a subterfuge. On the one hand, risk managers must ensure that organizational mishaps can only happen far from the realm of near-disaster. But on the other, they seek to highlight risks and drive action by actively constructing near-misses: by convincing people that the organization came dangerously close to something it should desperately avoid. While near-miss events have come a long way, most risk managers wouldn't want their organizations to realize that.

Carl Macrae is a Special Advisor to the National Patient Safety Agency.

Near-Miss Management:
Managing the Bottom of the Risk Pyramid

Ulku G Oktem, Rafael Wong and **Cigdem Oktem**

Analyses of serious accidents reveal that prior to an accident, a number of related incidents occur with limited or no impact. Collectively, these incidents are called near-misses, close calls, near accidents or accident precursors. Near-misses are often indicators of system failures that can lead to serious outcomes. Therefore, a Near-Miss Management System (NMMS) can be a powerful tool to reduce risk and improve system reliability. The concept of a near-miss can be applied to almost any operation in any industry.

Although there is not a single, agreed-upon definition of a 'near-miss', for our discussions we will embrace the following broad definition: a near-miss is an event, observation, or situation that possesses the potential for improving a system's safety and/or operability by reducing the risk of upsets, some of which may eventually cause serious damage.

The concept can also be illustrated through a new version of the Safety Pyramid discussed in earlier studies by U Oktem and A Meel (2008, *'Near-Miss Management: A Participative Approach to Improving System Reliability', in Encyclopedia of Quantitative Risk Assessment and Analysis, Melnick, E, and Everitt, B (eds). John Wiley & Sons Ltd. Chichester, UK, pp 1154-1163*), which we have re-named the Risk Pyramid. Oktem and Meel extended the bottom region of the pyramid to include events with no damage (incidents on the cusp of accidents), which we have referred to as 'Foreshadowing Events and Observations.' We modify the pyramid further by adding: (a) another layer to the bottom of the pyramid where the organization believes there are no risks: 'Positive Illusions, Unsafe Conditions and Unobserved Problems – Unawareness, Ignorance, Complacency'; and (b) a new dimension called 'risk perception' along the height of the pyramid. The Risk Pyramid includes all the elements of the Safety Pyramid but extends beyond visibly unsafe operations to include misleading or non-visible (non-obvious) conditions (such as Positive Illusions, Unsafe Conditions and Unobserved Problems). Of course, it is worth noting that the categories in the risk pyramid represent a continuum, with uneven overlapping areas as one moves up the progression.

As in the Safety Pyramid, the Risk Pyramid illustrates that serious adverse events are often preceded by a large number of related incidents with lesser impact and an even larger number of incidents with no adverse effects. Near-misses form the bottom portion of this pyramid while accidents form the top.

The very bottom level of the Risk Pyramid (Positive Illusions, Unsafe conditions and Unobserved Problems), or what we can think of as the 'False Comfort Zone,' describes the conditions in which management, employees and/or customers are under the impression that they are not facing any risks. Problems go unobserved, as do unsafe conditions, aided by general attitudes of ignorance or unawareness. But the false comfort can often go beyond ignorance of existing risks and turn into something more pernicious, a belief that the organization has 'risk immunity' – that everything is proceeding so successfully according to plan that risks cannot exist, thus complacency prevails. In this 'False Comfort Zone,' near-miss events take the form of positive illusions, unsafe conditions and unobserved events.

A recent example of positive illusions, where seemingly good results were near-misses in disguise, comes from the well-publicized Madoff Scandal. As has been described extensively in the media, Bernard L Madoff's firm created a sense of risk-free investing through a combination of consistently high returns, images of trustworthiness (Madoff himself was non-executive chairman of the NASDAQ stock market and served as the Chairman of the Board of Directors and on the Board of Governors of the National Association of Securities Dealers (NASD)), and a brand of exclusivity. All of these factors created positive illusions that masked the true risks.

Other examples of positive illusions that in retrospect were clearly near-misses include the cases of Bearings in 1990 and AIG in 2008. One can argue that failure to notice and act on the weak signals around subprime lending describes the recent case of Fannie Mae's 'accident,' where they lost $59 billion in 2008, leading to a $15 billion cash injection from the government (ref: *Washington Post*, Feb 27, 2009, p D1). Another example of near-miss oversight is the recent Salmonella outbreak incident of the Peanut Corporation of America, which resulted in the loss of lives and massive recalls. Based on the FDA's report (FDA 2009), the Peanut Corporation of America repeatedly shipped peanut butter that had initially been identified as having Salmonella, but then approved on the second round of testing. By doing this repeatedly without suffering adverse consequences, the company developed an attitude of 'no risk,' or complacency, toward batches initially identified as

contaminated. Thus they did not make any effort to change the manufacturing conditions.

