Analysing Medical Incident Reports by Use of a Human Error Taxonomy

Kenji Itoh[1] and Henning Boje Andersen[2]
[1] Tokyo Institute of Technology, Tokyo, Japan
[2] Risø National Laboratory, Roskilde, Denmark

Abstract

The present paper describes analysis results of incident reports collected from three Japanese hospitals. We developed a medical incident taxonomy by adapting an existing scheme in air traffic management. Its application allowed us to elicit characteristics of adverse events and errors that have occurred in health care. The taxonomy was evaluated in terms of inter-rater reliability by use of raw agreement and a tentative measure of chance corrected agreement between two judges.

1 Introduction

A variety of research methods for risk management have been developed and applied to modern human-machine system areas such as aviation, maritime operations and nuclear power production. It is to be expected that medical activities share many characteristics with high-tech human-machine system operations. Therefore, it would seem useful to adapt some of the research methods that have been developed for application in such human-machine system domains to investigating potential factors involved in patient safety. Similar to their use in these high-tech industries, incident reporting systems are also utilised in all Japanese hospitals, and they are regarded as a primary methodology for risk management.

The present study develops an incident and human error taxonomy system for analysing medical incident reports. The taxonomy system is partly adapted from a system used in air traffic management [1]. We applied the taxonomy to analysis of incident reports collected from three Japanese hospitals. The primary purpose of its application is to identify critical risk factors of medical accidents as well as characteristics of incidents and human errors that have occurred in health care. Reliability of ratings with the taxonomy was evaluated based on raw agreement and a tentative chance corrected measure of agreement between two judges.

2 Taxonomy of Medical Incidents

The medical incident taxonomy was developed in part on the basis of taxonomies targeted at human error in air traffic control [1]. The taxonomy is divided into five overall dimensions, each of which contains one or several (sub-)dimensions comprising individual categories (shown in Table 1). The *task* dimension specifies medical and nursing activities involved in the incident such as injection, medication, diagnosis, treatment and tube troubles. The *action* dimension is a description of healthcare staff action at the time when an error was made.

Table 1 Outline of dimensions included in the taxonomy

Dimensions	Example categories
<Task performance>	
Task	Injection, medication, treatment, inspection, tube troubles, patient falls, etc.
Action	Monitoring, communication, manipulation, decision making, etc.
Info./Equip.	Patient, team member, medicine, medical document, device & materials, etc.
<Error occurrence>	
Error type	Omission, timing error, qualitative error, selection error, sequential error, etc.
Violation	Unintentional rule contravention, routine, exceptional, and situational violation
<Contextual conditions>	Communication factors, staff [or patient] human factors, task factors, equipment factors, organisational factors, environmental factors
<Outcome and recovery>	
Severity	From level 0 (near-miss) to 5 (death)
Error detector	Patient him/herself, staff him/herself, team member, another patient, family, etc.
Error recovery	Inform. to leader, explanation to pt., apologise to pt., additional inspection, etc.
<Learning from the case>	Reinforcement of rule, command, communication & check, fool-proof, checklist, manual, change of procedure, work condition, training, etc.

The *information/equipment* dimension specifies the type of information, if any, that healthcare staff misperceived, forgot or misjudged or what human-machine element, if any, was used in the error. The *error occurrence* aspect involves two dimensions referring to the *type of error*, i.e., how it occurred in terms of errors of omission and commission and the *type of violation*. The dimension of *contextual conditions* is divided into several factors, as indicated in Table 1. Each of these factors includes several elements that potentially affect safety and quality of activities. For example, typical elements of organisational factors are manuals and checklists, medical documents and records, rules and procedures, training, commands and indication and management style. The *outcome and recovery* aspect comprises four dimensions: *severity* of the outcome, *error detection*, and *error recovery*. Finally, *lessons learned* from the individual incident case include the types of recommendations for reinforcing or revising work aspects to avoid repeated occurrence of the event.

3 Application of the Taxonomy

In each incident report, we identified 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. First, for each report it was checked whether the cause-event relations described was sufficiently depicted or contained any discrepancies. If no human error or causal factor could be identified from the report, the case report was returned to a risk manger to acquire more details.

2. One or multiple items were selected for each dimension from a category list based on the open-ended description of the event. In particular, we sought to identify what errors took place and then judge what types of errors were involved and what actions had failed.
3. Additional category items were elicited with reference to a guideline regarding a matching of the categories of the taxonomy with error items and incident types that are specific to each hospital's reporting form. Any category that was judged to have a highly likely, direct – but non-stated and therefore implicit – relationship with a stated fact would be assigned. For example, when a nurse, having only a few months experience, is involved an incident in which she/he makes a wrong decision, it can be inferred that she/he may not have enough experience or knowledge; therefore, we select as an additional category "staff human factors" in the contextual conditions dimension.

