Journal of Risk Research
Vol. 12, Nos. 3–4, April–June 2009, 485–511

A human error taxonomy for analysing healthcare incident reports: assessing reporting culture and its effects on safety performance

K. Itoh[a]*, N. Omata[a] and H.B. Andersen[b]

[a]*Tokyo Institute of Technology, Tokyo, Japan;* [b]*Technical University of Denmark, Lyngby, Denmark*

The present paper reports on a human error taxonomy system developed for healthcare risk management and on its application to evaluating safety performance and reporting culture. The taxonomy comprises dimensions for classifying errors, for performance-shaping factors, and for the maturity of reporting culture contained in incident reports. Applying several dimensions in the taxonomy, we propose on the one hand two safety performance measures, i.e., the rate of near-miss reporting and the rate of near-miss detection by safety procedure, and on the other, measures for diagnosing reporting culture including average descriptive depth in reports.

We applied the taxonomy to a total of 3749 incident cases collected from two Japanese hospitals, which were at different stages of patient safety activities: Hospital A initiated organisation-wide initiatives several years before the survey period, while such safety-related activities had just commenced in Hospital B. The hospitals also differed in their reporting rates of incidents per nurse: 3.05 (A) vs. 0.65 (B). Results show that the taxonomy can identify differences between these hospitals both in terms of safety performance and reporting culture. In addition, a correlation trend was observed between these two measures.

Keywords: human error taxonomy; reporting culture; incident reporting; organisational learning; inter-rater reliability; patient safety

1. Introduction

A variety of approaches and methods for risk management have been applied to patient safety issues in healthcare, often adapted from human–machine systems research and development in other safety critical domains such as aviation, maritime operations and nuclear power production. Among such approaches, incident reporting systems have been regarded as a key methodology for risk management in Japanese healthcare organisations (Itoh 2003). One of the primary aims of incident reporting in healthcare is to manage *organisational learning* effectively within a healthcare organisation (Department of Health 2000) and to perform systematic analysis of incidents in order to improve safety and quality (Kohn, Corrigan, and Donaldson 1999; Aspden et al. 2004), similar to such applications in other domains (Barach and Small 2000; Itoh, Seki, and Andersen 2003). Thus, the primary function of reporting is not to collect the greatest number of cases, nor to assess the distribution of different types of mishaps; rather, it is to obtain stories from which we can learn; i.e., narratives contained in incident reports. For these requirements, human error taxonomy can be critically useful for supporting the incident analysis to enhance systematic organisational learning.

*Corresponding author. Email: itoh.k.aa@m.titech.ac.jp

ISSN 1366-9877 print/ISSN 1466-4461 online

DOI: 10.1080/13669870903047513
http://www.informaworld.com

A human error taxonomy in healthcare has several roles and aims. One of the most important roles is to provide a meaningful structure for describing, explaining and understanding specific events, and therefore a taxonomy shall support a framework for systematic analysis of incident and near-miss reports to support organisational learning (Kohn, Corrigan, and Donaldson 1999; Runciman et al. 2006). The taxonomy is also required to develop a database of incidents or adverse events as common terminology and classification schema for collecting data (Chang et al. 2005; van der Schaaf and Habraken 2005). It can be guided *retrospective* analysis for identifying causes, contributing factors to systems failure and adverse events (Chang et al. 2005). It may be further applied *prospectively* to identification of potential risk factors and prevention and mitigation strategies/actions to protect patient harm.

For the aforementioned applicability and potential benefits, a variety of taxonomy systems have been developed as common frameworks for incident analysis, e.g., Rasmussen's (1986) Skill-, Rule- and Knowledge-based (SRK) model, Step-ladder model and Abstraction hierarchy; Reason's (1990) Generic Error-Modelling System (GEMS); and Hollnagel's (1998) Cognitive Reliability and Error Analysis (CREAM), and as analytical tools for specific domains, not only in healthcare but also in human–machine operation domains, for example, aviation (Isaac et al. 2001; Isaac, Shorrock, and Kirwan 2002; Shorrock and Kirwan 2002), process industry (van der Schaaf 1996) and power industry (Taylor-Adams and Kirwan 1995). In recent years, taxonomy systems have also been developed for healthcare settings, among which some are applicable to all clinical medical work (Chang et al. 2005; World Health Organization (WHO) 2005; Zhang et al. 2004). Among them, some systems are strongly integrated with analysis and data managing features as part of incident management systems like the Australian Incident Monitoring System (AIMS) (Runciman, Helps et al. 1998; Runciman, Williamson et al. 2006) and PRISMA-Medical (van der Schaaf and Habraken 2005; Salo and Svenson 2003). There are also several studies on healthcare incident taxonomy which focus on specific areas of medicine such as general practice (Dovey et al. 2002; Kostopoulou 2006), paediatrics (Woods et al. 2005) and emergency medicine (Hammons et al. 2003).

Most healthcare taxonomies include the following dimensions or aspects, sometimes under slightly different labels: (a) event type (what happened); (b) domain (characteristics of staff and patient involved, and setting); (c) errors and causes; (d) contributing factors (hazards, root causes, latent conditions, contextual conditions); (e) impact (outcome, consequences, or level of harm); and (f) lessons learned (measures taken or proposed, or prevention and mitigation). In addition, some studies focus on more specific components such as cognitive failures and system factors (Kostopoulou and Delaney 2007) and work process and knowledge and skills (Makeham et al. 2002). For instance, the PRISMA-Medical taxonomy (van der Schaaf and Habraken 2005; Salo and Svenson 2003) applies the Eindhoven Classification Model to patient safety incidents, focusing on latent causes of events, and accordingly it includes classification of two types of errors: active failures and latent conditions. The latter type is further divided into two sub-classifications, technical and organisational factors, both of which have several categories. For instance, external, transfer of knowledge, protocols, management priorities and culture are included for the latter factor. The Joint Commission of Accreditation of

Healthcare Organizations (JCAHO) Patient Safety Event Taxonomy (Chang et al. 2005) is a comprehensive system that includes five primary classifications: impact, type, domain, cause, and prevention and mitigation. These root nodes of classification are divided into 21 sub-classifications, and further divided into more than 200 categories in total. The WHO classification scheme, which is the result of a wide-scale international effort at establishing an internationally applicable taxonomy, comprises the following top-level classes: incident type, contributing factors, event characteristics, patient characteristics, recovery factors, preventive factors, mitigating factors, patient outcome, organisational outcomes and actions taken (WHO 2006).

The purpose behind the development of the healthcare incident taxonomy to be described here was three-fold: (1) to identify by retrospective analysis of incident cases the risk factors (latent causal factors) that exist in a specific organisation or work unit; (2) to understand the nature and characteristics of errors and associated factors of various professional areas in healthcare – such as clinics, nursing and pharmacy – and in order to track them over time via analysis of incident reports; and (3) to assess current levels of risk or safety, and in particular, related measures to safety culture (Itoh, Andersen, and Madsen 2006; Madsen, Andersen, and Itoh 2006) for given work units and the entire organisation with the aim to proactively improve patient safety.

With respect to the last purpose, it may be possible to calculate several types of reporting rates, for example, ones applying to a set of all incident reports or to sub-sets that consist of reported events belonging to specific classes, for example, level of severity, type of event, and then use the rates as indicators of safety performance. However, it is difficult to obtain a valid estimate of the incidence of different types of events since incident reporting is selective (Battles and Lilford 2003). Thus, reporting rates of incidents may not only depend on external factors but may also reflect internal processes or risk factors (Itoh, Andersen, and Madsen 2006). It has become widely accepted that organisational factors are of critical importance for safety in human–machine system operations (Reason 1993). At the same time, it is well-known that organisational problems are typically latent factors that contribute or even lead to the occurrence of human error made by frontline personnel (Reason 1997) and healthcare (Kohn, Corrigan, and Donaldson 1999; Department of Health 2000).

A useful and comprehensive concept referring to organisational factors involved in safety performance is that of *safety culture* which, at the same time, emphasises the more or less shared understanding by staff and management of norms, work conditions, and constraints (cf. Helmreich and Merritt 1998; Reason 1997). To establish a 'safe' organisational culture, Reason (Reason 1997; Reason and Hobbs 2003) suggested that an organisation must improve the following three inter-connected component cultures: (1) *reporting culture* – organisational climate in which people are prepared to report their errors and near-misses; (2) *just culture* – atmosphere of trust in which people are encouraged, even rewarded, for providing essential safety-related information; and (3) *learning culture* – willingness and competence needed to draw right conclusions from safety information systems such as the incident reporting system, and the will to implement major reforms when their need is indicated. Safety culture is typically assessed by questionnaire-based surveys or safety audits. It may also be possible to estimate elements relating to safety culture or its component cultures, especially for the reporting culture, by using data from incident reports with no additional effort of surveys or audits.

