**DTU Library**

# Validating the Danish adaptation of the World Health Organization's International Classification for Patient Safety classification of patient safety incident types

Mikkelsen, Kim Lyngby; Thommesen, Jacob; Andersen, Henning Boje

[Link back to DTU Orbit](#)

# Validating the Danish adaptation of the World Health Organization's International Classification for Patient Safety classification of patient safety incident types

KIM LYNGBY MIKKELSEN[1], JACOB THOMMESEN[2] AND HENNING BOJE ANDERSEN[2]

[1]National Agency for Patients' Rights and Complaints, Copenhagen, Denmark, and [2]Technical University of Denmark, Management Engineering, Kgs. Lyngby, Denmark

Address reprint requests to: Kim Lyngby Mikkelsen, National Agency for Patients' Rights and Complaints, Copenhagen, Denmark, Frederiksborggade 15, Copenhagen DK-1360, Denmark. Fax: +45-7222-7411; E-mail: kilm@patientombuddet.dk

## Abstract

**Objectives.** Validation of a Danish patient safety incident classification adapted from the World Health Organizaton's International Classification for Patient Safety (ICPS-WHO).

**Design.** Thirty-three hospital safety management experts classified 58 safety incident cases selected to represent all types and subtypes of the Danish adaptation of the ICPS (ICPS-DK).

**Outcome Measures.** Two measures of inter-rater agreement: kappa and intra-class correlation (ICC).

**Results.** An average number of incident types used per case per rater was 2.5. The mean ICC was 0.521 (range: 0.199–0.809) and the mean kappa was 0.513 (range: 0.193–0.804). Kappa and ICC showed high correlation ($r = 0.99$). An inverse correlation was found between the prevalence of type and inter-rater reliability. Results are discussed according to four factors known to determine the inter-rater agreement: skill and motivation of raters; clarity of case descriptions; clarity of the operational definitions of the types and the instructions guiding the coding process; adequacy of the underlying classification scheme.

**Conclusions.** The incident types of the ICPS-DK are adequate, exhaustive and well suited for classifying and structuring incident reports. With a mean kappa a little above 0.5 the inter-rater agreement of the classification system is considered 'fair' to 'good'. The wide variation in the inter-rater reliability and low reliability and poor discrimination among the highly prevalent incident types suggest that for these types, precisely defined incident sub-types may be preferred. This evaluation of the reliability and usability of WHO's ICPS should be useful for healthcare administrations that consider or are in the process of adapting the ICPS.

**Keywords:** adverse events, patient safety, incident reporting and analysis, risk management, taxonomy

## Introduction

Since 1 January 2004, reports about patient safety incidents occurring in hospitals in Denmark have been reported to a national patient safety reporting system. In a 2010 amendment to the Health Act, the reporting system was extended to include incidents occurring at private practices and pre-hospital sector including municipal health services and pharmacies, and to allow patients and relatives to report safety incidents. The extension of reporting to non-hospital sectors offered an opportunity to enhance the electronic reporting system to improve incident management, retrieval and statistics. With this upgrade,

the classification of incidents began to use an international standard: the World Health Organization's International Classification for Patient Safety (ICPS-WHO) [1–4].

Over several years, observers have called for the collection and analysis of data on patient safety incidents in order to support learning from failures and thereby to mitigate risks to patients [5–9]. One key tool for analysing incidents and extracting useful data is a classification system or taxonomy [10, 11] to capture and distinguish different types of failures and their causal factors. The WHO's World Alliance has developed the International Classification for Patient Safety (ICPS) in order to establish 'a common format to facilitate

aggregation, analysis and learning across disciplines, borders and time' [12].

Because of reporting bias, counting types of incidents, failures, problems and causes does not provide a valid picture of the true distribution [13]. Nevertheless, a classification system can support the analysis of incidents, aid the discovery of trends (e.g. same problems with infusion pumps in several places) and facilitate learning if users can share narratives about 'similar' failures and problems. It is useful also when selecting 'similar' events for subsequent 'in-depth' analysis. The Danish National Board of Health adopted the ICPS in order to contribute to international cooperation on standardization of terminology and methods, on the planning of interventions and, in general, to engage in research collaboration on further development and use of a common incident reporting system [13].

## Objective

Adapting the ICPS's incident type classification to the Danish reporting system provided the opportunity to test the validity and reliability of a prototype of the ICPS. The intended users of the classification system (front-line staff and safety managers in hospitals and the primary sector including municipal health services) can be expected to receive limited training in use of the system. It was deemed essential that the system should be easy to use and require little training beyond a succinct user guide. The purpose of the pilot test was (i) to capture and possibly correct usability problems of the classification system before its finalization and (ii) to assess the inter-rater reliability of the use of the system.

