

Current Incident Reporting and Learning Systems Yield Unreliable Information on Patient Safety Incident Reporting Reliability

Mari Plukka*, Auvo Rauhala, Tuija Ikonen and Lisbeth Fagerstrom

Department of Clinical Medicine, Abo Academi University, Turku, Finland

Abstract

Background: Incident reporting systems are being implemented throughout the world to record safety incidents in healthcare. The quality of the recording and analysis of reporting systems is important for the development of safety promotion measures.

Methods: To assess the reliability of incident reporting ratings collected in a hospital setting, a three level interrater comparison was undertaken. The routine ratings of the frontline event handlers responsible for evaluating safety incident reports (n=495) were compared with the parallel ratings of two trained patient safety coordinators. The two patient safety coordinators then each separately reviewed about half of the 495 reports, followed by reclassification of a random sample (a random data subset of 60 reports) previously reclassified by the other coordinator during the first reclassification. The following seven patient safety variables were included: Nature of the incident, type of incident, patient impact, treating unit impact, circumstances and contributory factors, immediate actions taken, and risk category. Interrater agreement was tested with kappa, weighted kappa or iota.

Results: For the seven variables examined, event handlers had an average of 1.36 missing answers, patient safety coordinators 0.32. For the first interrater comparison, the average change between the three ordinal scale variables for all variables together was towards more serious in 29% (95% CI: 27%, 32%) and towards less serious in 2% (95% CI: 0%, 5%) of incidents. The net change for the first interrater comparison was 27% (95% CI: 25%, 30%) towards a more serious incident. The average selection of several categories, when allowed, increased from 7% (95%CI: 6%, 8%) to 33% (95% CI: 31%, 35%). For all three paired interrater comparisons, the average interrater agreements were in the range of 0.44 to 0.53 and considered moderate. While patient safety coordinators should in theory represent a 'golden standard', the coordinator interrater agreement seen in this study was moderate.

Conclusion: Consensus at national level on how to classify high risk incidents is needed to develop incident reporting reliability. Also, continuous training in common practices; terminology and rating systems should be given more attention. Having a patient safety coordinator reclassify incident reports can improve reporting accuracy and thereby corrective actions and learning.

Keywords: Patient safety • Incident reporting • Adverse event • Reliability • Validity • Classification

Introduction

The reporting of patient safety incidents is critical for understanding defects in care processes and improving safety [1]. Also referred to as 'occurrence reporting' or 'event reporting', voluntary patient safety incident reporting is the process of identifying and reporting events that could have or have led to a negative outcome, typically by those professionals directly involved in the

incident or events leading up to the adverse event [2,3]. There are three types of safety incidents: near miss, no harm incident and harmful incident [4].

Assessing patient safety is challenging because comprehensive standardised metrics have not yet been developed [5,6]. To realise reliable and consistent reporting, both common terminology and incident categorisation are needed [7]. To facilitate the consistent categorisation of safety incidents and thereby improve patient safety,

*Address for Correspondence: Mari Plukka, Department of Clinical Medicine, Abo Academi University, Turku, Finland, Tel: 408223573; E-mail: mari.plukka@ovph.fi

Copyright: © 2023 Plukka M, et al. This is an open-access article distributed under the terms of the creative commons attribution license which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Received: 01 March, 2023, Manuscript No. JCCR-22-76483; **Editor assigned:** 03 March, 2023, PreQC No. JCCR-22-76483 (PQ); **Reviewed:** 14 March, 2023, QC No. JCCR-22-76483; **Revised:** 20 March, 2023, Manuscript No. JCCR-22-76483 (R); **Published:** 27 March, 2023, DOI: 10.37421/2165-7920.2023.13.1557

the World Health Organization (WHO) has developed a conceptual framework for the international classification for patient safety, in which a standardised set of concepts and terms are presented [8].

At present, most perceive that the primary purpose of safety incident reporting systems is to expedite organisational learning and improvement, and the usefulness of a reporting system has been linked to the feedback that personnel and frontline event handlers provide. Some suggest that safety incident reports should not be used to monitor the rate of harm over time in hospital settings nor to appraise or compare hospital safety [9]. Some criticise the use of incident reporting systems or other methods for identifying adverse events, such as retrospective chart review, finding that they do not provide a true and reliable picture of the level of patient safety within an organisation [10,11].