Based on the aforementioned background, we developed a healthcare incident taxonomy system which seeks to provide not only a framework for understanding characteristics of patient safety incidents but which also can be applied to retrospective incident analysis and predictive identification of safety status. The present paper focuses on applicability of the taxonomy system to assessing safety level and safety culture status, and reports results of applying this to a number of incident reports collected from two hospitals in Japan. We also discuss correlations between levels of safety performance and reporting culture estimated by indicators proposed in this study. Inter-rater reliability using the taxonomy was also evaluated based on calculation of raw and chance-corrected agreements.

2. Taxonomy and patient safety indicators

2.1 Generic dimensions of taxonomy

A prototype of the healthcare incident taxonomy was developed by partly adapting from the Human Error in ATM [Air Traffic Management] (HERA) taxonomy (Isaac et al. 2001; Isaac, Shorrock and Kirwan 2002; Shorrock and Kirwan 2002), and by adding dimensions specific to healthcare setting (Itoh and Andersen 2004). Subsequently, an additional section was included to analyse maturity of reporting as well as several minor revisions of the original sections. The current taxonomy is divided into six overall sections, each of which contains one or several dimensions comprising individual categories. Descriptions of sections and dimensions in the taxonomy system are shown in Table 1 and a complete list of categories in each dimension is provided in Table 2.

One of the dimensions in the 'event outline' section is the 'problem type', which specifies medical and nursing activities involved in the incident, for example, injection, medication, diagnosis, treatment and tube troubles. The dimension of 'action failed' describes the staff actions at the time when an error was made. Actions in this dimension are further divided in terms of generic actions and typical 'healthcare' objects involved in the action, as shown in Table 3. The generic actions are classified into several types adapted from the Simple Model of Cognition (SMoC; Hollnagel 1998) and expanded with 'execution' divided into two modes, motor execution (handling) and verbal execution (communication). The information/equipment dimension specifies the type of information, if any, that healthcare staff misperceived, forgot or misjudged or the human–machine element, if any, that was used in the failed action.

The 'error occurrence' section involves two dimensions referring to the type of error, i.e., how it occurred in terms of errors of omission and commission (Swain and Guttmann 1983), and the type of violation (Mason 1997). The section of 'contextual conditions' that lists underlying factors behind errors, is similar to what other authors have called contributing factors, performing shaping factors or latent conditions. Contextual conditions are divided into seven types of 'background' factors, as indicated in Table 1. Each of these factors includes several elements or categories that potentially affect safety and quality of activities. For instance, typical elements of organisational factors are (poor or inadequate) manuals and checklists, medical documents, rules and procedures, training, staffing, scheduling, management style and safety climate. Besides these elements, communication failures may sometimes be classified under problems in organisational factors

Table 1. Sections and dimensions included in the taxonomy.

Dimensions	Descriptions
Event outline	Overall description about event occurred.
Problem type	What goes wrong, what kind of problem reached or was about to reach the patient.
Action failed	What action(s) failed, e.g., cognitive, communication or motor action.
Information/equipment	What information was misperceived or was not perceived, or what equipment or materials were mismanipulated or misused.
Error occurrence	Classifications about erroneous action(s) involving in the event.
Error type	What error(s) occurred in terms of omission/commission classification.
Violation	How and what violation of protocol, rule or procedure occurred.
Contextual conditions	Performance shaping factors or underlying causes that lead or contribute to the event.
Communication factors	Failure, lack of, no or low quality of communication between staff and patients, between team members, within an organisation, etc.
Staff human factors	Staff's human factors characteristics, including knowledge and skills, experience, stress, fatigue, and other team and individual factors.
Patient human factors	Patient's individual characteristics which may influence incidence of events.
Task factors	Factors relating to work conditions, schedule, requirements, etc.
Equipment/materials factors	Factors relating to medical equipment or materials, and their technical failures.
Organisational factors	Organisational aspects that directly or indirectly may influence safety and quality of medical and nursing activities, and their management.
Environmental factors	Factors relating to work setting and environment surrounding medical and nursing activities.
Outcome and recovery	Outcome or effects of the event, situations when detecting error, and intervention or recovery actions for the event.
Outcome severity	Level of outcome or effects of the event in terms of loss or harm to the patient.
Error capture cue	Who detected the error or the event or by what it was captured.
Error recovery	What intervention or recovery actions were taken when the error was detected.
Preventive mechanisms	Measures or means taken, proposed or learned to protect patients from the same or similar events or to reduce their occurrence or effects.
Maturity of reporting	Reporting characteristics and contents described in the report.
Timing of reporting	When the incident was reported in term of the duration after its detection.

Table 1. (*Continued.*)

Dimensions	Descriptions
Reported content	What was described about the case in the report.
Time-band of description	What statements were described in the report in terms of time or tense of things that happened or expected.
Descriptive level	How well, deeply or thoroughly the case was described in the report.

Table 2. Categories in each dimension of the taxonomy.

Dimensions	Categories
Event outline	
Problem type	Injection; medication; diagnosis; treatment; inspection; tube troubles; patient falls; equipment manipulation; patient care; office work or administration.
Action failed	(See Table 3)
Information/equipment	Patient (including patient's vital signs and other states); team member; medicine; medical document (chart, record, etc.); medical tool, device, machine or materials; computer interface (e.g., interface of electronic patient record); command or message; situation variables of sickroom; situation variables of workplace; and others.
Error occurrence	
Error type	Omission error; timing error; qualitative error; selection error; sequential error; extraneous error; no error made.
Violation	Unintentional rule contravention; routine violation; exceptional violation; situational violation; optimising violation; no violation included.
Contextual conditions	
Communication factors	Communication problem between different professional groups (e.g., doctor and nurse); between members in the same professional group (e.g., between nurses); between staff and patient; between staff and patient family or relatives; inter-department collaboration or communication problem.
Staff human factors	Knowledge and skills (including work experience); health; fatigue; memory failure; inattention; confusion; wrong assumption or preconception, time pressure, personal fear; emotional stress; decreased motivation; other psychological or emotional factors.
Patient human factors	Age (e.g., elderly or baby); dementia, cognitive failure; motor disability; perceptual impairment; reduced consciousness; anaesthetised or sleep; mental instability; other specific factors.
Task factors	Overload (high task demand); complicated task; novel task; multiple goals; work interruption; busy situation; exceptional work condition; midnight; holiday; other task-related factors.
Equipment/materials factors	Equipment/materials failure; obsolescent equipment/materials; access problems; bad interface design; bad materials design; inconsistent design of equipment/materials.

Table 2. (*Continued.*)

Dimensions	Categories
Organisational factors	Bad/inappropriate medical document; rules or procedures; training; manuals and checklists; task allocation or staffing; lack of resources; scheduling; decision making procedure; leadership; safety culture related problem; other organisational issues.
Environmental factors	Layout in sickroom/workplace; lighting in sickroom/workplace; available space in sickroom/workplace; patients with same or similar names; medicines with similar names/bottles/colour; other environmental factors.
Outcome and recovery	
Outcome severity	levels 0–5: Level 0: Near miss, error did not reach the patient, no harm; Level 1: No effect and no harm though error reached the patient; Level 2: Temporary effect was given to the patient but additional action was not required for the event, but no harm; Level 3: Minor outcome or temporary harm was suffered with additional treatment and/or longer hospitalisation for the event; Level 4: Major outcome or permanent harm was suffered; Level 5: Death.
Error capture cue	Built-in safety procedure; staff involving the case; patient or patient family; other patients; team member in the same shift; team member in another shift; leader, supervisor or superior; staff in other departments; other persons or detection cues.
Error recovery	Informing leader or superior/doctor in charge/associated departments; monitoring patient states; diagnosing patient by doctor; additional inspection; treatment for harm; suspension of the planned treatment; explanation about the event to patient or family; apology to patient or family; other actions for the event.
Preventive mechanisms	Compliance with existing procedures, rules or protocols; ensuring to check required actions; improvement of communication; fool-proof; staffing; work scheduling; addition or change of checklist/safety manual/safety procedure; improvement of work condition/work environment/sickroom environment/medical document/management style/human-machine interface; education and training; introduction of or update to new equipment/materials; root cause analysis for the case; improvement of safety climate; other actions.
Maturity of reporting	
Timing of reporting	Immediately after detection; within an hour; 3 hours; 12 hours; 24 hours; 3 days; 4+ days.
Reported content	Event type; situation surrounding reported event; error causes; contextual conditions (cf. 'contextual conditions' dimension; lessons learned; measures for avoidance.

Table 2. (*Continued.*)

Dimensions	Categories
Time-band of description	Past (long before the event); pre-event (immediately before the event); present (when the event happened); post-event (immediately after the event occurred); and future (situation to be expected).
Descriptive level	Levels 1–5: Level 1: No meaningful information contained in the report; Level 2: Described simply or briefly; Level 3: Described at more length but superficially; Level 4: Described in detail; Level 5: Described in depth.

(Chang et al. 2005). In this taxonomy, however, this type of underlying causes was included as an independent factor in view of the fact that failures of communication within and between teams or units are frequently observed in adverse event cases in healthcare.

