What Will We Do With the Data?
Issues in the Reporting of Adverse Healthcare Events

C.W. Johnson,

Department of Computing Science, University of Glasgow,
Glasgow, Scotland, G12 9QQ.
http://www.dcs.gla.ac.uk/~johnson, Email: johnson@dcs.gla.ac.uk

Abstract
Incident reporting has been proposed as a means of identifying and addressing the causes of human error in medicine. Politicians, regulators and professional bodies have started initiatives to implement these schemes in many different countries. It is time to take a rational look at the limitations of incident reporting. Many people have been too ready to believe the over-stated claims about the effectiveness of incident reporting in other domains. Others have not listened to the more limited claims made by the operators of these existing systems in aviation and in organizational health and safety applications.

Keywords: Incident reporting; human error; adverse events.

1. Introduction
On the 13th June 2000, the Chief Medical Officer for England, Liam Donaldson, announced the establishment of a centralized reporting facility for adverse incidents across the UK National Health Service (NHS) (BBC, 2000). The Chief Medical Officer said; “At the moment there is no way of knowing whether the lessons learned from an incident in one part of the NHS are properly shared with the whole health service". The UK Health Secretary Alan Milburn said; “Patients, staff and the public have the right to expect the NHS to learn from its mistakes so we can ensure the alarm bells ring when there are genuine concerns so they can be nipped in the bud". In May 2001, Tommy Thompson, Secretary at the US Department of Health and Human Services told the Senate Health, Education, Labor and Pensions Committee**: “**(We have) highlighted the need to establish a national focus to create leadership, research, and tools to enhance the knowledge base about safety; to identify and learn from medical errors through mandatory and voluntary reporting systems; to raise standards and expectations for improvements in safety through the actions of oversight bodies, group purchasers, and professional organizations; and to implement safe practices at the delivery level” (Thompson, 2001). The following pages provide a brief overview of the many problems that are emerging as clinicians, managers, regulators and research teams struggle to fulfill these political visions.

1.1 The Benefits of Incident Reporting

The political initiatives in the UK and the US are motivated by the apparent success of incident reporting in the field of aviation following patterns established by NASA's Aviation Safety Reporting System and the UK Confidential Human Factors Incident Reporting scheme (CHIRP). The focus moves away from the analysis of low frequency, high-consequence events to more frequent near-miss events. This offers numerous benefits. Incident information provided by pilots, ground crew, ATC helps regulators to find out why accidents DON'T occur. Reports describe how operators cooperate to detect and mitigate adverse events. The higher frequency of incidents also permits quantitative analysis in a way that is not possible with lower frequency accidents. Information about previous incidents is, typically, published in newsletters and increasingly on web sites. This reminds staff about potential hazards and helps to keep them "in the loop". The data (and lessons) from incident reporting schemes can be shared. Incident reporting systems provide the raw data for comparisons both within and between industries. If common causes of incidents can be observed then, it is argued that common solutions can be found. Incident reporting schemes are cheaper than the costs of an accident. A further argument in favor of incident reporting schemes is that organizations may be required to exploit them by regulatory agencies. There is also a widespread but erroneous perception that these systems are 'low cost' and simple to operate. (Johnson, 2003). Given this list of benefits in the aviation domain it is hardly surprising that incident reporting should be proposed as a key technique to improve patient safety.

2. Obvious Issues and Unasked Questions

This paper identifies the limitations of incident reporting. The intention is not simply to criticize these systems but also to identify areas where existing weaknesses can be addressed.

2.1 The Problems of Participation Bias

It can be difficult to elicit incident reporting within the medical domain. Those reports that are obtained may not be well distributed across all grades of staff. For example, a Scots adult intensive care unit found that nursing staff contributed 90% of all reports submitted over the last decade. Nurses submitted 621 reports; medical staff only reported 77 incidents (Busse and Wright, 2000). This bias is often assumed to indicate reluctance by senior staff to participate in reporting schemes. In consequence, schemes will pay undue attention to the execution of medical procedures by nursing staff rather than the planning, coordination and administration of treatment by senior personnel. However, things are more complex than this simplistic analysis would suggest. It is difficult to interpret the distribution of reports from different clinical disciplines and grades. It is important to consider the total number of staff who might contribute to such a system. Usually the team consisted of three medical staff, one consultant, and up to eight nurses per shift. The larger number of reports contributed by nursing staff can also be explained in terms of the involvement in, or exposure to, the types of workplace incidents that were solicited under this particular scheme. Nursing staff had the most direct contact with the patients who remain the focus of the reporting

system. Hence, it can be argued that they have a proportionately greater opportunity to witness adverse events.