Despite reported problems, limitations and criticism, safety incident reporting and learning systems are being implemented throughout the world [12]. Yet the regulatory and technical aspects of the reporting systems in use as well as the resulting reporting data vary. For example, a wide and diverse range of both mandatory and voluntary reporting systems are seen in the various European Union member countries [13].

In Finland, all health care organisations must maintain a safety incident reporting system. Fifteen years ago a systematic patient safety improvement project was started at Vaasa Central Hospital, a secondary care teaching hospital in Western Finland. In 2006 HaiPro®, a hospital electronic incident reporting system was introduced at the hospital. HaiPro® is a voluntary and anonymous system developed in line with international guidelines and used by over 80 percent of primary care providers and in all public hospitals in Finland [14]. In 2017, a systematic and comprehensive patient safety research and improvement programme was also implemented. As part of this programme, all available data on patient safety for 2017 was collected and the quality of patient safety incident reporting was investigated. Since 2020, the hospital has received funding from the ministry of social affairs and health for the establishment of a national coordination centre for patient safety improvement.

The main purpose of this study was to assess the reliability of the incident reporting ratings collected at Vaasa Central Hospital over three months in 2017 by comparing the routine ratings of the frontline event handlers responsible for evaluating safety incident reports with the parallel ratings of two trained patient safety coordinators. The parallel ratings were also compared, with the aim to illuminate the reliability of incident reporting, *i.e.* we sought to determine whether the parallel ratings should be considered concordant.

The research questions were:

- To what extent did the routine ratings align with the parallel ratings.
- How uniform were the parallel ratings, *i.e.* what is the interrater agreement between the two patient safety coordinators.

Materials and Methods

Study design and setting

The study was an observational descriptive cross-sectional study of the incident reporting data collected at Vaasa central hospital over three months in 2017 with the HaiPro® system, as part of a patient safety research programme. The chief medical officer of the hospital provided authorisation for the study. Identifiable patient data were not included; therefore approval from the hospital's ethics committee was not required.

Data collection

Personnel at the setting included in the study can report a safety incident or an adverse event with the HaiPro® system. All personnel are given a brief introduction to the system as part of their general orientation and provided with instructions on how to report an incident or observations they consider to be a 'near miss'.

Using a semi-structured electronic form, the person reporting an incident (the reporter) provides the following information: Reporter's professional group, date and time of incident, location of incident, circumstances and contributory factors, and description of incident as free text. All information is collected as structured data, except for the description of the incident.

Once a report is completed and submitted, an email notification is sent to an event handler and the senior staff nurse, deputy head and senior consultant for the unit where the incident took place. Drug related incident report notifications are sent to the medication safety officer overseeing the relevant unit. An event handler classifies the report, which includes categorisation of the nature of the incident, type of incident, patient impact, treating unit impact, circumstances and contributory factors, immediate actions taken, and risk category. Event handlers receive training in the handling of reports, and instructions are available on the hospital's Intranet.

The sample was based on practical possibilities. The overall data material was comprised of 495 reports recorded by personnel over three months (October-December) in 2017. For internal validation, about ten of the 495 reports were first reviewed and classified by members of a patient safety group to ascertain a common understanding of the rating policy. The patient safety group members included the director, director of operations and director of quality as well as a senior consultant and patient safety coordinators at the hospital included in the study. Broad agreement amongst the patient safety group members on the handling of the incident reports was seen, but even some differences of opinion.

A three-level comparison of the data was undertaken after the internal validation. As part of a first reclassification, the two patient safety coordinators included in this study each separately reviewed about half of the 495 reports. This was followed by a second reclassification of a random sample (a random data subset of 60 reports) from the original data material. During the second reclassification, the two patient safety coordinators reclassified the data subset previously reclassified by the other coordinator during the first reclassification. During each step, the patient safety coordinators were blinded to each other's reviews/opinions and followed the guidelines set forth by the patient safety group (Figure 1).