The 'outcome and recovery' section comprises three dimensions: outcome severity, error capture cue, and error recovery. According to our preliminary analysis of incident reports collected from several hospitals in Japan, a majority of cases, reported by nurses in particular, were 'no harm' incidents. The proportion of the reported 'no-harm' events is 90%, which is much higher than that described in other studies from Western countries; for example, 55.8% in the USA (Dovey et al. 2002). Also, results of questionnaire-based surveys of Japanese healthcare staff point in the same direction. The surveys, which asked Japanese healthcare staff to rate their likelihood of submitting a given incident report presented as a plausible but fictitious event, showed that 64% of doctors and 67% nurses agreed strongly or slightly that they would report a near-miss case in which an error was intercepted before the sequence of actions were completed (Itoh et al. 2005). Therefore, our taxonomy system is required to distinguish near-misses (Level 0 in severity level) from no-harm events, in which an error reached the patient but with no effect at all (Level 1), and those with a temporary effect (physical effect such as temporary pain and fever or psychological discomfort – Level 2), progressing to temporary harm (Level 3), permanent harm or death (Levels 4 and 5), as can be seen in Table 4.

The dimension of 'error capture cue', which refers to how a wrong act was detected, and includes in its categories the persons who detected the incident, and the cues used to detect the event. The 'error recovery' dimension includes categories not only about the actual error recovery actions, such as diagnosing the patient and therapeutic actions, but also interactions with patient such as explanation about the event and possibly apology as well as reporting actions.

The section of 'preventive mechanisms' is supposed to be used in two ways. One is to apply the classification to identify latent factors by conducting a detailed root cause analysis. Therefore, the section includes several categories of counter-measures to guard against repeated occurrence of the event through, for example, fool-proof design, staffing and improvement of work condition. The other purpose of applying this section is to specify what was leaned from the case by staff involved in the event or staff in the department/ward or even in the entire organisation.

Table 3. Categories for the action failed dimension.

	Patient (or family)	Team member		Device/tool/ materials	Medicine	Document
		Same profession	Other profession			
Observe (incl. *identify*)	*Observe* patient	*Observe* colleague	*Observe* other professional	*Monitor/check* dev./tool/mater.	*Check* medicine	*Check* document
Interpret (*decide* & *analyse*)	*Diagnose* patient	*Decide* colleague's work	*Decide* other professional's work	*Decide* on dev./ tool/mater.	*Decide* on (specifying) medicine	*Decide* with document
Plan (synthesis)	*Treat* patient	*Plan* colleague's work	*Plan* other professional's work	*Plan* on dev./tool/ mater.	*Plan* on medicine	*Plan* with document
Handle (motor execution)	*Take care* of patient	*Cooperate* with colleague	*Coordinate* with other professional	*Manipulate* dev./ tool/mater.	*Handle* medicine	*Write* document
Communicate (verbal execution)	*Communicate* with patient	*Communicate* with colleague	*Communicate* with other professional	–	–	–

Note: –: not applicable.

Table 4. Severity levels of outcome.

Harm			No harm		
Death	Permanent harm	Temporary harm	Minor effect, no harm	No effect	
	Major injury	Intervention required	Intervention not required	Error reaches patient, but no effect	Near-miss
					Error does not reach patient
Level 5	Level 4	Level 3	Level 2	Level 1	Level 0

The section of 'maturity of reporting' includes the four dimensions: timing of reporting, reported content, time-band of description, and descriptive level. The 'timing of reporting' dimension is used in identifying how promptly an incident report was submitted after the event. Several categories are included in terms of duration from detection to reporting of the event. By 'reported content' we refer to properties of the case as described in the incident report: the type of event involved (e.g., medication, injection, patient fall, and misuse of equipment), situation at the event occurred (e.g., busy ward, work interruption, and nightshift), causes of errors, contextual conditions behind errors (e.g., descriptions about human factors, organisational factors, and poor communication), and lessons learned.

Within the 'time-band of description' we identify the time span between the time of reporting and the time of the failed action and the discovery of the incident. Finally, the 'descriptive level' concerns the depth or thoroughness of the report and is rated on a five-point scale of: (1) no information; (2) described very briefly; (3) described at some length but superficially; (4) described in detail; and (5) described in depth.

2.2 Steps of classifying a report into categories of the taxonomy

In this sub-section, we shall mention a typical procedure for categorising results of analysis of incident reports. In Japan there is no nation-wide reporting system, and every hospital has its own reporting form for incidents. The systems differ and items included in the reporting forms are not the same. However, every form has a field for open-ended description of what and how the event happened. Each hospital form also includes several fields about causes and factors that must be selected from a list of alternative items.

For each incident report of this study, we would identify one or more errors, their causes and relating factors as well as relevant information on the event for all the possible dimensions comprising the taxonomy in the following steps:

(1) Discrepancy check of incident description: first, it is checked that any description of cause–event relations are coherent and not discrepant. If no human error or causal factor is stated in the report, it is returned to a risk manger for further details.
(2) Initial assignment of categories: one or multiple items are selected for each dimension from a category list (see Tables 2 and 3) based on the open-ended description on the event. In particular, one must identify what errors took place and then judge what types of errors they were and what actions were failed.
(3) Additional selection of categories: additional category items are elicited in accordance with a guideline that describes how to match categorised items on errors and incidents that are specific to each hospital's reporting form with categories classified in the taxonomy. In addition, a taxonomy category that obviously applies to a stated fact or action will be added, even though the report has not described this in direct terms. For instance, if a nurse involved in an incident has only a few months of experience, it can be inferred that she has little experience and limited skills.

When category assignment is completed for a certain batch of incident reports for a given period (e.g., monthly, quarterly), incident statistics are generated in terms of problem type, actions failed and other dimensions in the taxonomy to analyse trends of patient safety events within an entire hospital or each department/ward. Such data can be used to determine which events should be taken up for further detailed root cause analysis. Focusing on a specific type of 'problem' or an individual 'case' that has been elicited by the application of the taxonomy system, detailed retrospective analysis can be performed to identify latent factors and develop counter-measures to prevent re-occurrence – typically changing (improving) some of the 'contextual conditions' which the taxonomy has allowed one to identify.

2.3 Diagnosing indices of reporting culture

We suggest several indicators to assess the level of reporting culture by applying the dimensions in the 'maturity of reporting' section. As one of the most straightforward indicators, the level of reporting culture in a specific department/ward or an entire organisation can be assessed in terms of percentage of case reports that actually describe the items under 'reported content', and so, for instance, if a high percentage of reports do not mention error causes or contextual conditions or lessons learned, this is an indication that the reporting culture is at a low level.

Another possible indicator is the *average descriptive depth*, which represents the overall level of details of a report. This is indicated on a five-point scale of 'descriptive level' assigned to each report (see Table 2).

In addition to richness of description in a report, promptness in reporting an event is another candidate indicator of reporting culture level, if one assumes that the sooner an event is reported the greater is the value that staff puts on reporting.

2.4 Safety performance indices

To assess safety performance, we suggest reporting 'rates' calculated by the proportion of a specific sub-set of reports received, e.g., outcome severity. As pointed out by several authors (Edmondson 1996, 2004; Itoh, Andersen, and Madsen 2006), making an error is different from documenting an error which is based on detecting or capturing the error. Therefore, when comparing between organisations or between work units within a single organisation, we should be careful in interpreting the rate of reported incidents. In general, this index may be interpreted either (a) as a measure of risk – the greater the rate of reported incidents of a given type, the greater is the likelihood that a patient injury may take place; or (b) as an index of safety (the inverse of risk) – the more that staff is willing to report, the greater is their sensitivity to errors, the greater the learning potential, and so, the greater is the safety in their department.

Accordingly, it has been observed that the reporting rate of incidents may be viewed as a 'safety' index – more frequent reporting of incidents is a sign of safety (Edmondson 1996, 2004). As mentioned previously, we make a distinction between near-miss cases in which a wrong act did not reach a patient (Level 0 in outcome severity classification) and incidents at the other severity level. Reports of 'near-miss' events can therefore be considered to reflect the staff's sensitivity to errors or risk conditions. Based on this assumption, we introduce the percentage of near-miss cases overall reported incidents of all severity levels as one of the possible indicators of the safety level of a unit or organisation. We refer to this percentage as the *rate of near-miss reporting*.

Moreover, we classify ‘near-miss’ events in terms of the types of ‘error capture cues’ used to detect the error. For instance, a wrong act detected by a safety rule or procedure built into the work system can be expected to capture the same error on other occasions before it reaches a patient. In contrast, errors detected by other cues may possibly only be caught by chance. Thus, the rate of reported errors caught by the built-in safety procedure compared to all near-miss cases reported is, we argue, a sensible and practical indicator of safety performance. We call this the *rate of near-miss detection by safety procedures.*

Another safety performance indicator to consider is the rate of Level 3+ of outcome severity compared with all cases (see Section 2.1 and Table 2). A case at Level 0–2 with at most a temporary discomfort or minor temporary pain is usually regarded as a ‘no-harm’ incident in Japanese hospitals. According to several interviews with risk managers and nursing department leaders, staff members will nearly always – and for nurses it is strictly always – report cases involving harm, i.e., Level 3 or higher. Therefore, this proposed indicator is, we suggest, a plausible risk index. Yet, since the number of reported incidents of level 3+ is, fortunately, quite small (0.8% of all cases from nurses in our sample), it is not a useful indicator to be used to compare smaller units or even larger units for shorter time periods.