We collected a corpus of 1,327 incident reports from three hospitals located in different regions of Japan. All the incident reports analysed were submitted by nurses, there being very few cases reported by doctors or other staff groups.

4 Application Results

4.1 Characteristics of Medical Incidents

In this section, we discuss applications of the taxonomy system to incident reports. First, there is a common trend in the types of tasks involved in adverse events, although frequencies of categories comprising the task dimension are significantly different between these hospitals (χ^2=110.14, p<0.001). Top four "tasks" showed the same order in each hospital: injection-related incidents (21% of reported cases on average), medication incidents (19%), patient falls (16%) and removal of a tube by patient him/herself (11%).

The reporting rates of each category in the action dimension are shown in Figure 1. There is also a significant difference in categories of this dimension between the hospitals (χ^2=170.93, p<0.001). However, in each hospital approximately a half of the errors were made while the healthcare staff – only nurses reported in this study – was (or should be) monitoring patients or while the nurse was manipulating medical equipment or materials such as a tube, syringe and pump. Errors were also frequently made when nurses were (or should be) monitoring medical equipment, materials or medicine (6~13% for the three hospitals) and when they are reading or checking a nursing record or another medical document (3~13%). The percentages of these actions failed vary among the hospitals. A possible reason for the variation among the hospitals in the rates of the individual types of failed actions is that their staff might have somewhat different work situations. Thus, staff tasks depend on, among other things, the medical equipment that is used and, therefore, calls for different types of procedures, combinations of devices and, in general, different work routines. Similarly, the reporting rate of failed actions related to document handling might also be affected by the use (or non-use) of an electronic case record system which, when available, relieves nurses of the need for writing or transcribing any paper documents.

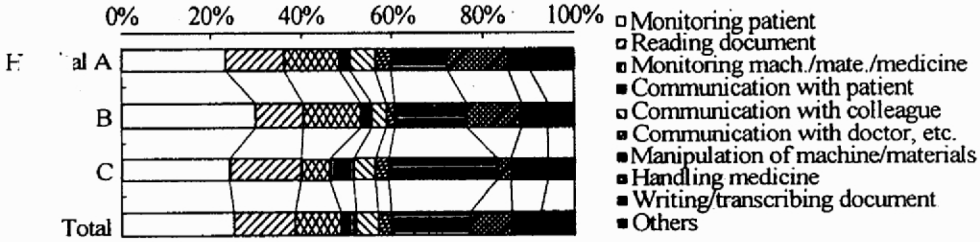

Figure 1 Actions failed in three Japanese hospitals

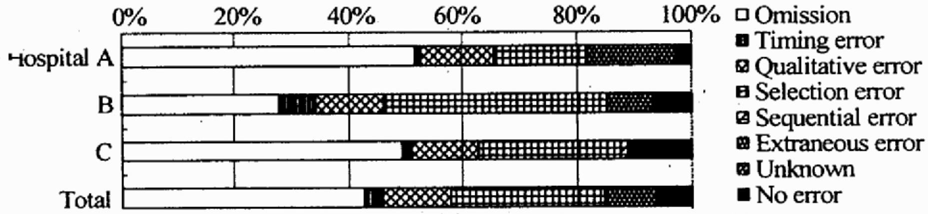

Figure 2 Human error types occurred in medical incidents

Rates of error types aggregated for all the tasks/actions are depicted in Figure 2. The most frequent error type found in our healthcare sample was *errors of omission* (41%) – similar to what is found in the case of maintenance activities in nuclear power plants [3], where one even finds almost the same rate (43%). Nevertheless, the rate of each error type significantly depends on the tasks (e.g., for Hospital A, $\chi^2=152.86$, $p<0.001$) and the actions (e.g., for Hospital A, $\chi^2=315.78$, $p<0.001$) as well as on the hospitals ($\chi^2=135.74$, $p<0.001$). Following omission errors, *selection errors* were the next most dominant type among errors of commission. The rate of this error type varies greatly between the hospitals: from 16% to 39%.