## Methods

A Patient Safety Classification Workgroup (see Acknowledgments) translated the ICPS-WHO Incident Type classification into Danish and adapted it to the Danish healthcare sector—henceforth referred to as ICPS-DK. In addition to translating the original ICPS-WHO classification terms into Danish, some elements were reorganized and the incident type 'Professional documentation' was expanded to include communication.

The core of the ICPS-DK consists of 13 main types and 16 subtypes, henceforth collectively referred to as *types* that form the mandatory part of the reporting system (see Table 1). In this mandatory part, incidents are classified according to the relevant healthcare process only, without specifying the problem or the contributing factors. Risk managers are obliged to classify any reported incident into one or more of the types defined by the mandatory part. In addition to the mandatory incident types, a detailed *optional* set of types is available to allow users to assign additional codes that may be helpful to learning (e.g. contributing factors). The rationale for defining a relatively small but mandatory set of incident types was to achieve a balance between succinctness and specificity, optimizing the information capture relative to the amount of effort required to classify cases.

### Selection of raters

Hospital safety managers were recruited to serve as raters in the study, as they had experience with classifying incidents in the prior reporting system and there were no raters available in non-hospital sectors (primary care, nursing homes, etc.). Each of the five Danish regions, who are responsible for the provision of hospital services in Denmark [14], was asked to recruit 10 safety managers and to nominate managers with prior experience classifying incidents in the previous system. There was no formal selection procedure. Two reminder e-mails were sent to non-responders. In a follow-up questionnaire, 70% of the raters reported that they were 'very experienced' in classifying adverse events suggesting that this convenience sample of raters will be typical of end users of the classification scheme.

### Selection of patient safety incident cases

The existing reporting system receives about 25 000 reports each year. A sample of 500 patient safety incident cases was selected at random from consecutive serious cases reported during 2009 with SAC score 1 or 2 (Safety Assessment Code) [15]. From this sample of 500 cases, further selection was done to produce two cases matched to each of the 29 incident types of the mandatory part of the classification. None of the 500 cases involved 'Self-harm' and so two additional cases involving 'Communication and documentation' were selected. Another two cases were selected to illustrate the systems to raters for a total of 60 cases. The cases presented to the raters were anonymized but otherwise exactly as reported.

### Test material

Participants received by e-mail instructions for the test, the user guide, the case descriptions of the 60 patient safety incident cases (average number of words = 113; range: 24–380) and the classification table (see Table 1). The user guide contained a short introduction to the system, and for each of the 29 incident types a short definition of the type and at least one example (narrative) of a typical case was given. The user guide explains briefly how to use the scheme, emphasizing that the classification is non-exclusive ('inclusive'), i.e. an incident may be assigned to one or more types. Participants were instructed to select incident types based on the information described in the case and not on speculation. Two of the 60 cases were provided as instructional examples along with classifications and explanations for the selection of types (made by the authors). Participants were promised anonymity.

**Table I** The mandatory part of the Danish adaptation of ICPS used in the pilot test (each case must be classified using at least one of the listed incident types)

| Type |
| --- |

|  |  |
| --- | --- |
|  | **Administrative processes** |
| 1 | *Handovers/shift changes/sector changes/referral* |
|  | Transfer of responsibility for patients |
| 2 | *Appointment* |
|  | An agreement or arrangement for a meeting between a patient and a healthcare professional |
| 3 | *Waiting list/waiting time/continuity break* |
|  | A queue of patients desiring appointments with a healthcare professional. Problems with continuity of care |
| 4 | *Admissions/reception* |
|  | The formal acceptance by a healthcare organization of a patient to receive health services |
| 5 | *Discharge* |
|  | Processes where the healthcare organization's or programme's active responsibility for the patient's care is terminated |
| 6 | *Patient identification* |
|  | The process of checking, confirming and/or validating who the patient is |
| 7 | *Informed consent* |
|  | The expressed, implied or documented permission of the patient to undergo a therapeutic intervention. |
| 8 | *Other/not known* |
|  | Other administrative processes |
|  | **Clinical processes** |
| 9 | *Screening/prevention/routine checkup* |
|  | Processes to identify, to minimize the impact of, or retard the progression of, a disease of a disease, as well as regular examinations |
| 10 | *Diagnosis/examination/assessment* |
|  | Processes of determining the nature of a disease or condition |
| 11 | *Treatment/intervention/monitoring* |
|  | Therapeutic actions taken to address diseases or injuries, including monitoring and control of the effects of the actions taken |
| 12 | *Care/rehabilitation* |
|  | Processes of patient's continuing care needs or strategies for providing services to meet those needs |
| 13 | *Test/survey/test results* |
|  | Processes related to the patient's tests, test specimens and/or diagnostic results, e.g. execution of, interpretation of and reaction on tests. |
| 14 | *Detention/fixation* |
|  | Processes of physical and pharmaceutical limitation of a patient |
| 15 | *Other/not known* |
|  | Other clinical processes |
| 16 | **Professional communication and documentation** |
|  | Incidents involving oral and written (including electronic) communication and documentation |
| 17 | **Medication** |
|  | Incidents involving any process related to the medication of a patient |
| 18 | **Medical equipment** |
|  | Incidents related to the use or misuse of medical equipment, including malfunctions of the equipment |
| 19 | **Infection** |
|  | Infections that are acquired in hospitals or as a result of healthcare interventions |
|  | **Blood and blood components** |
| 20 | Incidents involving any process related to the use of blood and blood components |
| 21 | **Gases and air for medical use** |
|  | Incidents involving any process related to the use of gases and air for medical use |
|  | **Self-harm, suicide attempts or suicide** |
| 22 | *Self-harm* |
|  | Incidents where a patient consciously performs self-harm without the intention to die |