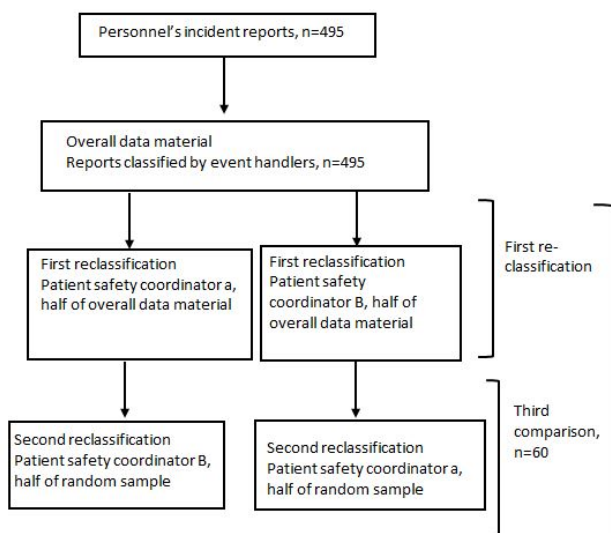


Figure 1. Three level comparison of the data.

This facilitated three pairwise comparisons between ratings (paired interrater comparisons) in the data material:

- **First interrater comparison:** Original vs. first coordinator rating (n=495).
- **Second interrater comparison:** Original vs. second coordinator rating (n=60).
- **Third interrater comparison:** First coordinator vs. second coordinator rating (n=60).

Analysis was based on free text descriptions of the incidents; the original patient charts were not traced.

Data analysis

Statistical analyses were performed with the IBM SPSS Statistics 25 and R 4.0.1 [15]. Software packages. Categorical variables are presented as frequencies and percentages and continuous data as median and range. For values, 95% Confidence Intervals (CI) were produced where appropriate and technically possible.

The following seven patient safety variables were included in our analyses: Nature of the incident, type of incident, patient impact, treating unit impact, circumstances and contributory factors, immediate actions taken, and risk category. The first four variables allowed the choice of several simultaneous alternatives; therefore for all classes of these variables a dummy variable with dichotomous values 1/Yes and 0/No was created.

A comparison of the incident severity estimate for all three paired interrater comparisons was undertaken. Three ordinal scale variables were used to assess rating change: More serious, less serious, and unchanged. Furthermore, to assess the interrater agreement of the

parallel ratings, a comparison of the number of categories that raters selected, where allowed, was undertaken for all three paired interrater comparisons.

Using software package 'rel', interrater agreement was tested with kappa for variables with nominal scale and with weighted kappa (quadratic weights) for variables with ordinal scale to allow production of CI [16]. For variables with multivariate data, i.e. those coded to dummy variables, interrater agreement was analysed with iota coefficient using software package 'irr'. The CI of percentages was calculated using software package DescTools. For an overview of agreement levels between separate comparisons, an average of the agreement coefficients of all seven original variables was calculated, although the coefficients were not of identical type.

Regarding interrater agreement, Landis and Koch characterise κ values of <0 as no agreement, 0 to 0.20 as slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1 as almost perfect agreement.

Results

In the overall data material (n=495), 214 reports (43%) were from 15 inpatient care units (median per unit 11, range 3 to 35), 165 (33%) from 16 outpatient care units (median per unit 10, range 2 to 130), and 116 (23%) from 26 other types of units (median per unit 3, range 1 to 13). Most reports from outpatient units (130/165) were recorded at the emergency department. Event handlers had an average of 1.36 missing answers (empty, '-', or 'not known') and patient safety coordinators and average of 0.32.

In the random data subset (n=60), 26 reports (43%) were from inpatient care units, 20 (33%) from outpatient care units and 14 (23%) from other types of units. Patient safety coordinators had an average of 0.27 missing answers. For the three paired interrater comparisons, there were differences in missing values, selected categories or selected multiple categories (where relevant).

