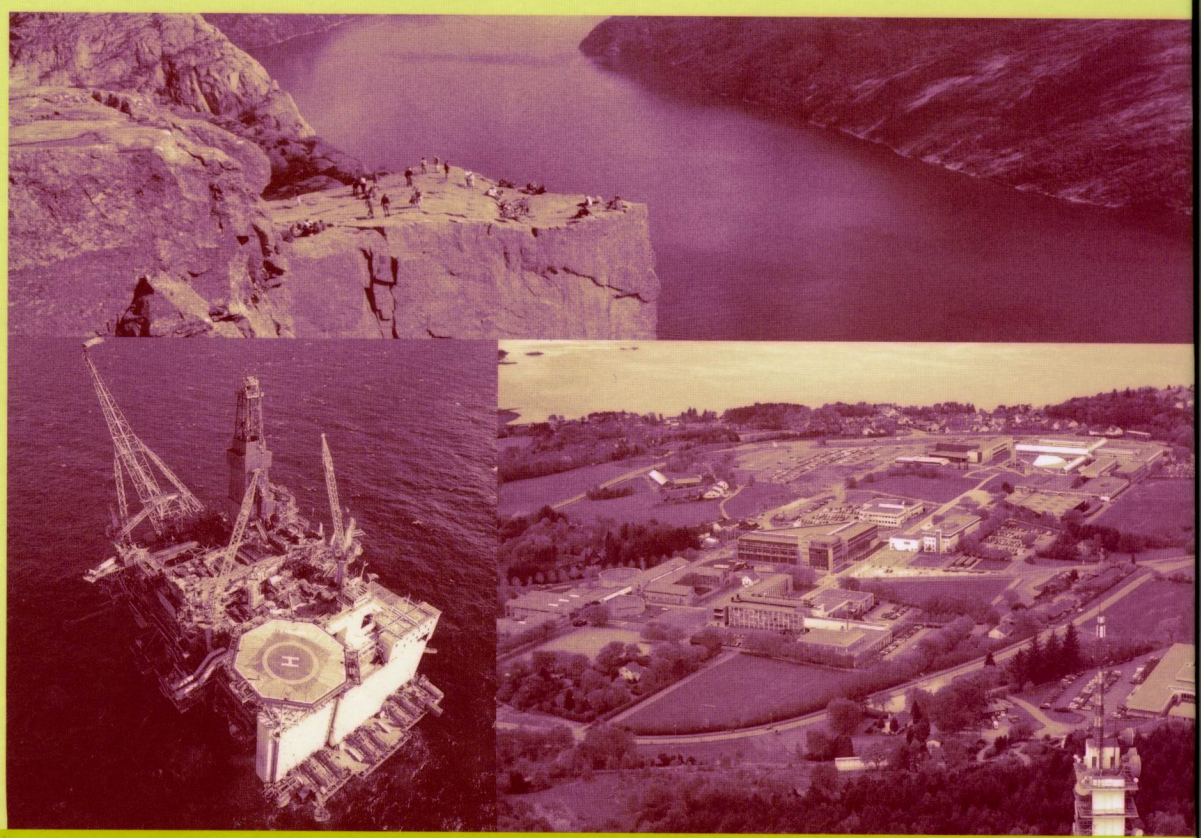
Risk, Reliability and Societal Safety

volume 1: Specialisation Topics

Terje Aven & Jan Erik Vinnem EDITORS



Taylor & Francis
Taylor & Francis Group

Risk, Reliability and Societal Safety – Aven & Vinnem (eds)
© 2007 Taylor & Francis Group, London, ISBN 978-0-415-44786-7

A human error taxonomy for analysing healthcare reporting culture and its effects on safety performance

K. Itoh & N. Omata
Tokyo Institute of Technology, Tokyo, Japan

H.B. Andersen
Risø National Laboratory, Roskilde, Denmark

ABSTRACT: In the present study, we developed an extended human error taxonomy system for healthcare risk management, aiming to apply it to evaluating safety performance and reporting culture in specific organisations. The taxonomy system contains, besides categories to classify errors and performance shaping factors, additional dimensions that seek to capture the maturity of case descriptions contained in incident reports. These dimensions are reporting content, time-band of description, and description level. Applying several dimensions in the taxonomy, we proposed two safety performance measures which can be calculated from assignments to several dimensions in the taxonomy system: rate of near-miss reporting; and rate of near-miss detection by safety procedure. The present paper reports application results of the extended taxonomy to a total of 3,749 incident cases collected from two Japanese hospitals, which were at different stages of patient safety activities.

1 INTRODUCTION

A variety of approaches and methods for risk management have been applied to patient safety issues in healthcare just as in modern human-machine system areas 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), similar to human-machine operation domain (e.g., Itoh et al., 2003). Thus, it is not numbers of cases that are required, but stories from which we can learn; i.e., narratives contained in incident reports. For these requirements, a *human error taxonomy* can be critically useful to support the analysis of incident reporting to enhance systematic organisational learning.

On this background, Itoh & Andersen (2004) developed a human error taxonomy system for analysing healthcare incident reports, partly on the basis of a system used in air traffic management (Isaac et al., 2001; 2002; Shorrock & Kirwan, 2002). The primary purpose of this taxonomy system is its focus on identifying risk factors involved in medical accidents as well as analysing characteristics of the associated human and organisational failures. In addition to these analyses, it is also of importance to diagnose and track current levels of risk or safety for given work units and the overall organisation to which they belong. It may also be possible to calculate several types of reporting rates, e.g., ones applying to all reports and or to their subsets selected from specific classes of severity or specific case descriptions as measures connecting to safety performance level. However, there is a difficulty for a valid estimate of incidence of different types of events by these measures mainly due to the fact that incident reporting is selective (Battles & Lilford, 2003).

Thus, reporting rates of incidents or accidents may be essentially dependent on external factors. Hence, reporting rates for different types of incidents may not reflect internal processes or risk factors (Itoh et al., 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 frequently *latent factors* that contribute or even lead to the occurrence of human error made by frontline personnel (Reason, 1997).

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 (compare Helmreich & Merritt, 1998; Reason 1997). To establish a "safe" organisational culture, Reason (Reason, 1997; Reason & Hobbs, 2003) suggested that an organisation must improve the following three inter-connected component cultures: (1) a *reporting culture* – an organisational climate in which people are prepared to report their errors and near-misses; (2) a *just culture* – an atmosphere of trust in which people are encouraged, even rewarded, for providing essential safety-related information; and (3) a *learning culture* – a willingness and the competence needed to draw right conclusions from safety information system 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.

In the present paper, we extend the previous framework of human error taxonomy to an analysis system of incident reports so that one can evaluate the present status of safety performance and the level of reporting culture in a specific work unit or organisation. We believe that the above-mentioned difficulty or limitation to using incident reporting data for risk prediction can be overcome by applying the taxonomy to estimating safety culture status as well as characteristics of incidents which occurred in an organisation. We applied the extended taxonomy to analysis of incident reports collected from two Japanese hospitals. Detailed application results are mentioned below, in particular, analysis of reporting culture status and its relation or contribution to the level of safety performance in healthcare organisations. Reliability of the taxonomy is also discussed based on calculation of inter-rater assignments.