2.2 The Problems of Elicitation and Form Completion

Jha et al (1998) illustrate the problems of eliciting incident reports. Their work has detected adverse drug events using three different techniques: voluntary incident reporting; the computer-based analysis of patient records and exhaustive manual comparisons of the same data. Both the automated system and the chart review strategies were independent and blind. Looking at admissions to nine medical and surgical units in an eight-month period, the computer monitoring strategy identified 2,620 incidents of which 275 were confirmed to be adverse drug events. The manual review found 398 adverse drug events. Voluntary reporting only detected 23. It is too simplistic to argue that this disparity stems from a reluctance to make voluntary reports. It can be difficult for staff to detect incidents when they occur. Adverse consequence may only emerge over a matter of days, weeks or months. It can also be difficult for busy clinical staff to find the time to complete a report for minor incidents, for mishaps that they are unsure about or where issues of anonymity are important. There are also more practical problems, in particular printed reporting forms are often difficult to obtain. Staff must be motivated to find one, fill it in and then post it to the appropriate safety managers. Many national schemes have responded to these practical problems by introducing web-based systems. However, I have found most of these applications to be deeply flawed in terms both of their engineering and, in particular, their usability. Typically, they have large logos announcing that the user is making a submission through the voluntary reporting system. For many medical staff, this can be very off-putting if they have to submit a form using a shared PC in a busy ward. Further problems stem from the design of these new computer-based reporting forms. For example, one system forced the user to enter the date of an incident using pull-down menus entitled Date, Month and Year. They were also forced to specify a single time at which the incident occurred. This created enormous problems when I conducted usability tests with healthcare staff. They were concerned to enter information about incidents, for instance where a medication had been incorrectly administered several times. Should they enter the first date when this occurred or the date on which it was identified? One group of clinicians even entered six separate forms for each instance; thereby creating problems for the safety managers who must identify they all involved the same patient. Many of these problems affected the design of previous generations of paper based reporting forms. However, computer-based systems are less forgiving in terms of the ad hoc annotations and inclusions that can be submitted about adverse events and near misses.

2.3 Problems of Analysis and Classification

Assuming that the new generation of reporting systems can overcome the barriers to submission, they must face a number of additional problems associated with the analysis of adverse, clinical events. The US Aviation Safety Reporting System relies upon dedicated teams of coders who will analyse each incident

submitted to the system. This is appropriate because coders can be trained and monitored to ensure consistency in their work. Unfortunately, the analysis of adverse events illustrates further differences between the aviation and healthcare domains. In particular, it is unlikely that any national system could afford to employ dedicated teams of coders for all medical events. Leape (2000) notes that the Aviation Safety Reporting System spends about $3 million annually to analyze approximately 30,000 reports. This equates to about $100 per case. At this rate, it would cost around £50 million to conduct a similar level of analysis for the 850,000 adverse events that are estimated to occur each year within the NHS. It has been estimated that the cost of clinical negligence to health authorities and NHS Trusts was approximately £ 200 million in 1995-1996. The cost of claims is expected to rise by 20%p.a. over the next 5 years. The NHS summarized accounts for this period include provision totaling £80 million with contingent liabilities of £1.6 billion. The costs of a centralized analysis service have persuaded many regulators to encourage local investigation by safety managers. This is a sensible approach given that these individuals often have a clear understanding of the context in which an adverse event occurred. However, in the UK they lack standardized training. In a series of empirical studies, it has been shown that there is no agreement over the causal analysis of a sample of adverse events by safety managers within the NHS (Jeffcott, 2003). This has led to the development of analytical tools and classification schemes, similar to flowcharts that help safety managers to identify appropriate causes from the incident information. Unfortunately, previous experience has shown numerous limitations with these techniques. There are theoretical problems. For instance, the more exhaustive the taxonomy then the more likely it is that analysis will find a classification that matches their incident. It is also likely, however, that the use of an extended classification will reduce the consistency of any analysis of similar events by different investigators. The FDA describes further problems. A violent patient in a wheelchair was suffocated through the use of a vest restraint that was too small. The risk manager, JC, continued:

> “She finds the list of event terms, which was detached from the rest of the coding manual... She muses: ‘Mr. Dunbar had OBS which isn't listed in these codes; he had an amputation which is listed; he had diabetes which isn't listed; and he had hypertension which is listed'. JC promptly enters 1702 (amputation) and 1908 (hypertension) in the patient codes. She then finds the list for Device-Related Terms... She reviews the terms, decides there was nothing wrong with the wheelchair or the vest restraint, and leaves the device code area blank." (FDA, 1996).