The comparison of the incident severity estimate for all three paired interrater combinations (more serious, less serious, unchanged) revealed noticeable differences between the first and second interrater comparisons for the variables nature of the incident, patient impact and risk category, with most changes towards a more serious incident (Table 1). For the first interrater comparison, the average change between the three ordinal scale variables for all variables together was towards more serious in 29% (95% CI: 27%, 32%) and towards less serious in 2% (95% CI: 0%, 5%) of incidents. The net change for the first interrater comparison was 27% (95% CI: 25%, 30%) towards a more serious incident. Some differences were seen between the second and third interrater comparisons, but to a lesser extent.

| Variables | Rating change | First interrater comparison: Original vs. first coordinator rating | Second interrater comparison: Original vs. second coordinator rating | Third interrater comparison: First coordinator vs. second coordinator rating |
|-----------|---------------|---|---|--|
|-----------|---------------|---|---|--|

| | | n | % | 95% CI | n | % | 95% CI | n | % | 95% CI |
|------------------------|--------------|------|------|----------|-----|------|----------|-----|------|----------|
| Nature of the incident | More serious | 113 | 27% | 23%, 31% | 17 | 34% | 22%, 48% | 5 | 9% | 2%, 18% |
| | Less serious | 5 | 1% | 0%, 6% | 0 | 0% | 0%, 13% | 2 | 4% | 0%, 12% |
| | Unchanged | 304 | 72% | 68%, 76% | 33 | 66% | 53%, 79% | 46 | 87% | 79%, 95% |
| | Total | 422 | 100% | | 50 | 100% | | 53 | 100% | |
| Patient impact | More serious | 162 | 43% | 38%, 48% | 29 | 73% | 60%, 86% | 8 | 17% | 6%, 28% |
| | Less serious | 1 | 0% | 0%, 6% | 0 | 0% | 0%, 14% | 3 | 6% | 0%, 18% |
| | Unchanged | 213 | 57% | 52%, 62% | 11 | 27% | 14%, 41% | 37 | 77% | 67%, 88% |
| | Total | 376 | 100% | | 40 | 100% | | 48 | 100% | |
| Risk category | More serious | 80 | 19% | 15%, 23% | 13 | 29% | 18%, 44% | 15 | 26% | 15%, 38% |
| | Less serious | 21 | 5% | 1%, 9% | 1 | 2% | 0%, 17% | 7 | 12% | 4%, 21% |
| | Unchanged | 317 | 76% | 72%, 80% | 31 | 69% | 58%, 84% | 35 | 62% | 49%, 74% |
| | Total | 418 | 100% | | 45 | 100% | | 57 | 100% | |
| All variables together | More serious | 355 | 29% | 27%, 32% | 59 | 44% | 36%, 52% | 28 | 18% | 11%, 24% |
| | Less serious | 27 | 2% | 0%, 5% | 1 | 1% | 0%, 10% | 12 | 7% | 1%, 14% |
| | Unchanged | 834 | 69% | 66%, 71% | 75 | 55% | 47%, 64% | 118 | 75% | 68%, 81% |
| | Total | 1216 | 100% | | 135 | 100% | | 158 | 100% | |
| | Net effect | 328 | 27% | 25%, 30% | 58 | 43% | 35%, 52% | 16 | 10% | 6%, 15% |

Table 1. Comparison of incident severity estimate, seen as rating change, three ordinal scale variables, all three paired interrater combinations. Changes in severity in number (n), percentage (%) and 95% Confidence Interval (CI). Each report classified once by either first or second coordinator.

In the comparison of the number of categories that raters selected, where multiple category selection was allowed, as seen in the first interrater comparison the patient safety coordinators clearly selected more categories than the original event handler: on average 33%

(95% CI: 31%, 35%) versus 7% (95% CI: 6%, 8%), respectively (Table 2). Similar differences were seen in the second and third interrater comparisons as well as between the patient safety coordinators, but to a lesser extent.