2 TAXONOMY FOR REPORTING CULTURE

2.1 *Diagnosing indices of reporting culture*

The original healthcare human error taxonomy (Itoh & Andersen, 2004) was developed by partly adapting a classification scheme developed for analysing aviation near-misses, the HERA (Human Error in ATM) taxonomy (Isaac et al., 2001; 2002; Shorrock & Kirwan, 2002). The healthcare taxonomy was extended to include an additional section concerning the maturity of case description with several minor revisions in the original sections. In this additional section, the following three dimensions were included: report content, time-band of description, and description level.

We suggest that an important indicator of the level of reporting culture in a particular unit or organisation can be based on incident reports submitted during a specific period, applying the three dimensions mentioned above.

By "report content" we refer to properties of the case as described in the incident report: the type of event involved (e.g., medication, patient fall, equipment, injection), background factors (situational context, e.g., busy ward, nightshift), causes of errors, contextual conditions behind errors (e.g., descriptions about human factors, organisational factors, and poor communication), and lessons learned.

In the dimension of time-band of description, we identify the time or tense of things which happened or are expected or look-ahead from statements contained the incident report, i.e., things occurred in the past (long before the event occurred), just before the event, present (when the event occurred), immediately after the event, and expected to occur in the future (situation to be expected).

The description level of a report concerns the depth or thoroughness of the report and is rated on a five-point scale, ranging from 1 to 5: no information contained (1), very briefly described (2), described at more length but superficially (3), described in detail (4), and described in depth (5). As one of the several indicators of the level of the reporting culture, that for a specific organisation or department/ward is then (in part) characterised in terms of percentage description of each "content" category over the entire set of incident reports submitted during a given period. The average descriptive depth will be calculated from the above-mentioned five point values of description level assigned to all written reports. This index will be also used as a rough measure of the reporting culture level of a specific work unit.

2.2 *Safety performance indices*

Concerning safety performance, we suggest that the reporting "rate" calculated by a specific subset of the submitted incident reports such as one selected on the basis of severity level or one captured by a particular cues could be possible indicators of safety level for an entire organisation or a given work unit. Of course, 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 incident rate. In general, the incident rate 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 the inverse of risk, that is, as an index of *safety* – that is, the more that staff is demonstrating willingness

to report, the greater is their sensitivity to errors, the greater the learning potential, and so, the greater is the safety in their department.

It has often been observed that the incident rate may be read as a "safety" index – more frequent reporting of incidents is a sign of safety (Edmondson, 1996; 2004). We make a distinction between near-miss cases in which a wrong act had no consequences because it did not reached a patient (categorised Level 0 in the outcome severity classification) and incidents at the other severity level. Error reporting for the former type of "near-miss" events can therefore be considered to reflect the staff's safety awareness or sensitivity to errors or incidents. Based on this expectation, we introduce the percentage of near-miss cases over all 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 incidents in terms of the type of "error capture cues" that were used to detect or capture the error, e.g., a wrong act detected by a safety rule or procedure which was built into the work system can be expected to avoid similar incidents regularly. In contrast, errors may sometimes remain undetected, or the error may be caught or detected by chance. Thus, the rate of error detection cases by the built-in safety procedure over all near-miss cases reported – which is referred to as the rate of near-miss detection by safety procedure – would be another promising safety performance indicator, we propose.

As another index pertaining to safety performance, the rate of Level 3 or higher cases can be calculated by applying the dimension of "severity of outcome", in which accidents and incidents are classified into six levels, ranging from 0 to 5 – Level 0 means a near-miss case, and Level 5 indicates death. Usually, cases at Levels 0 through 2 are regarded as incidents in Japanese hospitals, whereas a Level 3 case is an event that actually or potentially involves or leads to a minor patient injury – which may require an extended hospitalisation of a few more weeks at most. According to a number of interviews with risk managers and leaders of nursing department, staff has nearly always – and nurses definitely always – reported severe cases of Level 3 or higher to their hospital's incident reporting system. Therefore, an index such as the one discussed may be a plausible indication of "risk": a greater number of errors take place in a workplace with a higher rate of this index since all incident cases are reported.

The present study applied the first two safety performance indices, i.e., the rate of near-miss reporting and rate of near-miss detection by safety procedure, to two hospital samples of incident reports. The other index was not adopted since at most a single case with Level 3 or higher had occurred in each of the departments or wards in involved.

Table 1. Collected reports and hospital attributes surveyed.

			No. of beds & staff		
Collected	Reports	Survey period	Beds	Dr.	Ns.
Hospital					
A	3,162	4/2002~9/2004	347	60	310
B	587	4/2000~9/2002	475	70	314
Total	3,749				

3 APPLIED CASES OF THE TAXONOMY

We applied the extended taxonomy to a total of 3,749 incident reports collected during a two-and-a-half year period from two hospitals located in different rural regions of Japan. The number of collected reports, survey periods and other hospital data are shown in Table 1. According to our interviews with the risk managers of both hospitals, organisation-wide initiatives for patient safety were initiated several years before the survey period, and in particular, special attention has 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. Difference in risk management and reporting culture between the hospitals, for instance, might well lie behind their annual reporting rates of incidents per nurse: 3.05 (Hospital A) vs. 0.65 (B).

In the present paper, we focus on applying results of the taxonomy to the incident reports collected from Hospital A. Results of the hospital-based analysis are also mentioned to make organisational comparisons of safety outcome, using the incident cases from both hospitals.

4 APPLICATION RESULTS

4.1 *Periodic changes in reporting culture*

In a healthcare organisation having a positive reporting culture we may expect that a staff member will submit an incident report unhesitatingly if he or she has become involved in any reportable event. To examine this expectation, we illustrate in Figure 1 when incident cases were reported to the local reporting system in Hospital A in terms of duration from the event detection to the report submission based on three half-a-year periods between April 2003 and September 2004. A significant difference was observed by the Chi square test in the timing of event reporting between the three half-year periods ($\chi 2 = 28.75, p < 0.05$). Compared with the earlier two periods, the percentage of reports made immediately after the event became greater and that of reporting after an hour was smaller in the last period of the survey. Approximately half

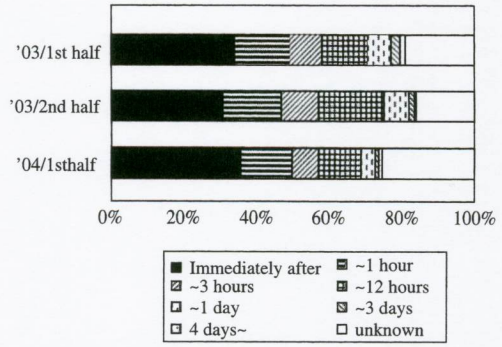


Figure 1. Periodic changes in percentage reporting of incident cases based on the timing from event detection (Hospital A).

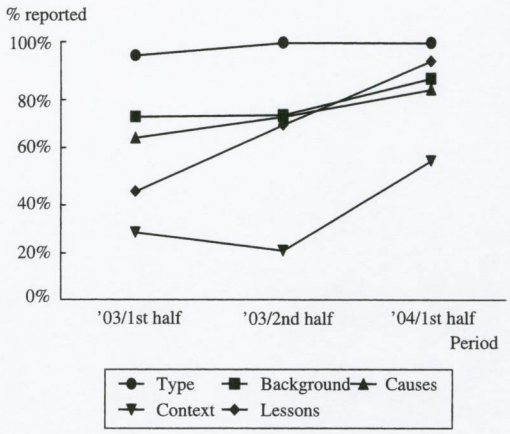


Figure 2. Periodic changes in percentage of descriptions included each category of reporting content (Hospital A).

of the incidents were reported within three hours after detection of the event although no information on submission time was found in approximately a quarter of reports.