The resulting classification of 1702 (amputation) and 1908 (hypertension) provided few insights into the nature of the incident. It is, however, characteristic of the analysis obtained from overstretched staff faced with complex incidents and inadequate training.

2.4 Lack of Computational Expertise

Several national governments have commissioned national patient safety agencies to encourage the reporting of adverse events systems. A key element in this national infrastructure is the provision of software tools to store the many thousands of incidents analyzed by regional safety managers. Unfortunately, there has been a lack of investment in appropriate technologies to support this task. This shortcoming can be explained in terms of a reliance on human factors experts rather than software engineers in the design and development of these systems. The consequences of this lack of engagement are far reaching. I have already mentioned the poor user interfaces to many electronic reporting forms. The lack of computational expertise also affects the design of storage and retrieval systems for incident records. Most of the proposed national schemes are based around relational database technology. In essence, this stores incident data according to the classification schemes that have been developed by human factors experts to classify the causes of human error in terms, for instance, of slips, lapses and mistakes. However, problems arise when the taxonomies change. This is likely when the human factors view of key concepts such as workload have developed radically over the last decade. The net effect is that in ten years time we may have to go back into our electronic databases and manually reclassify many hundreds of thousands of reports to reflect a revised taxonomy.

There are further technical issues in the development of reporting systems. Precision and recall are concepts that are used to assess the performance of all information retrieval systems. The precision of a query is measured by the proportion of all documents that were returned which the user considered to be relevant to their request to the total number of documents that were returned. In contrast, the recall of a query is given by the proportion of all relevant documents that were returned to the total number of relevant documents in the collection. It, therefore, follows that some systems can obtain high recall values but relatively low precision. In this scenario, large numbers of relevant documents will be retrieved together with large numbers of irrelevant documents. This creates problems because the user must then filter these irrelevant hits from the documents that were returned by their initial request. Conversely, other systems provide high precision but poor recall. In this situation, only relevant documents will be returned but many other potential targets will not be retrieved for the user. In most other areas of software engineering, the trade-off between precision and recall can be characterized as either performance or usability issues. In incident reporting schemes, these characteristics have considerable safety implications. For instance, low-recall systems result in analysts failing to identify potentially similar incidents. This entirely defeats the purpose of compiling national and international collections. Conversely, low-precision approaches leave the analyst with an increasing manual burden as they are forced to continually navigate "another 10 hits" to slowly identify relevant reports from those that have no relation to their information needs. Again this can result in users failing to accurately identify previous records of similar incidents. There has been relatively little investment by patient safety agencies into these issues. In consequence, it can be argued that we are relatively unprepared for the problems that will lie ahead when increasing amounts of data is obtained about adverse events and near-miss incidents.

2.1 Unrealistic Expectations and Political Pressure

Charles Billings (1998), the former Chief Scientist at NASA Ames has identified some of the limitations that might affect healthcare schemes based on his experience with the Aviation Safety reporting System:

> "… there are enough reports of mishaps with potassium chloride, lidocaine, vincristine and other drugs and devices to have made it very clear that a problem with these exists. The information that these events occur is already present. We may well ask what it is that keeps us from making progress on safety, given that we already know about the existence of these problems. What is added by more formal, elaborate (and expensive) incident reporting?"

This is a crucial insight. National governments have already established high targets and aspirations for their healthcare systems. For instance, the NHS (1998) has been asked to meet the following objectives:

> "...by 2001, reduce to zero the number of patients dying or being paralyzed by maladministered spinal injections (at least 13 such cases have occurred in the last 15 years); by 2005, reduce by 25% the number of instances of negligent harm in the field of obstetrics and gynecology which result in litigation (currently these account for over 50% of the annual NHS litigation bill); by 2005, reduce by 40% the number of serious errors in the use of prescribed drugs (currently these account for 20% of all clinical negligence litigation); by 2005, reduce to zero the number of suicides by mental health inpatients as a result of hanging from non-collapsible bed or shower curtain rails on wards (currently hanging from these structures is the commonest method of suicide on mental health inpatient wards)."