Many recent events in the financial industry have been driven by operational failures. The risk and control self-assessment process (RCSA) on which the industry continues to rely has proved inadequate for managing these risks.

A near-miss management system instituted properly can reduce the risk of accidents by catching the 'near-miss' signals early and alerting the organization to the potential danger. Although many institutions have some type of near-miss system under various names, their effectiveness in preventing events with high negative impacts varies widely. The Wharton Risk Center's Near-Miss study, conducted in 2000, shows that in order for a near-miss management system (NMMS) to be effective, it must cover the entire range of operations and must contain the essential components of eight steps, all of which should be implemented successfully. These steps are:

Step 1 Identification and recognition of a near-miss

Step 2 Disclosure (reporting) of the identified information/incident

Step 3 Prioritization and classification of information for future actions

Step 4 Distribution of the information to proper channels

Step 5 Analyzing causes of the problem

Step 6 Identifying solutions (remedial actions)

Step 7 Dissemination of actions to the implementers and general information to a broader group for their knowledge

Step 8 Resolution of all open actions and review of system checks and balances

While a near-miss programme must be an integral component of any risk management system, it cannot alone provide a comprehensive risk prevention mechanism. It should be strengthened with other preventive intelligence tools/actions.

Integration of near-misses into risk assessment methods can be accomplished in two different, and complementary, ways:

a) By using near-miss data to develop better, more realistic, estimate of failure probabilities;

b) By revising and modifying the results of conventional risk analysis, such as fault trees, to make the system more robust.

Near-miss management systems can improve the outcome of risk evaluation in financial industries by providing more complete risk evaluation for each of the components of the operational risk management framework (resource failure likelihood, business impact analysis and risk shield analysis).

For example, in recent years, before the financial crisis hit, companies pressured their loan officers to sell high volumes of credit cards and mortgages, and, as a result, loan officers made simple mistakes while processing quantities of paperwork. Had these near-miss events been recognized, they could have alerted the financial services companies to the non-desirable exposure of uncompleted documentation and helped them avoid significant losses due to incomplete documentation that prevented the legal procedures from being implemented during recovery of funds.

Although there is a wealth of research in this area, there is still a lot to be learned about why management and workers fail to recognize the real risks. The question still remains: What are the conditions in different industries that will enable organizations to reside at the bottom of the Risk Pyramid?

Ulku G Oktem is Senior Fellow at the Risk Management Center, and Adjunct Professor, Operations and Information Management Department, Wharton School. **Rafael Wong** was Operational Risk Director and is currently Collection Officer of Banco de Credito del Peru. **Cigdem Oktem** is Senior Principal, Value Engineering, SAP America.

Intelligent Assessment Tools in Heathcare:

Technological Fix or the Potential for Unintended Consequences?

Nicola Mackintosh and **Jane Sandall**

Introduction

It has been estimated that approximately 23,000 in-hospital cardiac arrests in the UK and at least 20,000 unanticipated intensive care unit admissions in England, Wales and Northern Ireland could be prevented with better care. Clinical deterioration of patients on general wards is often preceded by changes in physiological observations in the period six to 24 hours before an adverse effect. However, these changes in clinical signs are often missed, misinterpreted or mismanaged. A number of safety solutions have been introduced principally aimed at facilitating earlier identification and treatment of deterioration in ward-based patients. This paper focuses on intelligent assessment tools and explores their potential for providing not only early warnings, but also unintended consequences.

The Tool

Intelligent assessment tools aim to facilitate an appropriate, graded medical response based on the severity of the condition of the patient. They use personal digital assistants (PDAs), tablet PCs and hospital intranets to replace traditional paper observation charts with real-time data and electronic charting. Vital signs data such as physiological parameters of blood pressure, heart rate and respiratory rate are directly entered into the PDAs. By allotting points to these vital sign measurements on the basis of physiological derangement from a 'predetermined range', a cumulative score or 'Early Warning Score' (EWS) is electronically generated. When the score reaches an arbitrarily predefined threshold the PDA triggers prompts for action e.g. 'increase frequency of observations to at least every 30 mins' and calls for help, e.g. 'involve registrar from the patient's team immediately'. The raw physiology data together with EWS and vital signs charts is made available to members of the healthcare team via wireless networking and linked to the hospital intranet.