To address another aspect of the reporting culture, namely, whether incidents were "properly" reported, we analysed reports for every half-year period by applying the "report content" dimension of the taxonomy. Figure 2 depicts periodic changes in percentages of description of each category of reporting content, applying to incident cases which occurred in Hospital A. The percentage of description for each reporting category was significantly increased every half-a-year period ($F = 4.42$, $p < 0.05$; applying the two-way ANOVA of the periods and the categories of reporting content). As illustrated in this figure, there was also a significant difference in the percentage of description between the reporting content categories ($F = 45.53$,

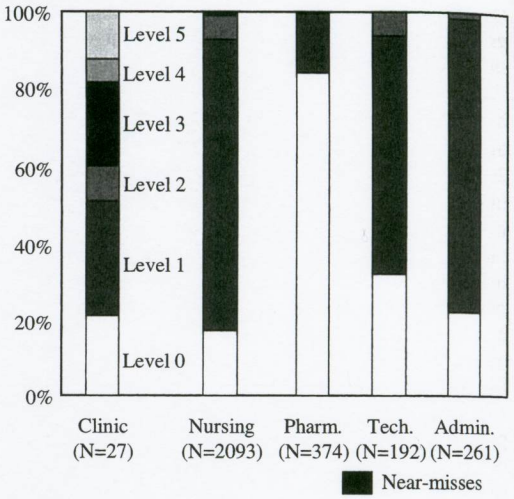


Figure 3. Department-based reporting percentage for each level of severity (Hospital A).

$p < 0.001$). In the last half-year period of the survey (April-September 2004), healthcare staff in this hospital were diligent in describing the type of event involved (100%), background factors (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 of errors.

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

4.2 *Department-based severity*

Regarding the severity of events collected from Hospital A, Figure 3 indicates percentages of each severity level of incidents for five departments, i.e., clinic, nursing, pharmacy, technical and administration departments. We did not apply a statistical test to these department-based data due to very small samples of reports from clinic and administration departments. Incident cases with severity index >2 were reported only from the clinic department where its proportion was 22% – a small number of severity level 2+ cases, i.e., 0.7% of proportion, were also reported from the nursing department. Figure 3 demonstrates the other extreme of severity distribution that more than 80% of cases were near-misses, i.e., Level 0 incidents, reported from the pharmacy department. 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 consequence reached the patent.

Comparing the rate of near-miss reporting between the departments, the value of this index for the nursing department is the lowest, it is slightly higher for the clinic department, and is 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. For instance, nurses usually work longer hours with patients than any other healthcare group, while doctors may perform more hazardous tasks with patients than other professionals. In addition, as also pointed out in Section 1, incident reporting is selective. In particular, reporting rate of incidents are lower from doctors than other healthcare 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 with no adverse effects to patients. Therefore, it may be safer to interpret the differences in reporting with a view to differences in tasks and work routines. For instance, the nature of the 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 between departments or organisations having identical or similar external factors, and should be tracked within a department or an organisation across several periodic intervals.

Periodic changes in the reporting indices within the same organisation or work unit were analysed by applying the safety performance measures mentioned in Section 2.2 to the hospital-wide set of reports including its subsets selected from each department. Results of periodic analysis will be mentioned in the next subsection.

4.3 *Periodic changes in safety level*

Selecting only near-miss cases from an entire set of incident reports in Hospital A, proportions of "error detection cues" are depicted in Figure 4 based on half-year periods in the same manner as Figure 1. Frequencies of categories that represent the "error detection 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 over the periods. In contrast, the percentage of error detection by the staff member involved in the near-miss case became smaller in the first half of 2004, compared with that in the earlier periods. Percentages by other detection cues or persons, e.g., team members, staff in other departments, and patients or

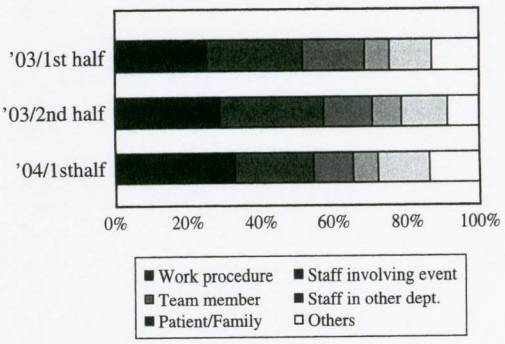


Figure 4. Periodic changes in percentages of error capture cues for near-miss cases every half a year (Hospital A).

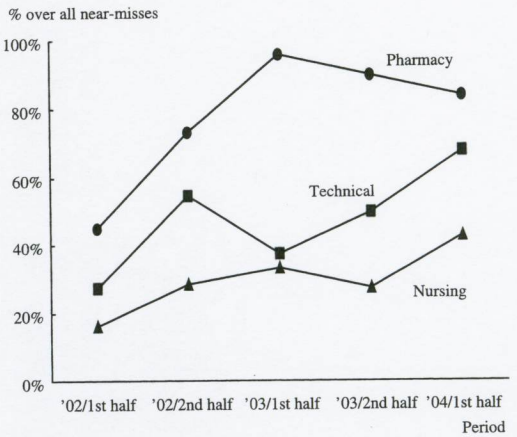
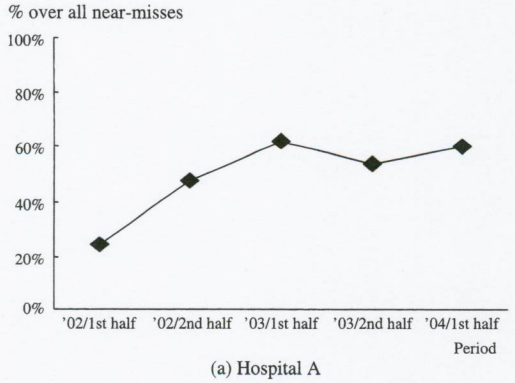


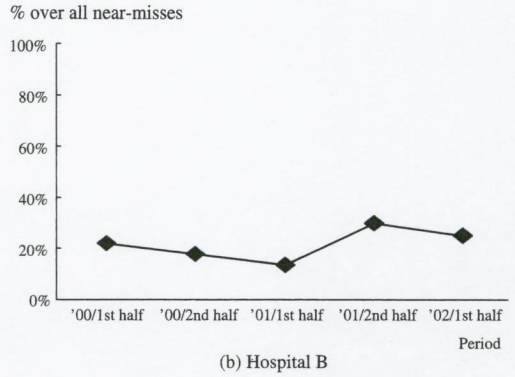
Figure 5. Department-based periodic changes in rate of near-miss detection by safety procedure (Hospital A).

their families, did not seem to have changed throughout the entire one-and-a-half-year period. Thus, the rate of near-miss detection by safety procedures had been gradually improved over the period in this hospital. Recalling this issue, we discuss the hospital-wide safety level in the next subsection, comparing this with the level of the other hospital surveyed.