We do not wish to suggest that these indicators can be used to compare different types of departments. For instance, some doctors are required, of course, to perform actions on patients that are inherently more hazardous than those required from nurses.

3. Applied cases of the taxonomy

We applied the taxonomy system developed in this study to a total of 3749 incident reports collected during two overlapping 2.5-year periods in two regional hospitals Japan. The number of collected reports, survey period and other hospital data are shown in Table 5. According to our interviews with the risk managers, organisation-wide initiatives for patient safety were initiated several years before the survey period, and special attention had been devoted to the incident reporting system in Hospital A. In contrast, such safety-related activities had just commenced at the end of the survey period in Hospital B. There appear to be differences in risk management and reporting culture in the hospitals, and they might well lie behind their annual incident reporting rates per nurse: 3.05 cases (Hospital A) vs. 0.65 (Hospital B).

Table 5. Collected reports and hospital attributes surveyed in this study.

Collected reports		Survey period	Number of beds and staff		
			Beds	Doctors	Nurses
Hospital A	3162	Apr. 2002–Sept. 2004	347	60	310
Hospital B	587	Apr. 2000–Sept. 2002	475	70	314
Total	3749				

4. Application results

4.1 Changes in reporting culture

When a healthcare organisation has a positive reporting culture we may expect that a staff member will submit an incident report unhesitatingly as soon as possible if he or she has become involved in any reportable event. To examine this expectation, we calculated distributions of the 'timing of reporting' categories, i.e., when incident cases were reported to the local reporting system. Figure 1 illustrates the result based on three half-year periods between April 2003 and September 2004 in Hospital A (the first half of fiscal year begins 1 April and ends 30 September, and the second half from 1 October to 31 March next year). There was a significant difference in the timing of reporting between the three half-year periods (χ^2=28.75, p<0.05). Compared with the earlier two half-year periods, the rate of 'immediate reporting' became greater and that of reporting after an hour was smaller in the last period of the survey. Approximately 60% of the incidents were reported within three hours after detection of the event, but for about a quarter of the reports, there was not enough information to determine the timing.

To address another aspect of the reporting culture, we examined whether incidents were 'properly described' by analysing the results of categorising the 'reported content' dimension for every half-year period. Figure 2 depicts changes in percentage description for each category of this dimension in Hospital A. The percentage description for each category of reported contents increased significantly every half-year period (F=4.42, p<0.05; analysed by the two-way ANOVA with the periods and the reported contents). As illustrated in Figure 2, there was also a significant difference in the percentage description between the categories of reported contents (F=45.53, p<0.001). In the last half-year period of the survey (April–September 2004), staff members in this hospital were diligent in describing the type of event involved (100%), situation when the event occurred (84%), lessons learned (81%), and causes of errors (76%) in their incident reports, while only a quarter of reports (27%) included the description of contextual conditions.

Integrating the results mentioned so far, the reporting culture of this hospital seems to have improved gradually over the period that was sampled since staff members reported their errors more promptly and more properly.

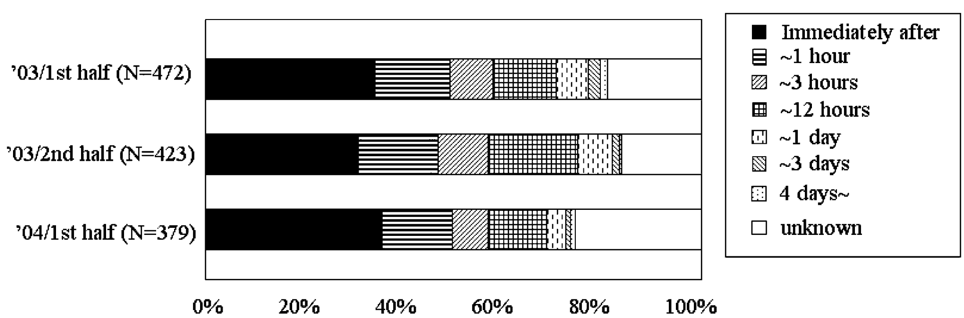

Figure 1. Periodic changes in percentage of incident cases based on the timing of reporting (Hospital A).

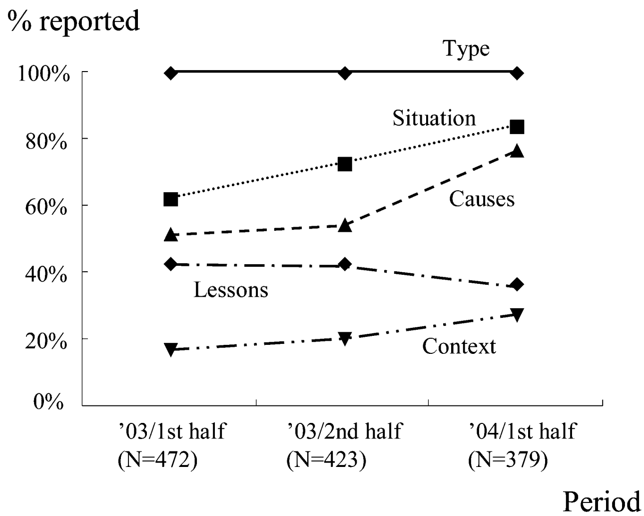

Figure 2. Periodic changes in percentage of descriptions included for each category of reporting content (Hospital A).

4.2 Department-based severity

Regarding the outcome severity of events, Figure 3 indicates percentages of each severity level for five departments in Hospital A, i.e., clinic, nursing, pharmacy, technical and administration departments. In each department corresponding there is a single professional group, i.e., doctors, nurses, pharmacists, technicians, or administrators. We did not apply a statistical test to these department-based data

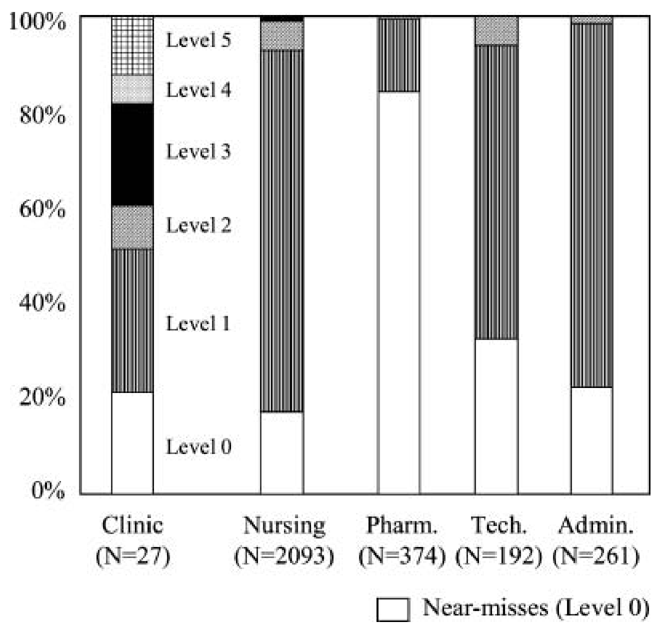

Figure 3. Department-based reporting percentage for each level of severity (Hospital A).

due to very small samples of reports collected from clinical and administration departments. Fifteen incident cases with 'severity level' 3 or higher were reported from doctors, corresponding to 22% of reports from the clinical department. From the nursing department 0.7% of Level 3 or higher were reported (16 cases). Figure 3 demonstrates the other extreme of severity-level distribution in that more than 80% of cases from the pharmacy department were 'near-misses' (Level 0 incidents) reported. The other three departments were located in the mid-point between these two extremes: the majority of incidents were Level 1 in which no adverse consequences arose for the patient.

Comparing the rate of near-miss reporting between the departments (Figure 3) we see that the value of this index for the nursing department was the lowest, was slightly higher for the clinic department, and was the highest for the pharmacy department. This, however, does not necessarily imply that the safety level of the low-scoring departments is lower than that of the high-scoring, since the reporting rate of incidents may be influenced by both internal and external factors, as mentioned previously. Nurses usually work longer hours with patients than any other healthcare professional group, while doctors may perform more hazardous tasks with patients than others. In addition, as also pointed out in Section 1, incident reporting is selective. In particular, the reporting rate of incidents of doctors is lower than that of other groups (Beckmann et al. 1996). According to our interviews with risk managers, there is a tendency to view doctors as likely to hold back on reporting near-misses or incident cases that have no adverse effects on patients. This may be reflected in the distribution of the 27 cases from the clinical department in Figure 3. Therefore, the differences in distribution of severity of cases across departments and professional groups reflect, it may be argued, differences in their tasks and required work routines. For instance, the nature of work in the pharmacy department is such that it is checked repeatedly before a pharmacist's service reaches a patient and, at the same time, the work can be performed routinely following safety procedures. More importantly, incident reporting indices as those computed in this study should be compared only between departments or organisations that have similar tasks and external factors.