The Evidence Base

These tools are predicated on the idea that there is increased accuracy and recording of physiological data, and that there can be a correct attribution of a weighted value (EWS) according to the degree of physiological derangement. There is some evidence that utilization improves EWS error rates. However, to date there is little empirical research to illuminate whether these tools trigger remedial actions at the right time. Data has not yet emerged regarding their ability to reduce rates of 'failure to rescue' and their capacity to reduce avoidable adverse events or death. Interestingly, although EWSs have been recommended to identify patients at risk for the past ten years, there is still limited evidence of their ability to predict patient outcomes or impending deterioration. Ensuring effective use of EWS has proved problematic and its adoption does not invariably result in improved clinical outcomes. Poor methodological quality standards and wide variations between different systems locally restricts comparison of outcomes and standardization of care.

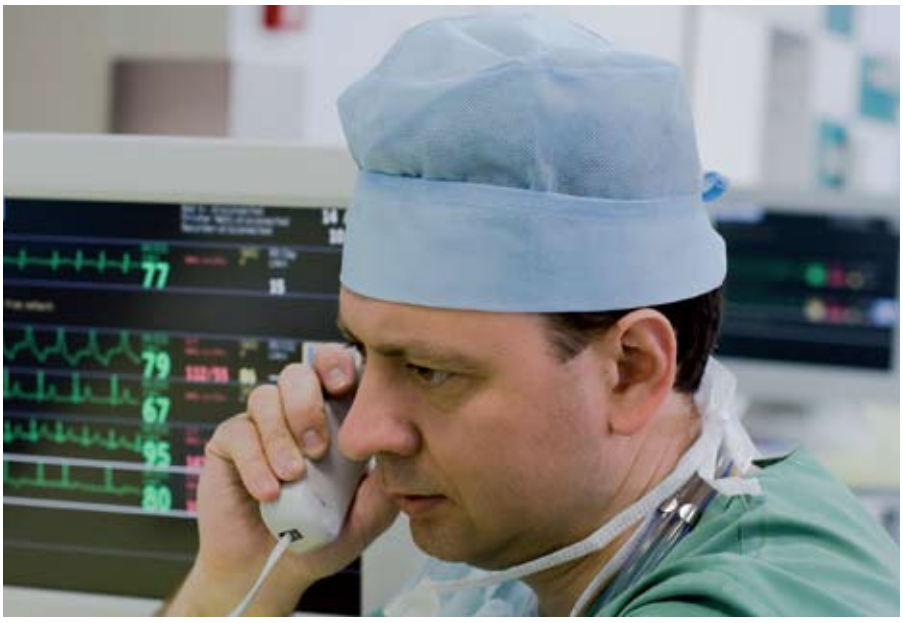
The Opportunities

Evidence suggests that failure to detect deterioration and thus institute timely management is frequently linked to poor patterns of taking and recording observations. These include partial observations, absence of observations at night, incomplete charts and EWS not being completed or being miscalculated. These tools offer a means of increasing the completeness, accuracy and legibility of vital signs data. A nurse taking a set of observations is directed by the PDA to enter physiological values in a certain chronological order. The nurse can choose to bypass a prompt, but the device will default back to the missing value and request its entry. The PDA therefore directs data collection and facilitates the gathering of a complete set of observations and EWS each time a measurement is made. Each data set is displayed in real-time, accurately dated and timed, together with the name of the staff member who recorded the observations. Repeat observation times are directed on the screen potentially prompting timeliness of observation recordings. Paper observation charts traditionally kept at the bottom of the patient's bed are replaced by 'high quality' electronic displays on PCs.

Importantly, the tool aims to improve response behaviour and to design out individual variation in response to the data received. Prompts on the PDA direct the nurse to repeat the observations at certain time points and to call for help. Remote access to the datasets can aid medical prioritization when medical teams are 'offsite'.