Changes over time in the rate of near-miss detection by safety procedures during two and a half years for three departments in Hospital A are depicted in Figure 5. The departments included are the nursing, the pharmacy and the technical department, whereas the clinical department is excluded due to its very small number of near-miss reports A two-way ANOVA analysis of this index, the rate of near-miss detection by safety procedure, showed that the pharmacy was significantly higher than the other two departments. The nursing department achieved the lowest percentage throughout the two-and-a-half year period ($F = 27.10, p < 0.001$). As mentioned previously, this



(a) Hospital A



(b) Hospital B

Figure 6. Comparison of periodic changes in rate of near-miss detection by safety procedure between two hospitals.

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

4.4 *Organisational comparisons of safety levels*

To compare safety performance measures at the overall organisational level, we applied the extended taxonomy to a set of incident reports collected from the other hospital, B. Figure 6 illustrates the periodic changes in the rate of near-miss detection by safety procedure during a different two-and-a-half year period for both Hospitals A (between April 2002 and September 2004) and B (two 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 the other hospital.

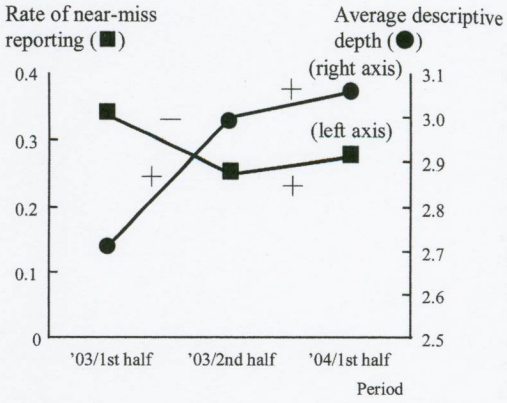


Figure 7. Periodic changes in and correlation between two hospital-wide measures: rate of near-miss reporting and average descriptive depth.

In addition, this safety performance rate was gradually increased in Hospital A (an 8% average increase every half year during the 2.5 year survey period). We analysed the correlation between the rate and the period to statistically examine this trend, finding a correlation coefficient of $r = 0.86$ ($p = 0.06$). In contrast, the rates for Hospital B during the two-and-a-half year period were nearly constant with a slight increase in the last year (2% increase; $r = 0.47, p = 0.43$). The trends in the periodic changes in this measure of the two organisations seem to be, at least in part, due to the hospital-wide activities for patient safety in organisation. As mentioned in Section 3 in connection with the review of some results of interviews with the risk managers, continuous patient safety activities may have contributed to a higher and increased rate of near-miss detection by safety procedures in Hospital A compared to Hospital B where no special safety-related activities took place and no increase in this rate appeared during the survey period.

4.5 *Correlations between reporting culture and safety performance*

In this subsection, we discuss relationships between reporting culture and safety performance based on the applications of the taxonomy system. Figure 7 illustrates changes over the periods observed in the rate of near-miss reporting and the average descriptive depth (cf. Section 2.1) for the three half-year periods in Hospital A. While one might expect data to show a relationship between reporting culture measures and safety level measures, simple and straightforward statistical analyses did not yield a significant correlation between these two measures. As can be seen 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

Table 2. Ward-based analysis of correlation between safety performance and reporting culture.

Wards/ Sections	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

Sign (+/−): Periodic transition from second half of 2003 to the first half of 2004.

to the second period. This discrepancy between the two measures may, it can be speculated, be due healthcare staff possible tendency to report in greater detail when an incident has a more severe outcome. Also, as mentioned previously, the safety level may be affected not only by internal factors such as reporting culture but also by various external factors.

Table 2 shows the results of a ward-based analysis to examine concordance of the above-mentioned two measures in terms of increasing or decreasing trend between the two successive periods, using incident reports selected only from nurses. 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 (including operating room and outpatient staff) improved the descriptive depth, and the rate of near-miss reporting increased in eight wards during the successive two half-a-year periods. The trend of improvement or degradation of the two indices matched for nine out of ten wards. This may suggest that positive reporting culture contributes to higher level of safety performance in healthcare.

4.6 *Inter-rater reliability*

Reliability of the taxonomy was evaluated on basis of raw agreement and chance corrected measure of agreement, applying the same procedure used in the former study (Itoh & Andersen, 2004) and in Bove (2002). Two judges, a human factors expert (not a medical domain expert) and an undergraduate student in human factors and industrial engineering, performed category assignments for 138 incident cases which were randomly selected from the Hospital A sample by applying the taxonomy. Raw agreements and chance corrected agreements were calculated using category assignments rated by the two judges.

Table 3. Reliability evaluation of ratings with the taxonomy.

Dimensions	Raw agreement	Chance corrected
Severity	0.91	0.87
Event capture cue	0.51	0.47
Time of detection	0.41	0.27
Reporting content*	0.58	0.35
Time-band of description*	0.36	0.36
Description level	0.59	0.42

* Calculation of reliability indices were applied modified raw agreement and chance corrected measures of agreement (Itoh & Andersen, 2004).

The results of calculating the inter-rater reliability (Cohen's *kappa*) are shown in Table 3 for the additional dimensions of the taxonomy extended in this study. As shown in this table, there were moderate and high chance corrected agreements between the two judges for three, near-moderate for two others, and slight agreement for one. Landis & Koch (1977) suggest that levels between 0.41 and 0.60 indicate 'moderate' agreement, 0.61–0.80 'substantial', and above 0.80 'almost perfect' agreement, whereas 0.21–0.40 only fair agreement. Thus, the values of the chance corrected measures indicate that the extended taxonomy did not obtain an acceptable level of agreement for several additional dimensions: time of detection, reporting content, and time-band of description.

5 CONCLUSION

The present paper reported the additional dimensions of the healthcare incident taxonomy extended from the original system (Itoh & Andersen, 2004) for the special purposes of assessing reporting culture and safety performance. Feasibility of the extended taxonomy for risk analysis was illustrated through application results to a number of incident reports collected from two Japanese hospitals. We believe that the extended taxonomy can be applied to analysis of incident reports on the basis of establishing positive reporting culture that will in turn contribute to building a safe healthcare organisation. For further extension of the taxonomy and its application to risk management, several additional dimensions and measures or indices are under development for the purpose of assessing "learning culture", another component of safety culture.

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 – who do not have human factors expertise – can easily and consistently apply the taxonomy system. In

addition to such a user manual, it may be required to create a short guideline 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 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

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

Beckmann, U., West, L.F., Groombridge, G.J., Baldwin, I., Hart, G.K., Clayton, D.G., Webb, R.K. & Runciman, W.B., 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(3): 314–319.

Bove, T. 2002. Development and Validation of a Human Error Management Taxonomy in Air Traffic Control. Risø National Laboratory, Risø-R-1378(EN). http://www.risoe.dk/rispubl/SYS/syspdf/ris-r-1378.pdf.

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, Stationery Office, London.

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(1): 5–28.

Edmondson, A.C. 2004. Learning from failure in health care: frequent opportunities, pervasive barriers. *Quality & Safety in Health Care*, 13(Suppl. 2): ii3–ii9.

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

Isaac, A., Shorrock, S.T., Kennedy, R., Kirwan, B. Andersen, H.B. & Bove, T. 2001. The human error in ATM (HERA) technique. Eurocontrol Headquaters.