| Material | Data material, n=495 | | | | | | Random sample data material, n=60 | | | | | | | | |
|--|-----------------------------|----|---------|--|-----|----------|-----------------------------------|-----|---------|--------------------------|-----|----------|---------------------------|-----|----------|
| | First interrater comparison | | | Second interrater comparison-Third interrater comparison | | | | | | | | | | | |
| Variable | Original | | | First coordinator rating | | | Original | | | First coordinator rating | | | Second coordinator rating | | |
| | n | % | CI | n | % | CI | n | % | CI | n | % | CI | n | % | CI |
| Type of incident | 30 | 6% | 4%, 8% | 111 | 22% | 19%, 26% | 1 | 2% | 0%, 13% | 9 | 15% | 3%, 28% | 28 | 47% | 35%, 60% |
| Treating unit impact | 25 | 5% | 3%, 7% | 291 | 59% | 54%, 63% | 4 | 7% | 0%, 18% | 34 | 57% | 45%, 70% | 44 | 73% | 63%, 85% |
| Circumstances and contributory factors | 45 | 9% | 7%, 12% | 188 | 38% | 34%, 43% | 6 | 10% | 0%, 22% | 21 | 35% | 23%, 49% | 30 | 50% | 38%, 64% |
| Immediate actions taken | 41 | 8% | 6%, 11% | 58 | 12% | 9%, 15% | 8 | 13% | 7%, 22% | 6 | 10% | 0%, 24% | 8 | 13% | 7%, 22% |
| On average | 35,25 | 7% | 6%, 8% | 162 | 33% | 31%, 35% | 4,75 | 8% | 5%, 11% | 17,5 | 29% | 24%, 35% | 27,5 | 46% | 40%, 53% |

Table 2. Comparison of multiple number of categories raters selected, where allowed, all three paired interrater combinations, in number (n), percentage (%) and 95% Confidence Interval (CI). Each report classified once by a single coordinator, as either first or second coordinator.

Regarding interrater agreement for the seven variables examined and their categories, where multiple category selection was allowed and according to coefficients (kappa, weighted kappa or iota), for

most variables moderate agreement was seen in the first interrater comparison, likewise in the second and third interrater comparisons (Table 3).

| Patient safety variable and subcategories | Coefficient | First interrater comparison: Original vs. first coordinator rating | | | Second interrater comparison: Original vs. second coordinator rating | | | Third interrater comparison: First coordinator vs. second coordinator rating | | |
|--|----------------|--|--------------|-----|--|--------------|----|--|--------------|----|
| | | Coeff. | 95% CI | n | Coeff. | 95% CI | n | Coeff. | 95% CI | n |
| Nature of the incident | kappa | 0.455 | 0.386, 0.525 | 495 | 0.265 | 0.041, 0.490 | 60 | 0.537 | 0.253, 0.821 | 57 |
| Type of incident | iota | 0.826 | | 495 | 0.687 | | 60 | 0.602 | | 57 |
| Information flow or management | kappa | 0.801 | 0.745, 0.857 | | 0.514 | 0.297, 0.732 | | 0.471 | 0.253, 0.688 | |
| Medicinal, blood transfusion, shadow or marker | kappa | 0.846 | 0.793, 0.900 | | 0.75 | 0.536, 0.964 | | 0.846 | 0.678, 1.015 | |
| Related to the device or its use | kappa | 0.884 | 0.782, 0.985 | | 0.783 | 0.480, 1.085 | | 0.782 | 0.491, 1.072 | |
| Other treatment | kappa | 0.739 | 0.639, 0.839 | | 0.615 | 0.317, 0.914 | | 0.612 | 0.342, 0.883 | |
| Laboratory, Radiology, or Other diagnostic examination | kappa | 0.829 | 0.728, 0.929 | | 0.813 | 0.603, 1.000 | | 0.708 | 0.469, 0.948 | |
| Patient impact | Weighted kappa | 0.428 | 0.341, 0.515 | 376 | 0.237 | 0.069, 0.406 | 40 | 0.733 | 0.557, 0.908 | 48 |
| Treating unit impact | iota | 0.228 | | 495 | 0.149 | | 50 | 0.219 | | 57 |
| Reputation damage | kappa | 0.106 | 0.043, 0.169 | | 0.124 | 0.082, 0.330 | | 0.097 | 0.000, 0.366 | |
| Extra work | kappa | 0.326 | 0.251, 0.400 | | 0.141 | 0.105, 0.388 | | 0.307 | 0.063, 0.551 | |
| Circumstances and contributory factors | iota | 0.49 | | 495 | 0.491 | | 50 | 0.368 | | 57 |
| Communication or data flow | kappa | 0.544 | 0.464, 0.623 | | 0.483 | 0.221, 0.744 | | 0.524 | 0.299, 0.749 | |
| Education, orientation, skills | kappa | 0.481 | 0.322, 0.640 | | 0.408 | 0.098, 0.913 | | 0.354 | 0.067, 0.815 | |
| Procedures | kappa | 0.348 | 0.253, 0.443 | | 0.42 | 0.126, 0.713 | | 0.606 | 0.398, 0.815 | |
| Working environment, tools, resources | kappa | 0.645 | 0.547, 0.747 | | 0.652 | 0.355, 0.949 | | 0.539 | 0.274, 0.804 | |