Aggregated patient data can provide ward managers with an electronic overview of the ward, identifying 'high risk' patients at a glance and contributing to 'situation awareness', an important precursor for safety performance. The tools, like other technologies, whilst designed to impact at individual level offer opportunity for coordination of activities and sensitivity to events and emerging problems. Their performative function ensures that the software flags up overdue observations. Similarly the system warns if erroneous values, partial data or 'unlikely observations' are entered or the same data is regularly recorded. Like other information technologies, the tools have a generative capacity, offering opportunities to bring together larger numbers of entities and events in more distinct spaces and times, enabling performance monitoring across wards.

Nurses have been found to use vital signs and EWS to effectively 'package' information about patient deterioration, providing doctors with persuasive referral language. These tools could provide nurses with the license to demand a review from the medical team, overcoming professional hierarchies. They could offer an opportunity for boundary work, enabling nurses to gain authority and 'symbolic capital', thus improving their social position.



The Risks

Underpinning the tools is a belief that incompleteness of vital signs data and overlapping data (duplicate charts) are problems that need designing out. The PDA orders particular vital signs and facilitates collection of a complete EWS dataset for all patients. However, this routinization and ordering, which is irrespective of a patient's individual condition, may promote 'mindless' utilization, suppressing active vigilance and attention with the risk of missing the exceptional cases. The regulation of observations may have the potential to stifle information and interpretations that may be considered intuitive and experiential; devaluing the merit of subjective data in defining and managing patients at risk. There may be times when pragmatism and the application of contingent standards is required when staff need to override the system e.g. around end of life care and chronic illness. Conversely, over-sensitivity of EWS may result in 'trigger fatigue', reducing faith and belief in the system. Overdue observations and high early warning scores, flagged as a means to trigger changes in practice, could lose meaning and significance over time, especially if past experience of the triggers has cast doubt on their validity. Deviance, manifested as failure to respond to the triggers, may become normalized and recast as an acceptable risk.

Overlapping data is perceived as a form of redundancy and a problem to be solved rather than recognizing duplication of effort in recording data as a source of reliability. Loss of paper records means there is no back-up system enabling access to vital sign data if the network goes down. Access to data is contingent on availability of PCs which during ward rounds and peak activity on the wards may be problematic. Viewed remotely, observations are disconnected from the patient, removing the practitioner's capacity to contextualize these signs with supplementary data collected from visual cues and patient narratives. Remote access to vital sign data may downplay the paper record's important role as a 'key material structuring device' and the face-to-face communication that often happens around the observation chart. When the 'instrumental rationality' of information systems are prioritized above the 'communicative rationality' of cooperative inter- and intra-professional working practices, there is a danger that staff become controlled by the very technology installed to facilitate working routines.

Lastly, the interpretation of numerical data generated may determine the nature of generalizable knowledge about social phenomena. These tools can be perceived as a form of surveillance, opening up opportunities for a network of accountabilities and blame of particular professional groups. The data generated provides a 'good story', enabling the organization to focus on those aspects of behaviour that are more easily auditable (eg, taking of observations), rather than others (eg, response behaviour). System design faults may be marginalized.

Conclusion

Technological determinism underpins the rationale for these tools, the belief that technology has the capacity to determine how people act. Whilst the tools offer potential for managing patients whose conditions are deteriorating, their 'agency is dependent upon and embedded in the practical interactional circumstances' in which they are deployed. Questions still remain about how we promote 'mindful' use of the tool.

Nicola Mackintosh, NIHR Patient Safety & Service Quality Research Centre, King's College London.
Jane Sandall, NIHR Patient Safety & Service Quality Research Centre, King's College London.

Biases in Organizational Learning Produce Vulnerabilities in Safety-Case Regulation

Johan M Sanne

Safety-critical industries, such as transportation, energy production or the chemical industry, are complex, socio-technical systems that cause intertwined, and often latent risks. These risks require a sufficiently varied response, including recognizing and making sense of their cause, significance and how they should be appropriately addressed. Such industries are usually regulated through safety-cases: regulatees produce a safety case, including risk analysis and learning from incidents, with associated means to manage identified risks. The regulator scrutinizes whether the safety case complies with relevant regulations, approves of it, or requests revisions. Learning from risk analysis and near misses is therefore a salient part of safety-case regulation.