Isaac, A., Shorrock, S.T. & Kirwan, B. 2002. Human error in European air traffic management: the HERA project. *Reliability Engineering and System Safety*, 75(2): 257–272.

Itoh, K. 2003. Safety culture and it's 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 H. Luczak & K.J. Zink (eds.). *Human Factors in Organizational Design and Management IV*. Santa Monica, CA: IEA Press, 35–45.

Itoh, K. & Andersen, H.B. 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*, Vol. 5, 2714–2719, Berlin, Germany, June 2004.

Itoh, K., Andersen, H.B. & Madsen, M.D. 2006. Safety culture in healthcare, In P. Carayon, (ed.), *Handbook of Human Factors and Ergonomics in Healthcare and Patient Safety*, Mahwah, NJ: Lawrence Erlbaum Associates, 199–216.

Itoh, K., Seki, M. & Andersen, H.B. 2003. Approaches to transportation safety: Methods and case studies applying to track maintenance train operations, In E. Hollnagel (ed.). *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum Associates, 603–632.

Landis, J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* (33), 159–174.

Reason, J. 1993. Managing the management risk: New approaches to organisational safety. In B. Wilpert and T. Qvale (eds.), *Reliability and safety in hazardous work systems*. Hove: Lawrence Erlbaum Associates.

Reason, J. 1997. *Managing the risk of organizational accidents*. Aldershot, UK: Ashgate.

Reason, J. & Hobbs, A. 2003. *Managing maintenance error: A practical guide*. Aldershot, UK: Ashgate.

Shorrock, S.T. & Kirwan, B. 2002. Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*. 33(4): 319–336.

Risk, Reliability and Societal Safety – Aven & Vinnem (eds)
© 2007 Taylor & Francis Group, London, ISBN 978-0-415-44786-7

A human error taxonomy for analysing healthcare reporting culture and its effects on safety performance

K. Itoh & N. Omata
Tokyo Institute of Technology, Tokyo, Japan

H.B. Andersen
Risø National Laboratory, Roskilde, Denmark

ABSTRACT: In the present study, we developed an extended human error taxonomy system for healthcare risk management, aiming to apply it to evaluating safety performance and reporting culture in specific organisations. The taxonomy system contains, besides categories to classify errors and performance shaping factors, additional dimensions that seek to capture the maturity of case descriptions contained in incident reports. These dimensions are reporting content, time-band of description, and description level. Applying several dimensions in the taxonomy, we proposed two safety performance measures which can be calculated from assignments to several dimensions in the taxonomy system: rate of near-miss reporting; and rate of near-miss detection by safety procedure. The present paper reports application results of the extended taxonomy to a total of 3,749 incident cases collected from two Japanese hospitals, which were at different stages of patient safety activities.

1 INTRODUCTION

A variety of approaches and methods for risk management have been applied to patient safety issues in healthcare just as in modern human-machine system areas 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), similar to human-machine operation domain (e.g., Itoh et al., 2003). Thus, it is not numbers of cases that are required, but stories from which we can learn; i.e., narratives contained in incident reports. For these requirements, a *human error taxonomy* can be critically useful to support the analysis of incident reporting to enhance systematic organisational learning.

On this background, Itoh & Andersen (2004) developed a human error taxonomy system for analysing healthcare incident reports, partly on the basis of a system used in air traffic management (Isaac et al., 2001; 2002; Shorrock & Kirwan, 2002). The primary purpose of this taxonomy system is its focus on identifying risk factors involved in medical accidents as well as analysing characteristics of the associated human and organisational failures. In addition to these analyses, it is also of importance to diagnose and track current levels of risk or safety for given work units and the overall organisation to which they belong. It may also be possible to calculate several types of reporting rates, e.g., ones applying to all reports and or to their subsets selected from specific classes of severity or specific case descriptions as measures connecting to safety performance level. However, there is a difficulty for a valid estimate of incidence of different types of events by these measures mainly due to the fact that incident reporting is selective (Battles & Lilford, 2003).

Thus, reporting rates of incidents or accidents may be essentially dependent on external factors. Hence, reporting rates for different types of incidents may not reflect internal processes or risk factors (Itoh et al., 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 frequently *latent factors* that contribute or even lead to the occurrence of human error made by frontline personnel (Reason, 1997).

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 (compare Helmreich & Merritt, 1998; Reason 1997). To establish a "safe" organisational culture, Reason (Reason, 1997; Reason & Hobbs, 2003) suggested that an organisation must improve the following three inter-connected component cultures: (1) a *reporting culture* – an organisational climate in which people are prepared to report their errors and near-misses; (2) a *just culture* – an atmosphere of trust in which people are encouraged, even rewarded, for providing essential safety-related information; and (3) a *learning culture* – a willingness and the competence needed to draw right conclusions from safety information system 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.

In the present paper, we extend the previous framework of human error taxonomy to an analysis system of incident reports so that one can evaluate the present status of safety performance and the level of reporting culture in a specific work unit or organisation. We believe that the above-mentioned difficulty or limitation to using incident reporting data for risk prediction can be overcome by applying the taxonomy to estimating safety culture status as well as characteristics of incidents which occurred in an organisation. We applied the extended taxonomy to analysis of incident reports collected from two Japanese hospitals. Detailed application results are mentioned below, in particular, analysis of reporting culture status and its relation or contribution to the level of safety performance in healthcare organisations. Reliability of the taxonomy is also discussed based on calculation of inter-rater assignments.

2 TAXONOMY FOR REPORTING CULTURE

2.1 *Diagnosing indices of reporting culture*

The original healthcare human error taxonomy (Itoh & Andersen, 2004) was developed by partly adapting a classification scheme developed for analysing aviation near-misses, the HERA (Human Error in ATM) taxonomy (Isaac et al., 2001; 2002; Shorrock & Kirwan, 2002). The healthcare taxonomy was extended to include an additional section concerning the maturity of case description with several minor revisions in the original sections. In this additional section, the following three dimensions were included: report content, time-band of description, and description level.

We suggest that an important indicator of the level of reporting culture in a particular unit or organisation can be based on incident reports submitted during a specific period, applying the three dimensions mentioned above.

By "report content" we refer to properties of the case as described in the incident report: the type of event involved (e.g., medication, patient fall, equipment, injection), background factors (situational context, e.g., busy ward, nightshift), causes of errors, contextual conditions behind errors (e.g., descriptions about human factors, organisational factors, and poor communication), and lessons learned.

In the dimension of time-band of description, we identify the time or tense of things which happened or are expected or look-ahead from statements contained the incident report, i.e., things occurred in the past (long before the event occurred), just before the event, present (when the event occurred), immediately after the event, and expected to occur in the future (situation to be expected).

The description level of a report concerns the depth or thoroughness of the report and is rated on a five-point scale, ranging from 1 to 5: no information contained (1), very briefly described (2), described at more length but superficially (3), described in detail (4), and described in depth (5). As one of the several indicators of the level of the reporting culture, that for a specific organisation or department/ward is then (in part) characterised in terms of percentage description of each "content" category over the entire set of incident reports submitted during a given period. The average descriptive depth will be calculated from the above-mentioned five point values of description level assigned to all written reports. This index will be also used as a rough measure of the reporting culture level of a specific work unit.