4.3 Changes in safety level

In Figure 4 we show the distribution of 'error capture cues' of only near-miss cases from an entire set of incident reports in Hospital A divided half-year periods in the same manner as Figure 1. Distributions of categories in the 'error capture cue' dimension were significantly different between the three half-year periods ($\chi^2 = 20.06$, $p < 0.05$). As clearly illustrated in this figure, the percentage of near-misses detected by safety procedures has increased with the periods. In contrast, there were a smaller proportion of cases in which the error was detected by the staff member involved in the first half of 2004, compared with the earlier periods. Percentages of near-miss cases by other cues or persons, for example, team members, staff in other departments, and patients or their families, did not seem to have changed across the three periods. Thus, the rate of near-miss detection by safety procedures gradually increased. As we have suggested previously, this may be taken to indicate an improvement of safety level over the period sampled in this hospital.

Changes over time in the rate of near-miss detection by safety procedures in the 2.5-year interval for three departments in Hospital A are depicted in Figure 5. The

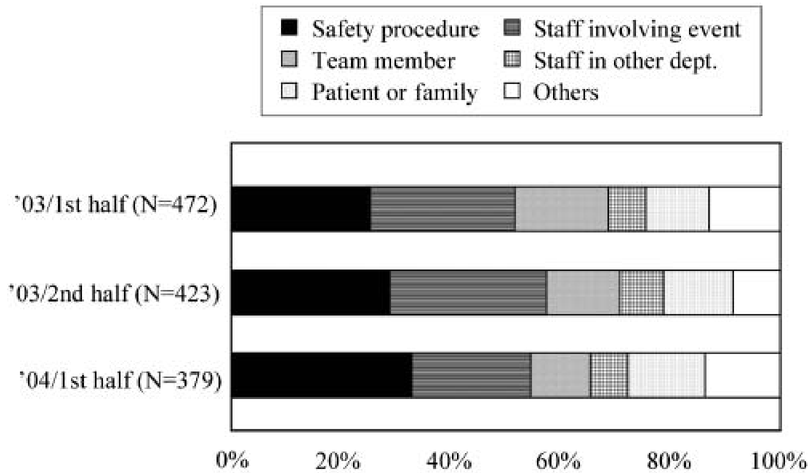

Figure 4. Periodic changes in percentages of error capture cues for near-miss cases every half-year (Hospital A).

departments included are the nursing, the pharmacy and the technical department, whereas the clinical department is excluded due to their relatively small number of near-miss reports. A result of the two-way ANOVA on this index showed that the rate of the pharmacy department was significantly higher than those of the other two departments. The nursing department achieved the lowest percentage throughout the 2.5-year period (F=27.10, p<0.001). As mentioned previously, this might suggest that pharmacists were more likely to work with routinely established safety rules and procedures than other healthcare professionals. There was also a significant

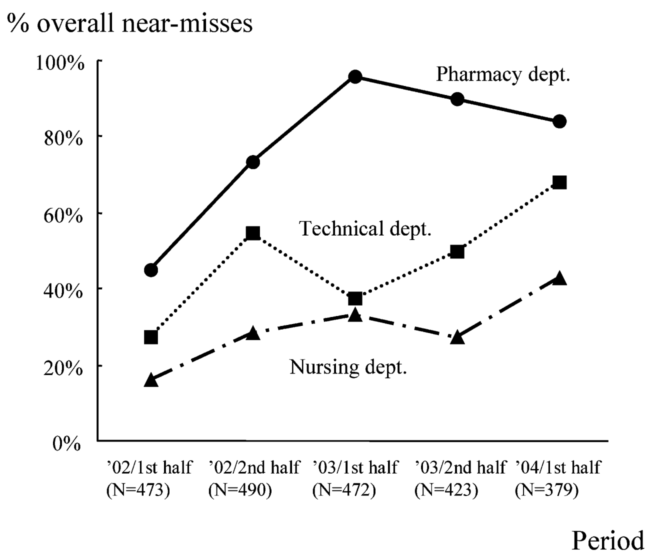

Figure 5. Department-based periodic changes in rate of near-miss detection by safety procedure (Hospital A).

difference in this index between the five half-year periods (F=4.81, p<0.05). As an overall trend, Figure 5 demonstrates that the rate of near-miss detection by safety procedures continuously increased in each department for every half-year period, although this index was slightly reduced in the pharmacy department during the last two half-year periods.

4.4 Organisational comparisons of safety levels

To compare safety performance measures at the hospital level, we applied the taxonomy to a set of incident reports collected from both hospitals. Figure 6 illustrates the changes in the rate of near-miss detection by safety procedure during different 2.5-year periods for Hospitals A (April 2002–September 2004) and B (2

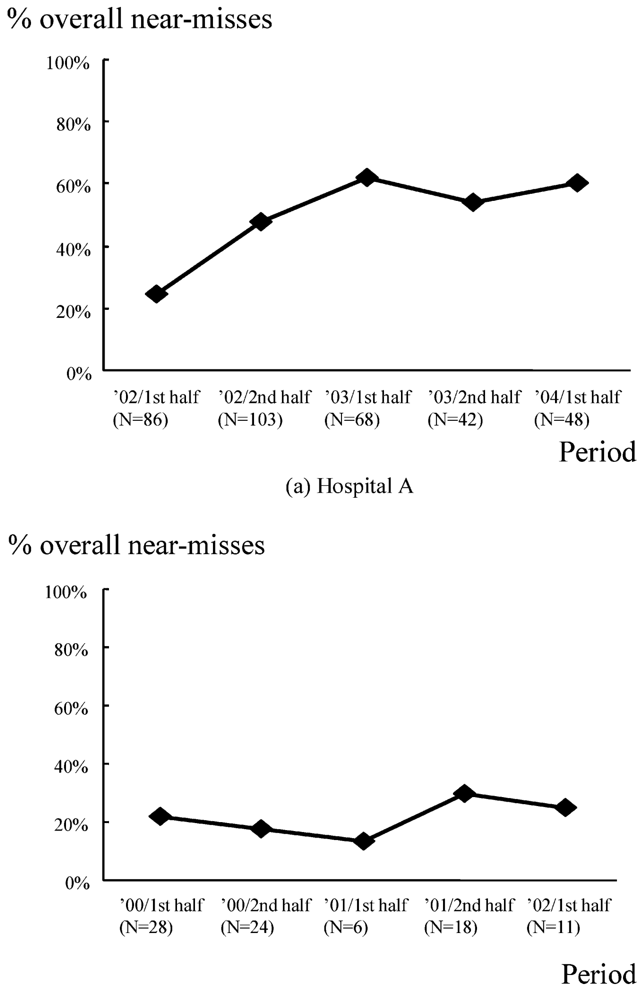

Figure 6. Hospital-based periodic changes in rate of near-miss detection by safety procedure: (a) Hospital A; (b) Hospital B.

years earlier than the survey period of Hospital A). As an overall trend, the values of this index for Hospital A were much higher than those for Hospital B.

In addition, this safety performance index value gradually increased in Hospital A (an 8% average increase per half-year step over the 2.5-year period). We performed the correlation analysis between the rate and the period to examine statistically this trend, and obtained a correlation coefficient of $r=0.86$ ($p=0.06$). In contrast, the rates for Hospital B were nearly constant during the 2.5-year period with a slight increase in the last year (2% increase; $r=0.47$, $p=0.43$). The changes of the two hospitals in this safety performance measure seem to be, at least in part, due to the organisation-wide activities for patient safety. As mentioned in Section 3, continuous patient safety activities may have contributed to a higher and increased rate of near-miss detection by safety procedures in Hospital A, while there was no increase in this index in Hospital B, which, as mentioned, had no special safety-related activities in the survey period.

4.5 Correlations between reporting culture and safety performance

In this sub-section, we discuss relationships between reporting culture and safety performance based on the application results of the taxonomy system to the nurse sample of incident reports collected from Hospital A. Figure 7 illustrates changes observed in the rate of near-miss reporting and the average descriptive depth (see Sections 2.1 and 2.3) for the three half-year periods. While one might expect data to show a relationship between reporting culture and safety performance measures, simple and straightforward statistical analyses did not yield any significant correlation between these measures.