Sometimes, though, the opportunities for regulatory reform might not be exploited sufficiently because the means and outcome of organizational learning and risk management do not fit the various threats that appear in a given industry. I will illustrate my argument by analyzing how the Swedish nuclear power industry addressed an incident in the Forsmark plant in 2006.

According to the industry report, the reactor brushed with disaster when one reactor was shut down and

the emergency power supply (to cool residual heat) was almost blocked due to the interaction between maintenance error and the failure of three technical subsystems: maintenance error, which caused an outside switchyard to uncouple one of the reactors from the national grid; inadequately installed low frequency protections for the turbines, which caused a transient electrical surge through the plant; and the failure of two out of four emergency power systems due to faulty design.

Instrumentation in the control room was misleading due to deficiencies in the man-machine interface, training and manuals. Some indicators fed from the failing emergency power systems were missing. The operators did not know if the control rods had been fed into the reactor and were ignorant about the water level and the reactor pressure. However, owing to instructions, experience and training, they were able to conclude what had happened and after 22 minutes they manually connected the regional grid. If more than two emergency power systems had failed, the operators' intervention would have been essential to save the core from damage.

The accident was caused by inadequate design, stemming from an insufficient risk analysis of the consequences of certain external disturbances for the internal electrical grid, as well as an inadequate installation of the low frequency protections. However, rather than investigating the processes that really took place, (either those that caused the incident or those that could have prevented it), the official investigation focused on engineering safety. It was informed by a 'human error' framework and a safety culture concept developed after the Three Mile Island and Chernobyl accidents. In this way, the event was normalized and the potential for revitalizing regulation was missed.

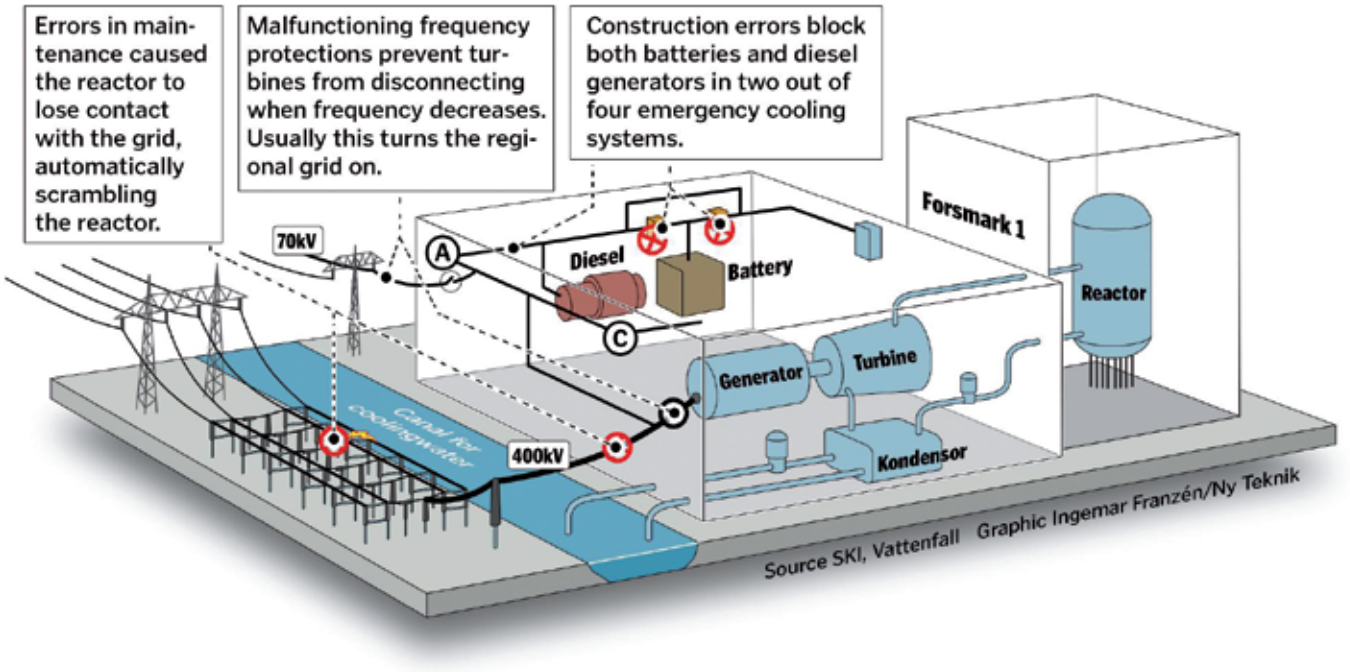
Through interviews and focus groups with plant safety managers and regulatory agents, I found a number of data and contributory processes that were left unaddressed in the report. The safety manager at Forsmark, for example, showed that he understands the real practices in control room work in ways consistent with ethnographic research of this work:

> Of course you can have conflicting data, since two instruments show the right level, so to speak. Two instruments showing [divergent outputs from two others] can seem odd […]. On the other hand, you can judge that these are unusual indications, that there are two instruments that show [correct readings] and thus we probably should act upon these.

The control room team does not just wait for alarms before taking action, but needs to prepare for unusual situations by continually making sense of normal operations. But current analysis of the control room operators' work is theoretically underdeveloped (the safety manager's story may be heard in corridors or at conferences but is not sufficiently widespread) and so it becomes difficult to generalize the implications.

The industry learnt from Chernobyl the need for a trustworthy 'safety culture' defined in terms of attitudes or behavior. The Forsmark incident report found operational safety culture deficiencies in many areas, and the plant board and CEO were replaced. The report also attributed the causes for the faulty redesign processes of the frequency protections to missing or inadequate instructions, arguing that these followed from the safety culture deficiencies. However, in interviews, regulatory agents told me that instructions existed but were not used. They assigned the faulty redesign to a risk assessment procedure that did not have enough scope. They also claimed that deficiencies in quality routines did not cause the incident, although they made it more difficult for the control room operators to manage, because the deficiencies contributed to the inadequate instrumentation, instructions and training. Thus, the report both makes erroneous assumptions about the causes behind the faulty design process and fails to address a potential clue about why things went wrong. This is probably due to an inadequate understanding of management as being a rational implementation of plans and instructions, rather than something that stems from unquestioned procedures and beliefs about relevant risk objects.

The industry initially struggled to make sense of the failed design of the plant's internal grid and found it



The incident at Forsmark 1: only one of the two reactor turbines and the corresponding generator are shown as well as only one of the four emergency cooling systems (A-D).
Copyright: Ingemar Franzén/Ny Teknik.

troublesome, as explained by the safety manager for the Ringhals nuclear power plant:

> If it had been a simple construction error… [for instance] if one should have constructed two but constructed one [or] if somebody made a mistake; someone did not follow an instruction; someone passed over that page; someone did not look into that paragraph. If that had been the problem, then one might have pointed it out. Then it would have been a quality defect. From what I understand in Forsmark, the construction that they had functioned the way it was intended. But the intended way was, as we know, not sufficient; it did not provide sufficient margins for this disturbance.

The design error was not a human error the way it is normally defined and managed. Rather, engineers had not understood the functionality of the power plant properly and had not foreseen what might happen. Moreover, the error violated required design principles that should provide defence in depth, causing a 'common cause failure' for which one error causes several breakdowns to occur, a design error that Forsmark shared with many other plants around the world.

Eventually, the design error became normalized as an example of engineering uncertainty for which a redundant design provided necessary protection, and a new risk object, the plant's electrical grid, was added to the checklist of safety-critical items. This explanation, however, left unexplained the causes behind the failed design process. The interviewees gave me three potential explanations, each pertaining to a different process and different time periods. First, when the nuclear plant was built around 1980, a transient surge was expected by the engineers who designed the national grid but this was not communicated to those who designed the plant grid since the former had not reflected on its significance for the plant. Second, when the emergency cooling system was redesigned in 1994, the plant likewise failed to realize the potential for interaction between the national grid and the plant grid. Third, the safety manager at the nuclear plant, Ringhals, argued that two incidents concerning electricity systems and diesel generators in other plants in 1999 and 2000 might be interpreted as precursors to the Forsmark incident, and were therefore incidents that they failed to learn from.

In summary, organizational learning for safety-case regulation in the nuclear industry is a discipline framed by engineering and human factors. This causes poor conceptualization of organizational and man-machine relations, stemming from the risk objects that the industry learnt from previous events. Institutionally, organizational learning is framed by an interest in resuming operation through fixing technical and organizational shortcomings. These biases caused both the design errors and the inadequate analysis of their causes. Accordingly, regulation focuses on technology rather than understanding social processes within engineering and operation. The investigation provides detailed analysis and recommendations for the technical errors encountered, while it is restricted to vague insights into faulty organizational processes – such as risk analysis – or learning from successful control room work.