| | | | | | | | | | |
|--|---------------------|-------|--------------|-----|-------|--------------|----|-------|--------------|
| Immediate actions taken | iota | 0.554 | | 444 | 0.829 | | 50 | 0.574 | 56 |
| Not known | kappa | 0.495 | 0.383, 0.607 | | 0.618 | 0.291, 0.944 | | 0.51 | 0.169, 0.851 |
| Action to mitigate the consequences and prevent further damage | kappa | 0.691 | 0.597, 0.786 | | 0.755 | 0.518, 0.991 | | 0.508 | 0.229, 0.787 |
| Deviation/error correction (treatment) action | kappa | 0.751 | 0.688, 0.814 | | 0.766 | 0.586, 0.947 | | 0.601 | 0.390, 0.812 |
| Patient monitoring / patient information | kappa | 0.873 | 0.813, 0.932 | | 0.947 | 0.843, 1.000 | | 0.64 | 0.396, 0.883 |
| Risk category | Weighted kappa | 0.729 | 0.672, 0.786 | 418 | 0.407 | 0.114, 0.701 | 45 | 0.291 | 0.051, 0.632 |
| All original variables together | Average coefficient | 0.53 | | | 0.44 | | | 0.47 | |

Table 3. Interrater agreement for variables and variable categories, where multiple category selection was allowed. Coefficient as kappa, weighted kappa or iota where appropriate. Estimate of agreement is presented as 95% confidence interval where technically possible.

For all three paired interrater comparisons, the average interrater agreements were in the range of 0.44 to 0.53 and considered moderate. The lowest coefficient about 0.1 in all three interrater comparisons was seen under the variable treating unit impact for the category damage to reputation. As seen in the overall data material (n=495), the event handlers estimated that reputational damage to a unit had occurred in only 15% (76) of cases, whereas the patient safety coordinators estimated that such had occurred in 77% (378) of cases.

Discussion

Statement of principal findings

Regarding interrater agreement, the test retest reliability between the event handlers and patient safety coordinators varied quite widely and is on average only moderate. However, this was not caused by random variation (reliability problems) solely. Even systematic variation (validity problems) in the ratings was seen. We found that the event handlers tended to underestimate the severity of incidents. Moreover, where analysis for such was allowed, we discerned that the event handlers also reported fewer multiple categories, only less than a quarter when compared to the patient safety coordinators. The differences were so substantial that for all classifications by event handlers the mean values were outside the 95% CI of the classifications by the patient safety coordinators. Likewise, moderate agreement was seen between the classifications by the two patient safety coordinators, but with less bias.

Strengths and limitations

All methods were carried out in accordance with relevant guidelines and regulations. The research was conducted following good scientific

practices. We sought to assess the reliability of the incident reporting ratings collected in a hospital setting by comparing the routine ratings of frontline event handlers responsible for evaluating such with the parallel ratings of two trained patient safety coordinators. The data material was collected from a secondary care teaching hospital in Western Finland, where a systematic patient safety research and improvement project has been on-going for the past 15 years. We conclude that the findings seen in this study are indicative and that the poor reliability and bias discerned from the material studied are not exaggerated. In the overall data material (n=495), the differences seen between the ratings of frontline event handlers and patient safety coordinators were of such magnitude that the mean values for each rater group were outside the 95% CI of the other. The results for the random sample (random data subset, n=60) used to assess the interrater agreement of the parallel ratings of the patient safety coordinators are more approximate. It may be beneficial to include a larger sample size and more settings, preferably several hospitals, in future [16].