As illustrated in Figure 7, although the average descriptive level increased steadily every half-year period, the index relating to safety level actually decreased from the first to the second period. This discrepancy between the two measures may, it can be speculated, be due to healthcare staff's possible tendency to report in

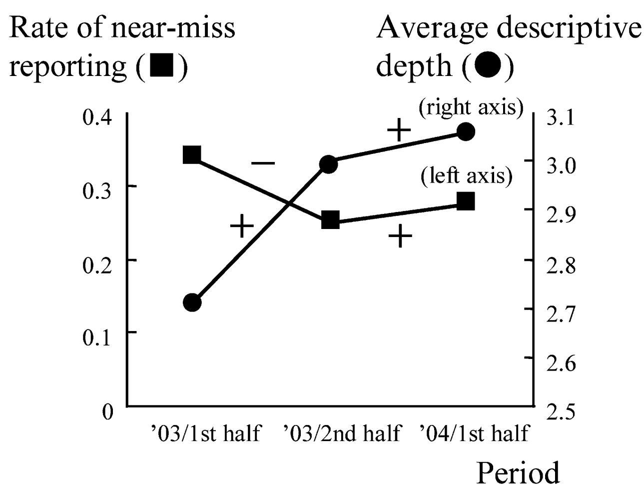

Figure 7. Periodic changes in two hospital-wide measures: rate of near-miss reporting and average descriptive depth.

greater detail events that have a more severe outcome. Also, as mentioned previously, the safety level index may be affected not only by internal factors such as reporting culture but also by various external factors.

Table 6 shows the results of a ward-based application of the same set of incident reports to examine the concordance of the aforementioned two measures in terms of increasing or decreasing trends between the two successive half-year periods. In the table, a positive sign indicates an increase of the given index from the second half of 2003 to the first half of 2004, and a negative sign exhibits the other direction. Nine out of ten wards or work units (including operating room and outpatient staff) improved the average descriptive depth, and the rate of near-miss reporting also increased in eight wards between the two successive half-year periods. The direction of the trend of improvement or degradation of the two indices matched for nine out of ten wards and units. The match between the two incident reporting measures suggests that positive reporting culture contributes to higher level of safety performance in healthcare.

5. Discussion

5.1 Characteristics of healthcare incidents

In this section, we discuss characteristics of incidents in a Japanese healthcare setting based on application results of other dimensions of the taxonomy system (Itoh and Andersen 2004). To highlight the characteristics of Hospitals A and B, the two medium-sized local hospitals of this study, we also collected incident data from another hospital (Hospital C), which is a large university hospital having approximately 1000 beds in a metropolitan area. We applied the taxonomy to a selected sub-set of incident reports from nurses in the three hospitals for a period varying from 2 to 19 months but including overlapping months. The total number of incidents in this sample was 1327 cases, selecting 537, 314 and 476 reports from Hospitals A, B and C, respectively.

Table 6. Ward-based analysis of correlation between safety performance and reporting culture (Hospital A).

Wards/work units	Rate of near-miss	Descriptive depth	Match
W1	+	+	1
W2	+	+	1
W3	+	+	1
W4	+	+	1
W5	+	+	1
W6	−	−	1
W7	+	+	1
W8	−	+	0
OR	+	+	1
Outpatient	+	+	1

Note: +/−, periodic transition from the second half of 2003 to the first half of 2004.

(1) Problem types and actions failed

There is a common trend in the problem types of incidents reported in the three Japanese hospitals we have surveyed while there are also significant differences in distribution of categories (χ^2=110.14, p<0.001). The top four types of 'problems' showed the same order in each hospital: injection-related incidents (21% of reported cases on average), medication incidents (19%), patient falls (16%) and tube troubles (11%).

The proportions of actions failed in reported incidents are shown in Figure 8. There was also a significant difference in categories of the 'action failed' dimension between the hospitals (χ^2=170.93, p<0.001). However, in the nursing department of each hospital approximately half of the errors were made while nurses were (or should be) monitoring patients or while they were manipulating medical equipment or handling materials such as a tube, a syringe or a pump. It was also frequently reported that errors were made when nurses were (or should be) monitoring medical equipment or materials, or checking medicine (6–13% for the three hospitals) and when they were reading or checking patient records or other documents relevant to their tasks (3–13%). However, the percentages of reporting these failed actions varied across the hospitals. A possible reason for the variation might be different work situations or conditions in which staff members were performing their tasks in each hospital.

The aforementioned results correspond well with the chief results of human factors about performance shaping factors, namely that the likelihood of human error depends to a large degree on the ergonomics of the equipment that is used, the procedures for operating equipment and, in general, the specific work routines. Similarly, the rate of failed actions with documents, i.e., checking and transcribing a document, might also be affected by the use (or non-use) of an electronic case record system which, when available, relieves nurses of from transferring information from paper to paper and checking possibly ambiguous hand-written documents.

(2) Error types

The distribution of error types across the three hospitals are depicted in Figure 9. The most frequent error type was errors of omission (41%) – similar to what was found in the case of maintenance activities in other domains, notably in nuclear power plants (Rasmussen 1980), where one found almost the same rate (43%). Nevertheless, the occurrence rate of each error type significantly depended on the problem types (e.g., for Hospital C, χ^2=152.86, p<0.001) and the actions failed (e.g., for Hospital C, χ^2=315.78, p<0.001); and it also differed significantly across the

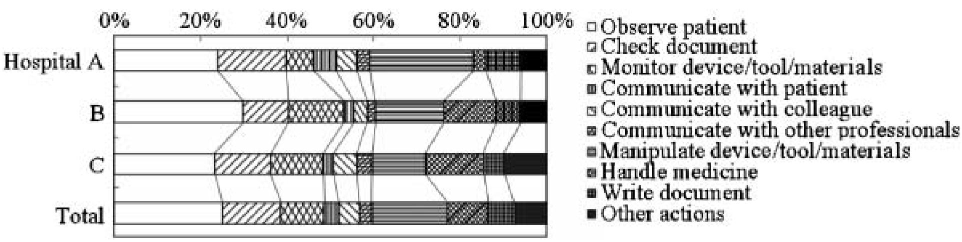

Figure 8. Proportions of actions failed in three Japanese hospitals.

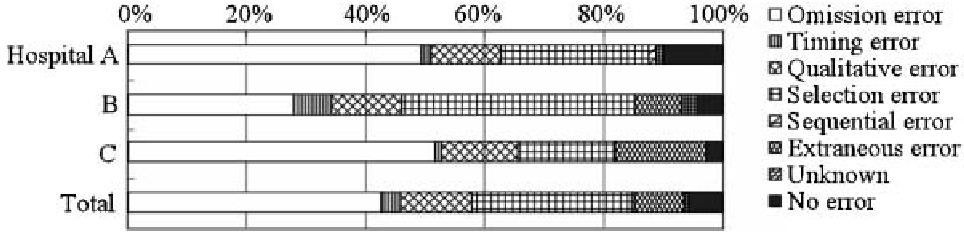

Figure 9. Proportion of human error types that occurred in three Japanese hospitals.

three hospitals (χ^2=135.74, p<0.001). Following errors of omission, selection errors were the next most dominant type among errors of commission – which is divided into timing error, qualitative error and sequential error as well as selection error (see Table 2). The rate of this error type varies greatly between the hospitals: from 16% to 39%.

(3) Contextual conditions

Within the section of contextual conditions, a great number of incident cases turned out to be characterised by multiple 'background' factors. Involvement of each factor in reported events depended significantly on the problem types (χ^2=265.79, p<0.001). Overall distributions of each factor are shown in Figure 10. As in the other dimensions, the rate of each category and its elements varied greatly between the hospitals. More than 60% of the cases involved staff human factors. The staff human factors category most often observed was lack of knowledge (34% on average), followed by wrong assumption or preconception (29%) and psychological factors such as time pressure and confusion (9%).

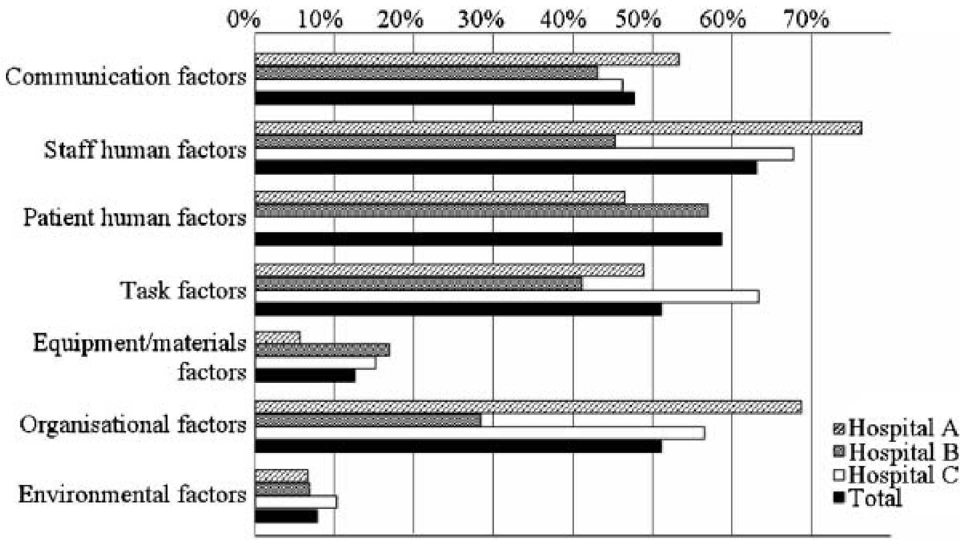

Figure 10. Contextual conditions behind healthcare incidents.