The case suggests that systematic biases in organizational learning are problematic in a complex safety-critical industry as they might not match the variety in the risks the industry faces. There is a need to reform the conceptual and institutional means for safety-case regulation by inviting new groups of experts and opening up the regulatory process to other stakeholders.

Johan M Sanne is Associate Professor, Department of Thematic Studies – Technology and Social Change, Linköping University.

Invisible Deaths:

The Challenge of Slow-Burning Mortality Crises for Public Health Agencies

Carine Vassy and **Robert Dingwall**

Many organizational failures arise from an inability to assemble and interpret known information correctly. The appropriate design response is to improve the means by which information is fed to critical decision-makers and to enhance their ability to use it in anticipation of, or in rapid response to, emerging problems. However, there is also a class of failures that are wholly unforeseeable, that appear from directions that are entirely unknown and unpredicted, where the only relevant response is to improve an organization's scanning of its environment in the hope of identifying the looming catastrophe before it actually happens. Such a strategy, though, faces the challenge that Plato labelled as Meno's Paradox, that if we knew what we were looking for, we would not have to search for it; while, if we do not know what we are looking for, we do not know how to look for it. Some two thousand years later, Donald Rumsfeld restated this as the problem of unknown unknowns, the things we do not know that we do not know.

As the recent influenza pandemic shows, public health is an area where 'early warnings' are particularly significant. An important aspect of this work is the attempt to identify population health challenges at an early stage, where they can be contained and controlled. One element starts from the known fact that people die. Deaths can be early warnings – but what are they early warnings of? The unknown is whether an individual death has a wider significance. Is it the first death in a cluster or an epidemic? Is it just a routine event? If it does seem to be part of a cluster, is this just random clumping or do these deaths share a common cause?

Since the middle of the 19th century, all developed countries have established civil registration systems that produce mortality data for their populations at various levels of aggregation. These data are the resources for epidemiological analyses that try to find patterns in these deaths that might allow their causes to be identified and suggest strategies for intervention. Such analyses are a potential source of early warnings. However, recent events in three developed countries – France, England and the USA – have exposed the limitations of this method. Slow-burning mortality crises were missed and organizational responses were inadequate or not triggered at all. As a result, all three countries have embarked on reforms of their civil registration systems intended to increase their responsiveness.

Reforms in France, USA and England

The trigger in France was the 2003 heatwave, 'La Canicule', which is thought to have caused about 15,000 excess deaths, and which provoked a major political scandal, leading to a senior public servant's resignation. Historically, death registration in France has two separate elements, which follow different tracks and are never reconciled. This system is designed to preserve the confidentiality of the cause of death, even from close relatives. When local authorities register deaths, they report the fact of the death to INSEE, an arm of the Ministry of Finance, and the cause, which comes under seal from the attending doctor, to a national health research institute (INSERM). Although some hospital physicians alerted regional health offices to an unusual number of deaths in early August 2003, this information was not recognized as evidence of a potential crisis. Many nursing homes saw extra deaths – but typically these rose by small amounts from a low base and their cumulative significance was overlooked. Only when overloaded emergency department physicians went to the media was the surge of deaths acknowledged, by which time intervention was pointless.

Subsequent reforms have concentrated on speeding up the reporting process through online communication between local government and INSEE. Currently, this covers about 70 per cent of deaths, although without data on their causes. However, there are still significant lags: only about 50 per cent of deaths covered by this system have been logged within three days. There is also more active surveillance through regional health networks, although this, again, has a lag of three to four days. More ambitious schemes for putting the entire death certification process online are being discussed, but face formidable financial, technical and professional obstacles.

These reforms will help the French know more rapidly whether they have experienced a mortality crisis, and something of its character, but it does not seem realistic to suppose that they will actually help with real-time surveillance and early warning.

The USA has had its own problems with heatwaves: July 1995 in Chicago closely resembled August 2003 in Paris. However, the main US driver for reform has been concern about identity theft, linked to the homeland security agenda. US death registration is complicated by the degree of state autonomy. In effect, each state operates its own recording system, often only sending annual summaries to the National Center for Health Statistics. Recognition of local mortality clusters tends to depend on informal networks between coroners or medical examiners, and their interactions with the medical community. Signals of concern have to emerge from this background conversation and then be communicated to civil authorities.