The patient safety coordinators were blinded to each other's reviews/opinions for their original assessments during each step, but not to the reviews/opinions of the other rater group included in this study, which could have caused some bias. Nevertheless, the interrater reliability was moderate, which supports the reasonable objectivity of the patient safety coordinators' assessments.

Interpretation within the context of the wider literature

During the search for literature, only one prior study was found that employed a method similar to the method used here. Our study findings are in line with previous research on the reliability of incident reporting systems, in which the reliability and accuracy of routine patient incident recordings has been found to be considerably low.

In recent guidance from the WHO, emphasis is placed on the need for personnel and event handlers to be given better instruction in the use of incidence reporting systems. Others find that incident report ratings can vary from unit to unit because there are different event handlers. By strengthening an organisation's safety culture, it is possible to influence individual attitudes and thereby safety incident handling processes. To achieve the highest possible quality of reporting, personnel must be trained in the evaluation and handling of safety incident reports.

As noted previously, safety incident report training and safety development can and should still be improved. Incomplete incident reports and analyses are a quality challenge that requires urgent safety development measures. The development of safety and thereby safety incident systems has previously been perceived in part to be an administrative issue. We argue that the entire underlying concept and associated training should be developed further, with particular attention being paid to an open safety culture in which all individuals 'dare' report safety incidents. Holmström et al. find that the reclassification of safety incident report notifications increases consistency and improves reliability. Walsh et al. maintain that generalise ability analyses may be one way to improve reliability. We find that differences in reliability cannot be resolved through cultural and educational change alone, but instead advocate for the establishment of improved terminology and incident categorisation classifications.

Implications for policy, practice and research

The patient safety reporting and assessment systems currently widely used, in which adverse incidents, near misses or unsafe conditions are identified, are unsatisfactory and demonstrate only moderate reliability. There is even a tendency to underestimate the severity of harm caused to the patient and/or the impact such has on the treating unit. Possible bias and limited reporting validity significantly hinder the assessment and measurement of patient safety, including the identification of relevant areas that can be improved and development measures.

Personnel and frontline managers need more detailed instructions on and continuous training in adverse event categorisation. Also, while patient safety coordinators should in theory represent a 'golden standard', the interrater agreement between the patient safety coordinators included in this study was moderate. Validated and more accurate ways to recognise and characterise potential safety incidents should be developed.

Conclusion

Consensus at national level on how to classify high risk incidents is needed to develop incident reporting reliability. Uniform standards for accident classification could constitute a reference at national level and even help define priorities for safety development. Our results confirm the need for such development measures at national level.

Incident reporting systems require human labour and costs; it is important to understand the validity and/or disadvantages of such systems when seeking to improve quality and security within an organisation. Even in hospital settings, where patient safety has

steadily improved, the primary handlers of safety incidents handle a wide variety of reports. Continuous training in common practices; terminology and rating systems should be given more attention. Having a patient safety coordinator reclassify incident reports within an organisation can improve reporting accuracy and thereby the targeting of corrective actions and learning.

Ethics Approval and Consent to Participate

This study received approval from the chief administrative physician of Vasa central hospital. Vaasan central hospital's ethics committee did not require the ethics board's approval during the research permit application phase. No further ethical approval was therefore necessary, which is in accordance with the regulatory regime for conducting health research in Finland.

The data used in the research is stored in the organization's database, and permission is granted to conduct research from database. Therefore, those who participated in the study are aware and accept their participation in the study.

There is no personal information in the used data and incident report was made anonymously. Strobe Statement checklist was used writing this study.

Conflict of Interests

No known conflict of interests.

Acknowledgments

None.

Author Contribution

All authors have read and approved the final manuscript.