For the patients involved in the incidents of this sub-sample (which, to recall, was submitted by nurses only) human factors issues were also involved in the majority of the cases (57%), and within this category the most often cited problem involved 'patient falls' (94%) and 'tube troubles' (91%). In approximately 50% of the cases, both organisational factors and task factors were involved. The most frequently reported categories in organisational factors were training issues (16%), followed by inappropriate rules and procedures (13%), issues on medical documents (10%) and inaccurate indication and instruction (7%).

5.2 Inter-rater reliability

The reliability of the taxonomy was evaluated in terms of inter-rater reliability between two judges, similar to other studies of taxonomies in various domains (Bove 2002; Kunac et al. 2006; Pedrali, Anderson, and Trucco 2002). If one, and only one, category is selected for each dimension for each incident, a 'raw agreement' rate can be calculated as n/N, where n=observed agreement and N=total possible agreement, which is equivalent to the number of incident cases in our application. Taking into account that part of the agreement between judges will be obtained by chance alone, we also calculated a Kappa statistic (Cohen 1960; Shoukri 2004) as a chance-corrected measure of agreement. The Kappa statistic ranges from −1.0 to +1.0, where a value of 0 indicates that the level of agreement is identical to that expected by chance. Landis and Koch (1977) suggested that a level between 0.41 and 0.60 indicates 'moderate' agreement, 0.61–0.80 'substantial', and above 0.80 'almost perfect' agreement, whereas 0.21–0.40 indicates only fair agreement.

In the taxonomy system developed in the present paper, ratings may be non-exclusive for several dimensions, that is, two or more categories can be assigned to a case. For instance, for a dimension such as 'staff human factors', several factors may simultaneously lie behind one and the same error (e.g., fatigue, experience, workload). More than one error might be made and reported in a single incident case and we can treat these errors as several objects, and will assign error types to all the errors. To accommodate this, we apply the standard procedure of raw agreements and the Kappa statistics whenever we measure agreement within a dimension that requires exclusive (single category) rating. Thus, for dimensions that allow multiple category rating, we define raw and chance-corrected agreement as follows: Let $a_k(i,j)$ (i=1–N; N=the number of cases; and j=1–m; m=the number of categories) be judge k's assignment of category j to case i in a given dimension, where $a_k(i,j)=1$ [=0] means that judge k assigns [does not assign] category j to case i. A matching rate, i.e., raw agreement, between two judges (k=1, 2) for the dimension is defined in Equation (1), where M=1 when $a_1(i,j)=a_2(i,j)$ and M=0 otherwise.

$$\frac{1}{m} \times \frac{1}{N} \sum_{j=1}^{m} \sum_{i=1}^{N} M(a_1(i,j), a_2(i,j)) \tag{1}$$

Further, let $p_k(j)$ be the probability of judge k's assignment of category j to all the cases; and $p'_k(j)$ be the probability of non-assignment, where $p'_k(j)=1-p_k(j)$. A mean chance matching rate for multiple category rating may then be calculated for any dimension as the mean matching rate over all the cases classified by the taxonomy. Therefore, for multiple category rating, the chance agreement is defined as Equation (2).

$$\frac{1}{m}\sum_{j=1}^{m}(p_1(j)\times p_2(j)+p'_1(j)\times p'_2(j)) \tag{2}$$

Accordingly, applying the original Kappa definition, i.e., Kappa=(raw agreement–chance agreement)/(1–chance agreement), we can calculate a chance-corrected measure of agreement for multiple category rating, using the previously defined Equations (1) and (2).

The results of calculating the inter-rater reliability by the aforementioned procedure are shown in Table 6 for the dimensions of the taxonomy system focused on in the present paper. In this evaluation, two judges, a human factors expert (the first author who is not a medical domain expert) and a student in human factors and industrial engineering (the second author), performed category assignments for 138 incident cases which were randomly selected from the Hospital A sample by applying the taxonomy.

As can be seen in Table 7, there were moderate and high chance-corrected agreements between the two judges for three dimensions, near-moderate for two others, and just fair agreement for one dimension, according to the aforementioned Landis and Koch's (1977) criteria. Thus, category assignment within outcome severity was performed perfectly, but only moderately within error capture and the descriptive level. The remaining dimensions achieved only fair agreement: time of reporting, reported content, and time-band of description.

There was a difference in values between raw agreement and chance-corrected agreement for the 'reported content' dimension. The low chance-corrected agreement for this dimension may partly stem from the nature of the definition that not only raw but also chance-corrected agreement are calculated higher – and consequently derives a low chance-corrected agreement – for a dimension that has many categories and some of which are rarely assigned to a case. There were a number of reports that did not describe explicitly when events were detected. Therefore, 'time of event detection' and 'time of reporting' were so vaguely described in some of the reports that agreement should not be expected. The two judges' assignments were not matched well for many of these cases.

Table 7. Inter-rater reliability for dimensions of the taxonomy.

Dimensions	Raw agreement	Chance-corrected
Outcome severity	0.91	0.87
Event capture cue	0.51	0.47
Time of reporting	0.41	0.27
Reported content*	0.58	0.35
Time-band of description*	0.36	0.36
Descriptive level	0.59	0.42

Note: *Calculation of reliability indices based on modified raw agreement and modified chance-corrected measures of agreement.

6. Conclusion

The present paper reported the healthcare incident taxonomy system developed for the special purposes of assessing reporting culture and safety performance as well as understanding characteristics of patient safety events based on incident reports. The feasibility of applying the taxonomy system to risk analysis was illustrated through a case study in which we analysed a number of incident reports collected from two hospitals in Japan. As results of this case study, we identified suggestive, but not statistically significant, evidence that there is a correlation between reporting culture and safety performance indicators, both of which were calculated from the data obtained when using the taxonomy; and we also identified characteristics of incidents and errors in the Japanese healthcare organisations surveyed.

Based on the results obtained in this study, we believe that the taxonomy system can be applied both for single case analysis and for analysing a collection of reports. It can be used to support the identification and analysis of latent causal factors involved in a single case; and it may support the analysis of a number of incident reports, for example, for the purpose of tracking safety level and nature of errors happed in specific work units or an entire organisation, and to identifying latent causal factors behind errors as its analysis framework. In particular, the taxonomy system may be expected to contribute to improving reporting culture if its results are used by leaders and patient safety managers in efforts to enrich the narrative quality of reports. For further extension of the taxonomy and its application to risk management, several additional dimensions are under development for the purpose of assessing 'learning culture', another component of safety culture (Reason 1997).

To perform root-cause analysis efficiently for any adverse event, the analysts need to obtain information on the relationships between errors, situations, performance shaping factors and consequences. For the purpose of supporting healthcare risk managers to conduct such analyses effectively, our challenge would be to develop several kinds of relationship diagrams or tables that represent likely and reliable relationships between dimensions comprising the taxonomy, similar to the consequent-antecedent relations in CREAM (Hollnagel 1998). For instance, as one of the most useful databases, we would suggest detailed lists of strongly connected relations which map from categories in combinations of 'action failed', 'information/equipment' and 'error types' dimensions to individual categories of any 'contextual condition', i.e., communication factors, staff human factors, organisational factors, etc., and then in turn to categories in the 'preventive mechanisms'. To develop such a database, we need to conduct further case studies in which we will apply the taxonomy to a number of incident cases collected from many healthcare organisations. Once we have obtained this kind of database, it could contribute not only to effective identification of latent factors through in-depth retrospective analysis but also to quick and reliable category assignment to the contextual conditions – which are the most difficult to determine when a report is not well described or contains little information. Finally, to improve on the low values of a chance-corrected measure for some dimensions, we need to develop a precise and easy-to-follow guideline for category assignment so that healthcare risk mangers with little human factors expertise can easily and consistently apply the taxonomy system.

In addition to such a user manual, it would be useful, it appears, to create a lesson material for healthcare staff to guide them in preparing reports which in terms of contents and structure lend themselves to critical incident analysis and, ultimately, to providing a basis for organisational learning.

Acknowledgements

This work was in part supported by Grant-in-Aid for Scientific Research A(2) (No. 18201029), Japan Society for the Promotion of Science, funded to the first author. We are grateful to risk managers in the two hospitals who did not only accept us to analyse incident reports but also provided us with valuable medical and nursing information.

References

Aspden, P., J.M. Corrigan, J. Wolcott, and S.M. Erickson, eds. 2004. *Patient safety: Achieving a new standard for care. Committee on data standards for patient safety: Board on health care services*. Washington, DC: The National Academic Press.

Barach, P., and S.D. Small. 2000. Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems *British Medical Journal* 320: 753–63.

Battles, J.B., and R.J. Lilford. 2003. Organizing patient safety research to identify risks and hazards. *Quality and Safety in Health Care* 12, suppl. 2: ii2–ii7.

Beckmann, U., L.F. West, G.J. Groombridge, I. Baldwin, G.K. Hart, D.G. Clayton, R.K. Webb, and W.B. Runciman. 1996. The Australian incident monitoring study in intensive care: AIMS-ICU. The development and evaluation of an incident reporting system in intensive care. *Anaesthesia and Intensive Care* 24, no. 3: 314–19.