While there has been an investment in accelerating reporting since 9/11, this has focused on trying to close the gap in which a dead person's birth certificate can be used to obtain other identity documents before it is linked to a death certificate. With federal support, some states have begun to introduce online death registration but there have been important technical and financial difficulties. Because of the mobility of the US population and the limited co-ordination between state systems, it seems unlikely that acceleration will do much to detect abuse of the certification processes and it is certainly not intended to increase early warning time for mortality events.

The United Kingdom also has separate registration systems in England and Wales, Scotland, and Northern Ireland, although their differences are more constrained than between US states and mainly relate to variations in their legal and institutional systems. In all three jurisdictions, for example, causes of death are publicly available data. England also experienced the 2003 heatwave in London and the South East, but did not notice 1,500-2,000 excess deaths until its mortality statistics were subsequently re-examined in the light of the French

experience. Unlike the other countries, England has long produced weekly estimates of mortality, although these are not considered to be accurate enough for policy until about three weeks after the event. Even the weekly estimate describes deaths that occurred about ten days previously. As elsewhere, concern is triggered mainly through informal networks between hospitals, primary care and public health agencies.

Reform has been driven by the failure to detect the activities of Dr Harold Shipman, who is believed to have murdered at least 200, and possibly as many as 400, of his patients over a twenty-year period. He covered his tracks by manipulating the process of death certification, particularly the requirement for a second physician to confirm cause of death if a body is to be cremated. Although England has accelerated online reporting from local offices to the national level, the main objective of reform has been to enhance the accuracy and auditability of death certification, rather than to increase its use as a means of identifying other kinds of mortality crisis. The English records will be much clearer about how people died – but this will still be some time after the event.

Conclusions

In effect, each country's reforms are those of generals fighting the last war. Each system has responded to its most recent challenge, rather than reflecting on the objectives of death certification. France has accelerated reporting but still does not reconcile deaths and causes. The USA strives to reconcile birth and death certificates. The UK attempts to prevent another serial killer going unobserved. Arguably, however, each is asking civil registration to assume a weight that it cannot bear. Clearly nation states need an authoritative means of accounting for deaths, in order to maintain reasonably accurate estimates of population size and distribution for national planning, to prevent homicides from going undetected, and to prevent identity theft. All of these pull in a different direction from the challenge of delivering valid and reliable early warnings of breaking mortality crises. However, public health systems need both, as recent disease outbreaks, from HIV/AIDS to pandemic influenza, have shown.

An effective early warning system will require public health agencies to embrace an 'intelligence gathering' model of working, rather than relying on the passive 'bureaucratic' model of civil registration. An example would be the GPHIN system run by Canada in partnership with WHO. This involves continuous real-time scanning of global news media, mostly online, to identify reports of potential public health significance. In the current influenza pandemic, we are aware of similar proposals to use other types of internet traffic as indicators of the distribution of emerging hot spots. Locally, it demands a more systematic approach to sharing information between health-related agencies and professionals, including groups conventionally excluded from information-sharing such as funeral directors.

None of this would come as a surprise to the nineteenth century pioneers of public health work, who recognised the importance of street-level knowledge and local outreach. However, it may challenge some of their more desk-bound successors. It also questions the drift towards a passive, consumer-led health system, dealing with expressed needs rather than searching for emerging problems, potentially compromising individual privacy interests in order to achieve collective benefits. Our experience of pandemics, however, should caution us against shutting our eyes and ears to the early warning signals that are out there if we know how to find them.

The work described in this paper was carried out in collaboration with Richard Keller (University of Wisconsin) and Anne Murcott (University of Nottingham) supported by funds from the Leverhulme Trust and the French Ministry of Health.

Carine Vassy is Senior Lecturer in Sociology at the University of Paris 13, and a researcher at the Interdisciplinary Research Institute on Social Issues (IRIS). **Robert Dingwall** is Director of Dingwall Enterprises, Nottingham.

carr
centre for analysis
of risk and regulation

LSE

ESRC Centre for Analysis of Risk and Regulation
The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7955 6577

Fax: +44 (0)20 7955 6578

Website: www.lse.ac.uk/collections/CARR/

Email: risk@lse.ac.uk