2.2 *Safety performance indices*

Concerning safety performance, we suggest that the reporting "rate" calculated by a specific subset of the submitted incident reports such as one selected on the basis of severity level or one captured by a particular cues could be possible indicators of safety level for an entire organisation or a given work unit. Of course, 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 incident rate. In general, the incident rate 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 the inverse of risk, that is, as an index of *safety* – that is, the more that staff is demonstrating willingness

to report, the greater is their sensitivity to errors, the greater the learning potential, and so, the greater is the safety in their department.

It has often been observed that the incident rate may be read as a "safety" index – more frequent reporting of incidents is a sign of safety (Edmondson, 1996; 2004). We make a distinction between near-miss cases in which a wrong act had no consequences because it did not reached a patient (categorised Level 0 in the outcome severity classification) and incidents at the other severity level. Error reporting for the former type of "near-miss" events can therefore be considered to reflect the staff's safety awareness or sensitivity to errors or incidents. Based on this expectation, we introduce the percentage of near-miss cases over all 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 incidents in terms of the type of "error capture cues" that were used to detect or capture the error, e.g., a wrong act detected by a safety rule or procedure which was built into the work system can be expected to avoid similar incidents regularly. In contrast, errors may sometimes remain undetected, or the error may be caught or detected by chance. Thus, the rate of error detection cases by the built-in safety procedure over all near-miss cases reported – which is referred to as the rate of near-miss detection by safety procedure – would be another promising safety performance indicator, we propose.

As another index pertaining to safety performance, the rate of Level 3 or higher cases can be calculated by applying the dimension of "severity of outcome", in which accidents and incidents are classified into six levels, ranging from 0 to 5 – Level 0 means a near-miss case, and Level 5 indicates death. Usually, cases at Levels 0 through 2 are regarded as incidents in Japanese hospitals, whereas a Level 3 case is an event that actually or potentially involves or leads to a minor patient injury – which may require an extended hospitalisation of a few more weeks at most. According to a number of interviews with risk managers and leaders of nursing department, staff has nearly always – and nurses definitely always – reported severe cases of Level 3 or higher to their hospital's incident reporting system. Therefore, an index such as the one discussed may be a plausible indication of "risk": a greater number of errors take place in a workplace with a higher rate of this index since all incident cases are reported.

The present study applied the first two safety performance indices, i.e., the rate of near-miss reporting and rate of near-miss detection by safety procedure, to two hospital samples of incident reports. The other index was not adopted since at most a single case with Level 3 or higher had occurred in each of the departments or wards in involved.

Table 1. Collected reports and hospital attributes surveyed.

			No. of beds & staff		
Collected	Reports	Survey period	Beds	Dr.	Ns.
Hospital					
A	3,162	4/2002~9/2004	347	60	310
B	587	4/2000~9/2002	475	70	314
Total	3,749				

3 APPLIED CASES OF THE TAXONOMY

We applied the extended taxonomy to a total of 3,749 incident reports collected during a two-and-a-half year period from two hospitals located in different rural regions of Japan. The number of collected reports, survey periods and other hospital data are shown in Table 1. According to our interviews with the risk managers of both hospitals, organisation-wide initiatives for patient safety were initiated several years before the survey period, and in particular, special attention has 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. Difference in risk management and reporting culture between the hospitals, for instance, might well lie behind their annual reporting rates of incidents per nurse: 3.05 (Hospital A) vs. 0.65 (B).

In the present paper, we focus on applying results of the taxonomy to the incident reports collected from Hospital A. Results of the hospital-based analysis are also mentioned to make organisational comparisons of safety outcome, using the incident cases from both hospitals.

4 APPLICATION RESULTS

4.1 *Periodic changes in reporting culture*

In a healthcare organisation having a positive reporting culture we may expect that a staff member will submit an incident report unhesitatingly if he or she has become involved in any reportable event. To examine this expectation, we illustrate in Figure 1 when incident cases were reported to the local reporting system in Hospital A in terms of duration from the event detection to the report submission based on three half-a-year periods between April 2003 and September 2004. A significant difference was observed by the Chi square test in the timing of event reporting between the three half-year periods ($\chi 2 = 28.75, p < 0.05$). Compared with the earlier two periods, the percentage of reports made immediately after the event became greater and that of reporting after an hour was smaller in the last period of the survey. Approximately half

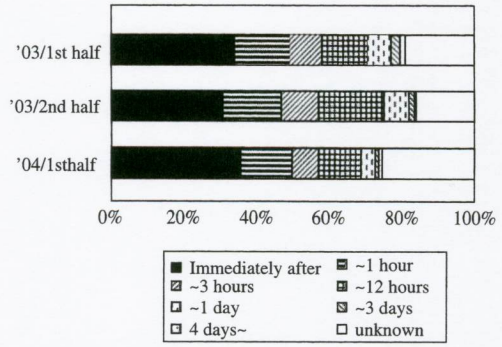


Figure 1. Periodic changes in percentage reporting of incident cases based on the timing from event detection (Hospital A).

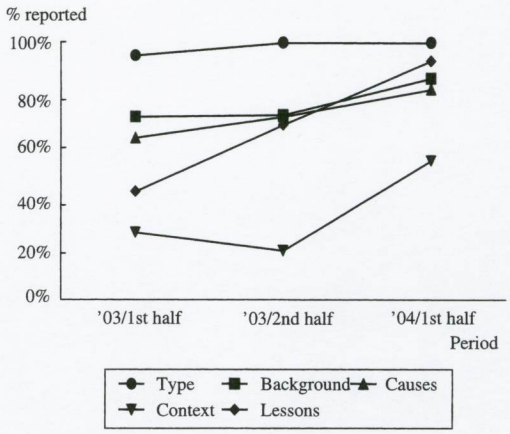


Figure 2. Periodic changes in percentage of descriptions included each category of reporting content (Hospital A).

of the incidents were reported within three hours after detection of the event although no information on submission time was found in approximately a quarter of reports.

To address another aspect of the reporting culture, namely, whether incidents were "properly" reported, we analysed reports for every half-year period by applying the "report content" dimension of the taxonomy. Figure 2 depicts periodic changes in percentages of description of each category of reporting content, applying to incident cases which occurred in Hospital A. The percentage of description for each reporting category was significantly increased every half-a-year period ($F = 4.42$, $p < 0.05$; applying the two-way ANOVA of the periods and the categories of reporting content). As illustrated in this figure, there was also a significant difference in the percentage of description between the reporting content categories ($F = 45.53$,

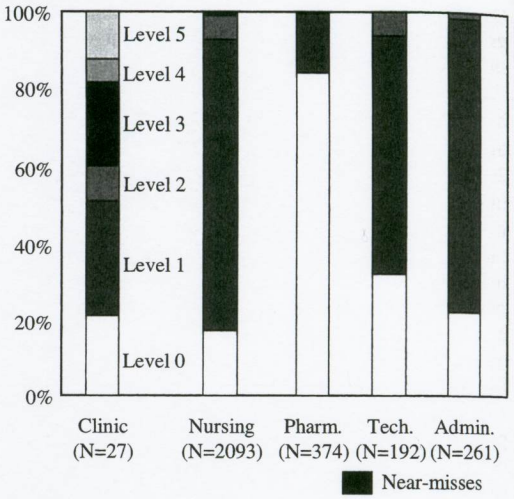


Figure 3. Department-based reporting percentage for each level of severity (Hospital A).