Bove, T. 2002. *Development and validation of a human error management taxonomy in air traffic control*. Risø-R-1378(EN). Roskilde: Risø National Laboratory.

Chang, A., P.M. Schyve, R.J. Croteau, D.S. O'Leary, and J.M. Loeb. 2005. The JCAHO patient safety event taxonomy: A standardized terminology and classification schema for near misses and adverse events. *International Journal for Quality in Health Care* 17, no. 2: 95–105.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurements* 20: 37–46.

Department of Health. 2000. *An organisation with a memory: Report of an expert group on learning from adverse events in the National Health Service chaired by the chief medical officer*. London: HMSO.

Dovey, S.M., D.S. Meyers, R.L. Phillips, Jr., L.A. Green, G.E. Fryer, J.M. Galliher, J. Kappus, and P. Grob. 2002. A preliminary taxonomy of medical errors in family practice. *Quality and Safety in Health Care* 11: 233–8.

Edmondson, A.C. 1996. Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science* 32, no. 1: 5–28.

———. 2004. Learning from failure in health care: Frequent opportunities, pervasive barriers. *Quality and Safety in Health Care* 13, suppl. 2: ii3–ii9.

Hammons, T., N.F. Piland, S.D. Small, M.J. Hatlie, and H.R. Burstin. 2003. Ambulatory patient safety: What we know and need to know. *Journal of Ambulatory Care Management* 26: 63–82.

Helmreich, R.L., and A.C. Merritt. 1998. *Culture at work in aviation and medicine: National, organizational and professional influences*. Aldershot, UK: Ashgate.

Hollnagel, E. 1998. *Cognitive reliability and error analysis method (CREAM)*. London: Elsevier Science.

Isaac, A., S.T. Shorrock, R. Kennedy, B. Kirwan, H.B. Andersen, and T. Bove. 2001. *The human error in ATM (HERA) technique*. Brussels: Eurocontrol Headquarters.

Isaac, A., S.T. Shorrock, and B. Kirwan. 2002. Human error in European air traffic management: The HERA project. *Reliability Engineering and System Safety* 75, no. 2: 257–72.

Itoh, K. 2003. Safety culture and its correlation with patient safety in Japanese hospitals: Proceedings of the 7th International Symposium on Human Factors in Organizational Design and Management (ODAM 7), Aachen, Germany, October 2003. In *Human factors in organizational design and management*. Vol. 4: ed. H. Luczak, and K.J. Zink, 35–45. Santa Monica, CA: IEA Press.

Itoh, K., and H.B. Andersen. 2004. Analysing medical incident reports by use of a human error taxonomy adapted from air traffic management: Proceedings of the 7th International Conference on Probabilistic Safety Assessment and Management – PSAM7-ESREL'04, Berlin, Germany, June 2004. In *Probabilistic safety assessment and management*. Vol. 5: ed. C. Spitzer, U. Schmocker, and V.N. Dang, 2714–19. London: Springer.

Itoh, K., H.B. Andersen, and M.D. Madsen. 2006. Safety culture in healthcare. In *Handbook of human factors and ergonomics in healthcare and patient safety*, ed. P. Carayon, 199–216. Mahwah, NJ: Lawrence Erlbaum Associates.

Itoh, K., H.B. Andersen, M.D. Madsen, and T. Abe. 2005. *Comparative results of cross-national surveys on hospital safety culture: The views and attitudes of healthcare staff towards safety-related issues and reporting of adverse events and errors*. Technical Report No.2005-1. Tokyo: Tokyo Institute of Technology.

Itoh, K., M. Seki, and H.B. Andersen. 2003. Approaches to transportation safety: Methods and case studies applying to track maintenance train operations. In *Handbook of cognitive task design*, ed. E. Hollnagel, 603–32. Mahwah, NJ: Lawrence Erlbaum Associates.

Kohn, L.T., J.M. Corrigan, and M.S. Donaldson, eds. 1999. *To err is human: Building a safer health system*. Washington DC: National Academy Press.

Kostopoulou, O. 2006. From cognition to the system: Developing a multilevel taxonomy of patient safety in general practice. *Ergonomics* 49, no. 5: 486–502.

Kostopoulou, O., and B. Delaney. 2007. Confidential reporting of patient safety events in primary care: Results from a multilevel classification of cognitive and system factors. *Quality and Safety in Health Care* 16: 95–100.

Kunac, D.L., D.M. Reith, J. Kennedy, N.C. Austin, and S.M. Williams. 2006. Inter- and intra-rater reliability for classification of medication related events in paediatric inpatients. *Quality and Safety in Health Care* 15: 196–201.

Landis, J., and G.G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 327–47.

Madsen, M.D., H.B. Andersen, and K. Itoh. 2006. Assessing safety culture in healthcare. In *Handbook of human factors and ergonomics in healthcare and patient safety*, ed. P. Carayon, 693–713. Mahwah, NJ: Lawrence Erlbaum Associates.

Makeham, M.A.B., S.M. Dovey, M. County, and M.R. Kidd. 2002. An International taxonomy for errors in general practice: A pilot study. *The Medical Journal of Australia* 177: 68–72.

Mason, S. 1997. Procedural violations: Causes, costs and cues. In *Human factors in safety critical systems*, ed. F. Redmill, and K.J. Rajan, 287–318. Oxford, UK: Butterworth-Heinemann.

Pedrali, P., H.B. Andersen, and P. Trucco. 2002. Are maritime accident causation taxonomies reliable? An experimental study of the human factors classification of the IMO, MAIB and CASMET Systems. In *Proceedings of the RINA Conference on Human Factors in Ship Design and Operation*, 117–129. London: RINA.

Rasmussen, J. 1980. What can be learned from human error reports? In *Changes in working life*, ed. K. Duncan, M. Gruneberg, and D. Wallis, 97–113. London: Wiley.

Rasmussen, J. 1986. *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: North-Holland.
Reason, J. 1990. *Human error*. New York: Cambridge University Press.
———. 1993. Managing the management risk: New approaches to organisational safety. In *Reliability and safety in hazardous work systems*, ed. B. Wilpert, and T. Qvale. Hove, UK: Lawrence Erlbaum Associates.
———. 1997. *Managing the risk of organizational accidents*. Aldershot, UK: Ashgate.
Reason, J., and A. Hobbs. 2003. *Managing maintenance error: A practical guide*. Aldershot, UK: Ashgate.
Runciman, W.B., S.C. Helps, E.J. Sexton, and A. Malpass. 1998. A classification for incidents and accidents in the health-care system. *Journal of Quality in Clinical Practice* 18: 199–211.
Runciman, W.B., J.A.H. Williamson, A. Deakin, K.A. Benveniste, K. Bannon, and P.D. Hibbert. 2006. An integrated framework for safety, quality and risk management: An information and incident management system based on a universal patient safety classification. *Quality and Safety in Health Care* 15, suppl. 1: i482–i490.
Salo, I., and O. Svenson. 2003. Mental causal models of incidents communicated in licensee event reports in a process industry. *Cognition, Technology & Work* 5: 211–17.
Shorrock, S.T., and B. Kirwan. 2002. Development and application of a human error identification tool for air traffic control. *Applied Ergonomics* 33, no. 4: 319–36.
Shoukri, M.M. 2004. *Measures of interobserver agreement*. Boca Raton, FL: Chapman & Hall/CRC.
Swain, A.D., and H.E. Guttmann. 1983. *A handbook of human reliability analysis with emphasis on nuclear power plant applications*. NUREG/CR-1278; SAND-80-0200. Albuquerque, NM: Sandia National Laboratories.
Taylor-Adams, S.E., and B. Kirwan. 1995. Human reliability data requirements. *Journal of Quality and Reliability Management* 12, no. 1: 24–46.
Van der Schaaf, T.W. 1996. PRISMA: A risk management tool based on incident analysis. Paper presented at the International Conference and Workshop on Process Safety Management and Inherently Safer Processes, 8–11 October, in Orlando, FL.
Van der Schaaf, T.W., and M.N.P. Habraken. 2005. *PRISMA-medical: A brief description*. Eindhoven: Eindhoven University of Technology.
WHO. 2005. *WHO draft guidelines for adverse event reporting and learning systems*. Geneva: World Alliance for Patient Safety, WHO.
———. 2006. *The conceptual framework of an international patient safety event classification*. Geneva: World Alliance for Patient Safety, WHO.
Woods, D.M., J. Johnson, J.L. Holl, M. Mehra, E.J. Thomas, E.S. Ogata, and C. Lannon. 2005. Anatomy of a patient safety event: A pediatric patient safety taxonomy. *Quality and Safety in Health Care* 14: 422–7.
Zhang, J., V.L. Patel, T.R. Johnson, and E.H. Shortliffe. 2004. A cognitive taxonomy of medical errors. *Journal of Biomedical Informatics* 37, no. 3: 193–204.