It is unlikely that incident reporting alone will yield the insights necessary to achieve these objectives. As Billings observes incident reporting systems often only serve to confirm existing suspicions about the causes of medical errors. They do not immediate suggest solutions. Fortunately, these schemes provide one component of broader initiatives to address healthcare incidents. National patient safety organizations are commissioning research in many different areas. These will, arguably, have a greater impact than the reporting systems. For example, smart infusion devices are being designed to recognize the drugs that they administer using indirect sensing techniques. Other initiatives are focusing on device labeling and communication protocols between teams of healthcare workers. Few of these initiatives stem uniquely from the insights of reporting systems.

3. Conclusion

Incident reporting systems do yield good insights into local problems and in ideal situations can be used to identify regional and national patterns of failure. However, my intention has been to encourage a critical reappraisal of the claims that are being made for these systems. A secondary aim has been to encourage

national patient safety agencies to look beyond the teams of human factors specialists who are developing taxonomies of medical error. In particular, it is essential to recruit user interface designers who can help clinicians and safety managers to enter information about complex adverse events, for example through direct user testing of on-line forms. Similarly, it is important to recruit specialist expertise in the storage and retrieval of large data sets. There are particular attributes of incident reporting systems that make it dangerous to rely upon conventional relational databases. Alternative techniques, including free text retrieval, offer greater flexibility and support the use of human error taxonomies that will change over time. It should be no surprise that these are key areas for recent investment within many of the major aviation reporting systems (Johnson, 2003). It would be better for patient safety agencies to invest now rather than relearn the lessons of existing reporting systems in other domains.

References

L.B. Andrews, C. Stocking, T. Krizek, L. Gottlieb, C. Krizek, and T. Vargish. (1997), An alternative strategy for studying adverse events in medical care. Lancet, (349):309-313.

C. Billings, (1998), Incident Reporting Systems in Medicine and Experience With the Aviation Safety Reporting System. Report from a National Patient Safety Foundation Workshop on Assembling the Scientific Basis for Progress on Patient Safety, http://www.npsf.org/exec/billings.html

British Broadcasting Corporation (2000). Plan to stop dangerous doctors. News Staff, BBC, London, United Kingdom, 13 June 2000. http://news.bbc.co.uk/hi/english/health/newsid 788000/788805.stm.

D. K. Busse and D. J. Wright. (2000), Classification and analysis of incidents in complex, medical environments. Topics in Health Information Management, 20(4):1-11,

Food and Drug Administration: M. Weick-Brady. Those codes! User Facility Reporting Bulletins, 1996. http://www.fda.gov/cdrh/issue18.pdf.

C.W. Johnson (2003), A Handbook of Accident and Incident Reporting, Springer Verlag, London (in press).

A.K. Jha, G.J. Kuperman, J.M. Teich, L. Leape, B. Shea, E. Rittenberg, E. Burdick, D.L. Seger, M. Vander Vliet, and D.W. Bates. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. Journal of the American Medical Informatics Association, 5(3):305-314, 1998.

S. Jeffcott (2003), Risk Perception in the UK NHS, PhD Thesis, Department of Computing Science, University of Glasgow, Scotland.

L.L. Leape (2000), Editorial: Reporting of Medical Errors - Time for a Reality Check, Quality in Health Care, (9)3:144-145.

NHS Expert Group on Learning from Adverse Events in the NHS. An organization with a memory. Technical report, National Health Service, London, United
Kingdom, 2000. www.doh.gov.uk/orgmemreport/index.htm.

T. Thompson (2001), Secretary at the US Department of Health and Human Services statement to the Senate Health, Education, Labor and Pensions Committee, May 24, 2001.
http://www.os.dhhs.gov/asl/testify/t010524.html

C. Vincent, S. Taylor-Adams, and N. Stanhope. (1998), Framework for analyzing risk and safety in clinical medicine. British Medical Journal, (316)7138:1154-1157.