$p < 0.001$). In the last half-year period of the survey (April-September 2004), healthcare staff in this hospital were diligent in describing the type of event involved (100%), background factors (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 of errors.

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

4.2 *Department-based severity*

Regarding the severity of events collected from Hospital A, Figure 3 indicates percentages of each severity level of incidents for five departments, i.e., clinic, nursing, pharmacy, technical and administration departments. We did not apply a statistical test to these department-based data due to very small samples of reports from clinic and administration departments. Incident cases with severity index >2 were reported only from the clinic department where its proportion was 22% – a small number of severity level 2+ cases, i.e., 0.7% of proportion, were also reported from the nursing department. Figure 3 demonstrates the other extreme of severity distribution that more than 80% of cases were near-misses, i.e., Level 0 incidents, reported from the pharmacy department. 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 consequence reached the patent.

Comparing the rate of near-miss reporting between the departments, the value of this index for the nursing department is the lowest, it is slightly higher for the clinic department, and is 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. For instance, nurses usually work longer hours with patients than any other healthcare group, while doctors may perform more hazardous tasks with patients than other professionals. In addition, as also pointed out in Section 1, incident reporting is selective. In particular, reporting rate of incidents are lower from doctors than other healthcare 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 with no adverse effects to patients. Therefore, it may be safer to interpret the differences in reporting with a view to differences in tasks and work routines. For instance, the nature of the 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 between departments or organisations having identical or similar external factors, and should be tracked within a department or an organisation across several periodic intervals.

Periodic changes in the reporting indices within the same organisation or work unit were analysed by applying the safety performance measures mentioned in Section 2.2 to the hospital-wide set of reports including its subsets selected from each department. Results of periodic analysis will be mentioned in the next subsection.

4.3 *Periodic changes in safety level*

Selecting only near-miss cases from an entire set of incident reports in Hospital A, proportions of "error detection cues" are depicted in Figure 4 based on half-year periods in the same manner as Figure 1. Frequencies of categories that represent the "error detection 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 over the periods. In contrast, the percentage of error detection by the staff member involved in the near-miss case became smaller in the first half of 2004, compared with that in the earlier periods. Percentages by other detection cues or persons, e.g., team members, staff in other departments, and patients or

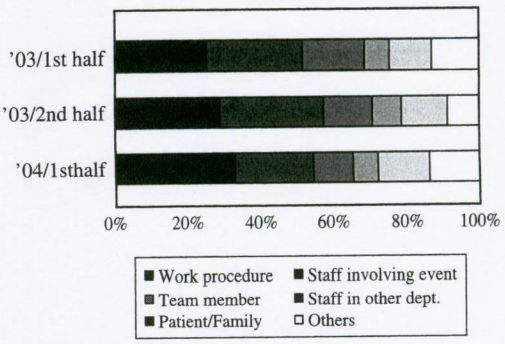


Figure 4. Periodic changes in percentages of error capture cues for near-miss cases every half a year (Hospital A).

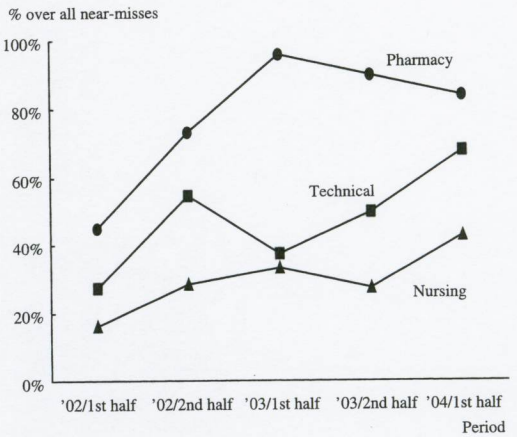
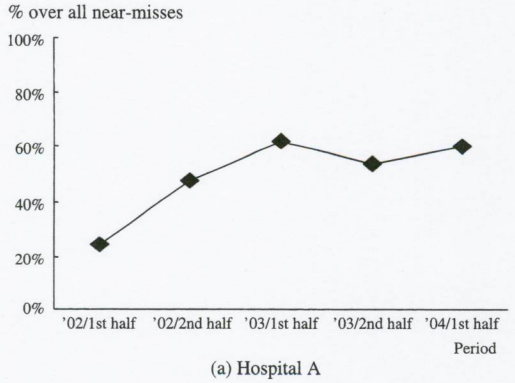


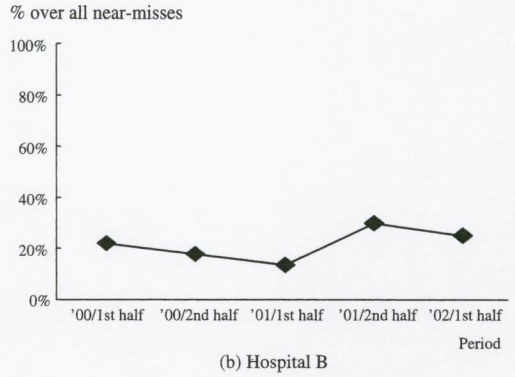
Figure 5. Department-based periodic changes in rate of near-miss detection by safety procedure (Hospital A).

their families, did not seem to have changed throughout the entire one-and-a-half-year period. Thus, the rate of near-miss detection by safety procedures had been gradually improved over the period in this hospital. Recalling this issue, we discuss the hospital-wide safety level in the next subsection, comparing this with the level of the other hospital surveyed.

Changes over time in the rate of near-miss detection by safety procedures during two and a half years for three departments in Hospital A are depicted in Figure 5. The departments included are the nursing, the pharmacy and the technical department, whereas the clinical department is excluded due to its very small number of near-miss reports A two-way ANOVA analysis of this index, the rate of near-miss detection by safety procedure, showed that the pharmacy was significantly higher than the other two departments. The nursing department achieved the lowest percentage throughout the two-and-a-half year period ($F = 27.10, p < 0.001$). As mentioned previously, this



(a) Hospital A



(b) Hospital B

Figure 6. Comparison of periodic changes in rate of near-miss detection by safety procedure between two hospitals.

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

4.4 *Organisational comparisons of safety levels*

To compare safety performance measures at the overall organisational level, we applied the extended taxonomy to a set of incident reports collected from the other hospital, B. Figure 6 illustrates the periodic changes in the rate of near-miss detection by safety procedure during a different two-and-a-half year period for both Hospitals A (between April 2002 and September 2004) and B (two 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 the other hospital.

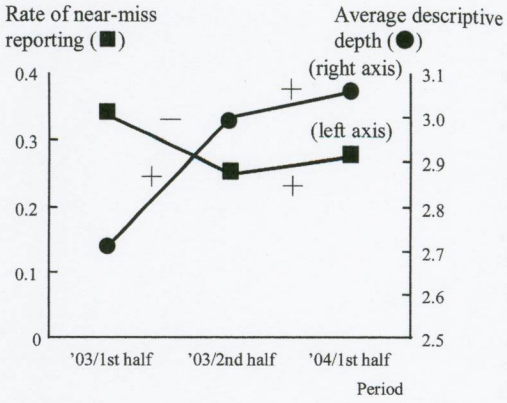


Figure 7. Periodic changes in and correlation between two hospital-wide measures: rate of near-miss reporting and average descriptive depth.