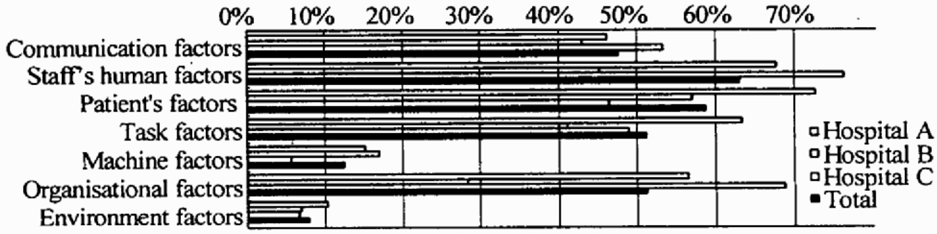

Figure 3 Contextual conditions behind adverse events

Within the dimension of *contextual conditions*, a great number of incident cases turned out to be characterised by multiple background factors. Involvement of each factor in adverse events is significantly dependent on categories of the task dimension ($\chi^2=265.79$, $p<0.001$). Overall percentages of each factor aggregated for all the cases are shown in Figure 3. Just like in the other dimensions, the rate of each category and its elements vary greatly *between the hospitals. More than 60% of the cases involved staff* human factors. The staff human factors category most often applicable was lack of knowledge (34% on average), followed by preoccupation (29%) and psychological factors such as time pressure and confusion (9%). The category of *patient human factors* was also involved in the majority of the reported incidents, viz., in 57%, and within this category the most often cited task involved "patient falls" (94%) and "tube troubles" (91%). In approximately 50% of the cases, both organisational and task factors were involved. The most frequently reported organisational factors were training issues (16%), followed by inappropriate rules and procedures (13%), issues on medical documents (10%) and inaccurate indication and instruction (7%).

4.2 Inter-rater Reliability

To evaluate the reliability of ratings with the taxonomy, we calculated raw agreements and a tentative chance corrected measure of agreement 34 incident cases randomly selected from one of the hospitals rated by two judges: a human factors expert (not a medical domain expert) and a graduate student in human factors. In case that one and only one category is selected in a dimension for each incident report, 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 cases in our application. Taking into account that part of the agreement between judges will be obtained by chance alone, we introduce a tentative measure of agreement applying a similar idea to the Kappa statistics [2]. Like the Kappa statistics, this measure ranges from -1.0 to +1.0, where a value of zero indicates that the level of agreement is identical to that expected by chance.

In the taxonomy developed in this study, ratings may be non-exclusive: for most of the dimensions, not only one but multiple categories can be assigned to a case. For example, more than one error might be made in a single incident case and we must assign error types to all the errors. To accommodate this, we apply the standard procedure of the Kappa statistics [2] whenever we measure agreement within a dimension that requires exclusive (single category) rating. But 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 of categories, where $a_k(i,j)=1$ [=0] means that judge k assigns [does not assign] j to i. A matching rate between two judges (k=1, 2) for a given dimension is defined in Equation (1), where $M=1$ when $a_1(i,j)= a_2(i,j)$ and =0 otherwise.

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

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 raw agreement (for multiple rating) may then be calculated for any dimension as the mean matching rate over all the cases classified by the taxonomy. Therefore, for multiple rating, chance agreement is defined as Equation (2)

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

Table 4 shows the results of calculating the inter-rater reliability for each dimension in the taxonomy, applying the above-mentioned modified measures. As indicated in this table, there are relatively large raw agreements between the two judges in most dimensions. However, the values of the chance corrected measure indicate that the taxonomy did not obtain an acceptable level of agreement for several dimensions: error types, action failed, and learning from the case.

Table 2 Reliability evaluation of ratings with the taxonomy

	Task	Action*	Info./equ.*	Err. type*	Severity	Detector	Learning*
Raw agreement	0.94	0.90	0.89	0.77	0.91	0.91	0.93
Chance correct.	0.87	0.24	0.44	0.01	0.63	0.88	0.04

	Contextual conditions (--- factors)*						
	Communi-cation*	Staff human*	Patient human*	Task*	Equip-ment*	Organisa-tional*	Environ-mental*
Raw agreement	0.92	0.92	0.95	0.94	1.00	0.89	0.98
Chance correct.	0.53	0.51	0.68	0.57	0.97	0.40	0.92

*: Results of calculating modified raw agreement and chance corrected measures of agreement

5 Conclusion

The present paper reported on the development and application of a medical incident taxonomy system designed to be used by healthcare risk managers in analysing incidents that have occurred in their organisation. We also mentioned some results of its first application to incident reports from three Japanese hospitals. Application of the taxonomy to more than 1,000 incident reports allowed us to partially uncover characteristics of errors and causal factors which have been made in Japanese hospitals. We believe that such results produced by the taxonomy application can be utilised for safety management activities to protect patients from repeated occurrence of accidents that will be caused by the same latent factors.

To overcome low values of chance corrected measure for some dimensions, we need to prepare an appropriate guideline or procedure for category assignment so that any healthcare risk manger can easily and consistently apply the taxonomy. Such improvement should also include creating a guideline for reporters when describing a case or event in an incident report as well as adapting a reporting form so that taxonomy application and critical incident analysis can be easily performed. Finally, it is also required to devise a more appropriate measure of chance corrected agreement for a taxonomy in which multiple categories may be assigned to a case.

References

1. Shorrock, S.T. and Kirwan, B. Development and application of a human error identification tool for air traffic control, *Applied Ergonomics*, 2002; 33(4): 319-336.
2. Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurements*. 1960; 20: 37-46.
3. Rasmussen, J. What can be learned from human error reports? In K. Duncan, M. Gruneberg and D. Wallis (Eds.), Changes in working life. Wiley, London, 1980.