References

1. Milch, Catherine E, Deeb N Salem, Stephen G Pauker, and Thomas G Lundquist, et al. "Voluntary Electronic Reporting of Medical Errors and Adverse Events." *J Gen Intern Med* 21 (2006): 165-170.
2. Howell, Ann-Marie, Elaine M Burns, Louise Hull, and Erik Mayer, et al. "International Recommendations for National Patient Safety Incident Reporting Systems: an Expert Delphi Consensus-Building Process." *BMJ Qual Saf* 26 (2017): 150-163.
3. Forster, Alan J, Burnand Bernard, Saskia E Drosler, and Yana Gurevich, et al. "A World Health Organization Field Trial Assessing A Proposed ICD-11 Framework for Classifying Patient Safety Events." *Int J Qual Health Care* 29 (2017): 548-556.
4. Hamilton, Emma C, Dean H Pham, Andrew N. Minzenmayer, and Mary T Austin, et al. "Are We Missing the Near Misses in the OR?—Underreporting of Safety Incidents in Pediatric Surgery." *J Surg Res* 221 (2018): 336-342.
5. Chang, Andrew, Paul M Schyve, Richard J Croteau, and Dennis S O'Leary, et al. "The JCAHO Patient Safety Event Taxonomy: a Standardized Terminology and Classification Schema for Near Misses and Adverse Events." *Int J Qual Health Care* 17 (2005): 95-105.

6. Moeller, Anders Damgaard, Kurt Rasmussen, and Kent Jacob Nielsen. "Learning and Feedback From the Danish Patient Safety Incident Reporting System Can be Improved." *Dan Med J* 63 (2016): 1-4.
7. Howell, Ann-Marie, Elaine M Burns, George Bouras, and Liam J Donaldson, et al. "Can Patient Safety Incident Reports be Used to Compare Hospital Safety? Results from a Quantitative Analysis of the English national Reporting and Learning System Data." *PloS one* 10 (2015): 0144107.
8. Classen, David C, Roger Resar, Frances Griffin, and Frank Federico, et al. "Global Trigger Tool'shows that Adverse Events in Hospitals may be ten Times Greater Than Previously Measured." *Health aff* 30 (2011): 581-589.
9. Klein, Dorte O, Roger Rennenberg, Rijk Gans and Roelien Enting, et al. "Limited External Reproducibility Restricts the use of Medical Record Review for Benchmarking." *BMJ Open Qual* 8 (2019): 000564.
10. Landis, J Richard, and Gary G Koch. "The Measurement of observer Agreement for Categorical Data." *Biometrics* 33 (1977): 159-174.
11. Holmström, Anna-Riia, Riina Jarvinen, Raisa Laaksonen, and Timo Keistinen, et al. "Inter-rater Reliability of Medication Error Classification in a Voluntary Patient Safety Incident Reporting System HaiPro in Finland." *Res Social Adm Pharm* 15 (2019): 864-872.
12. Sari, Ali Baba-Akbari, Trevor A Sheldon, Alison Cracknell, and Alastair Turnbull, et al. "Sensitivity of Routine System for Reporting Patient Safety Incidents in an NHS Hospital: Retrospective Patient Case Note Review." *Bmj* 334 (2007): 79.
13. Vincent, Charles. "Incident Reporting, and Patient Safety." *BMJ* 334 (2007): 51-51.
14. Lee, Kwangmi, Kyeongsuk Yoon, Byeongsook Yoonm, and Eunhee Shin, et al. "Differences in the Perception of Harm Assessment among Nurses in the Patient Safety Classification System." *PloS one* 15 (2020): 0243583.
15. Ridelberg, Mikaela, Kerstin Roback, Per Nilsen, and Siw Carlford, et al. "Patient Safety Work in Sweden: Quantitative and Qualitative Analysis of Annual Patient Safety Reports." *BMC Health Serv Res* 16 (2016): 1-9.
16. Lambrinos AH. "Patient Safety Learning Systems: a Systematic Review and Qualitative Synthesis." *Ont Health Technol Assess Ser* 17 (2017): 1.

How to cite this article: Plukka, Mari, Auvo Rauhala, Tuija Ikonen and Lisbeth Fagerström, et al.. "Current Incident Reporting and Learning Systems Yield Unreliable Information on Patient Safety Incident Reporting Reliability." *Clin Case Rep* 13 (2023): 1557.