In addition, this safety performance rate was gradually increased in Hospital A (an 8% average increase every half year during the 2.5 year survey period). We analysed the correlation between the rate and the period to statistically examine this trend, finding a correlation coefficient of $r = 0.86$ ($p = 0.06$). In contrast, the rates for Hospital B during the two-and-a-half year period were nearly constant with a slight increase in the last year (2% increase; $r = 0.47, p = 0.43$). The trends in the periodic changes in this measure of the two organisations seem to be, at least in part, due to the hospital-wide activities for patient safety in organisation. As mentioned in Section 3 in connection with the review of some results of interviews with the risk managers, continuous patient safety activities may have contributed to a higher and increased rate of near-miss detection by safety procedures in Hospital A compared to Hospital B where no special safety-related activities took place and no increase in this rate appeared during the survey period.

4.5 *Correlations between reporting culture and safety performance*

In this subsection, we discuss relationships between reporting culture and safety performance based on the applications of the taxonomy system. Figure 7 illustrates changes over the periods observed in the rate of near-miss reporting and the average descriptive depth (cf. Section 2.1) for the three half-year periods in Hospital A. While one might expect data to show a relationship between reporting culture measures and safety level measures, simple and straightforward statistical analyses did not yield a significant correlation between these two measures. As can be seen 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

Table 2. Ward-based analysis of correlation between safety performance and reporting culture.

Wards/ Sections	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

Sign (+/−): Periodic transition from second half of 2003 to the first half of 2004.

to the second period. This discrepancy between the two measures may, it can be speculated, be due healthcare staff possible tendency to report in greater detail when an incident has a more severe outcome. Also, as mentioned previously, the safety level may be affected not only by internal factors such as reporting culture but also by various external factors.

Table 2 shows the results of a ward-based analysis to examine concordance of the above-mentioned two measures in terms of increasing or decreasing trend between the two successive periods, using incident reports selected only from nurses. 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 (including operating room and outpatient staff) improved the descriptive depth, and the rate of near-miss reporting increased in eight wards during the successive two half-a-year periods. The trend of improvement or degradation of the two indices matched for nine out of ten wards. This may suggest that positive reporting culture contributes to higher level of safety performance in healthcare.

4.6 *Inter-rater reliability*

Reliability of the taxonomy was evaluated on basis of raw agreement and chance corrected measure of agreement, applying the same procedure used in the former study (Itoh & Andersen, 2004) and in Bove (2002). Two judges, a human factors expert (not a medical domain expert) and an undergraduate student in human factors and industrial engineering, performed category assignments for 138 incident cases which were randomly selected from the Hospital A sample by applying the taxonomy. Raw agreements and chance corrected agreements were calculated using category assignments rated by the two judges.

Table 3. Reliability evaluation of ratings with the taxonomy.

Dimensions	Raw agreement	Chance corrected
Severity	0.91	0.87
Event capture cue	0.51	0.47
Time of detection	0.41	0.27
Reporting content*	0.58	0.35
Time-band of description*	0.36	0.36
Description level	0.59	0.42

* Calculation of reliability indices were applied modified raw agreement and chance corrected measures of agreement (Itoh & Andersen, 2004).

The results of calculating the inter-rater reliability (Cohen's *kappa*) are shown in Table 3 for the additional dimensions of the taxonomy extended in this study. As shown in this table, there were moderate and high chance corrected agreements between the two judges for three, near-moderate for two others, and slight agreement for one. Landis & Koch (1977) suggest that levels between 0.41 and 0.60 indicate 'moderate' agreement, 0.61–0.80 'substantial', and above 0.80 'almost perfect' agreement, whereas 0.21–0.40 only fair agreement. Thus, the values of the chance corrected measures indicate that the extended taxonomy did not obtain an acceptable level of agreement for several additional dimensions: time of detection, reporting content, and time-band of description.

5 CONCLUSION

The present paper reported the additional dimensions of the healthcare incident taxonomy extended from the original system (Itoh & Andersen, 2004) for the special purposes of assessing reporting culture and safety performance. Feasibility of the extended taxonomy for risk analysis was illustrated through application results to a number of incident reports collected from two Japanese hospitals. We believe that the extended taxonomy can be applied to analysis of incident reports on the basis of establishing positive reporting culture that will in turn contribute to building a safe healthcare organisation. For further extension of the taxonomy and its application to risk management, several additional dimensions and measures or indices are under development for the purpose of assessing "learning culture", another component of safety culture.

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 – who do not have human factors expertise – can easily and consistently apply the taxonomy system. In

addition to such a user manual, it may be required to create a short guideline 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 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

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

Beckmann, U., West, L.F., Groombridge, G.J., Baldwin, I., Hart, G.K., Clayton, D.G., Webb, R.K. & Runciman, W.B., 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(3): 314–319.

Bove, T. 2002. Development and Validation of a Human Error Management Taxonomy in Air Traffic Control. Risø National Laboratory, Risø-R-1378(EN). http://www.risoe.dk/rispubl/SYS/syspdf/ris-r-1378.pdf.

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, Stationery Office, London.

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(1): 5–28.

Edmondson, A.C. 2004. Learning from failure in health care: frequent opportunities, pervasive barriers. *Quality & Safety in Health Care*, 13(Suppl. 2): ii3–ii9.

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

Isaac, A., Shorrock, S.T., Kennedy, R., Kirwan, B. Andersen, H.B. & Bove, T. 2001. The human error in ATM (HERA) technique. Eurocontrol Headquaters.

Isaac, A., Shorrock, S.T. & Kirwan, B. 2002. Human error in European air traffic management: the HERA project. *Reliability Engineering and System Safety*, 75(2): 257–272.

Itoh, K. 2003. Safety culture and it's 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 H. Luczak & K.J. Zink (eds.). *Human Factors in Organizational Design and Management IV*. Santa Monica, CA: IEA Press, 35–45.

Itoh, K. & Andersen, H.B. 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*, Vol. 5, 2714–2719, Berlin, Germany, June 2004.

Itoh, K., Andersen, H.B. & Madsen, M.D. 2006. Safety culture in healthcare, In P. Carayon, (ed.), *Handbook of Human Factors and Ergonomics in Healthcare and Patient Safety*, Mahwah, NJ: Lawrence Erlbaum Associates, 199–216.

Itoh, K., Seki, M. & Andersen, H.B. 2003. Approaches to transportation safety: Methods and case studies applying to track maintenance train operations, In E. Hollnagel (ed.). *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum Associates, 603–632.

Landis, J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* (33), 159–174.

Reason, J. 1993. Managing the management risk: New approaches to organisational safety. In B. Wilpert and T. Qvale (eds.), *Reliability and safety in hazardous work systems*. Hove: Lawrence Erlbaum Associates.

Reason, J. 1997. *Managing the risk of organizational accidents*. Aldershot, UK: Ashgate.

Reason, J. & Hobbs, A. 2003. *Managing maintenance error: A practical guide*. Aldershot, UK: Ashgate.

Shorrock, S.T. & Kirwan, B. 2002. Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*. 33(4): 319–336.