*(continued)*

Table 1 Continued

| Type | |
| --- | --- |
| 23 | *Suicide attempt* |
| | Incidents where a patient attempts suicide |
| 24 | *Suicide* |
| | Incidents where a patient commits suicide |
| | **Patient accident** |
| 25 | *Fall* |
| | Incidents where a patient falls |
| 26 | *Other* |
| | Other patient accidents |
| 27 | **Buildings and infrastructure** |
| | Problems involving the basic facilities and services needed for the functioning of the healthcare organization |
| 28 | **Resources and organization** |
| | Problems involving individual, team or organizational factors, including occupational factors |
| 29 | **Other incident type** |
| | Other incident types not otherwise classifiable |

Main types are in bold and subtypes in italics font.

## Statistical analyses

Two measures of the inter-rater agreement were used: kappa and ICC (intra-class correlation), the former because it is a widely used measure of inter-rater agreement and the latter because the interpretation of kappa is controversial [8–11] due to the way it handles chance agreement. The ICC measure used is the ICC (2,1) described by Shrout and Fleiss [16]. Statistics were calculated using Stata/MP 11.1 [17].

## Results

Thirty-three of the 43 raters returned their responses. Not all raters classified all the 58 cases. Of the possible 1914 rater–cases (33 raters times 58 cases), 1619 (85%) had been completed. Several raters noted that the task took longer than expected, as discussed later.

### Per-case analysis

The average number of types used per case per rater was 2.5. Eighty-five per cent of the 1619 rater–cases were classified using three or fewer types (99% using five or fewer types), and only one rater classified a case as having seven types. When all ratings by the 33 raters are considered, the mean number of types used per case was 8.9 (range: 4–15). ICC and kappa for all 58 cases are given in Table 2. The mean ICC was 0.521 (range: 0.199–0.809) and the mean kappa was 0.513 (range: 0.193–0.804). The pairwise correlation (Pearson's $r$) between ICC and kappa was 0.998.

The length of the case descriptions was positively correlated with the inter-rater agreement: ICC increased by 1.2 percentage points for every 10-word increase in the length of the case description ($P = 0.001$). An inter-quartile range increase in the number of words in the case description (80 words) was associated with an increase in ICC of 9.7 percentage points. The three cases with the lowest and the three with the highest inter-rater agreement are copied in Table 3.

### Per-type analysis

ICC and kappa for all 29 types are given in Table 4. The mean ICC was 0.454 (range: 0.006–1.000) and the mean kappa was 0.479 (range: 0.005–1.000). The pairwise correlation (Pearson's $r$) between ICC and kappa was 0.999. There was no association between self-rated experience with the former classification system and ICC measures (data not shown).

### Prevalence of incident type use

The prevalence of use of an incident type was defined, somewhat arbitrarily, as the proportion of cases used by 'at least one' rater. The prevalence of the incident types varied considerably. Six of the 29 types were used with a prevalence >50% (see Table 4; numbers in parentheses in this section refer to the type numbers in Table 4 and Fig. 1). The most prevalent type was 'Resources and organization' (28), which was used by at least one rater in 56 of the 58 cases (prevalence = 97%). Also, 'Professional communication and documentation' (16) and 'Treatment/intervention/monitoring' (11) were used frequently, with prevalences of 90 and 79%, respectively. In contrast, 'Gases and air for medical use' (8), 'Suicide attempt' (23), 'Suicide' (24) and 'Fall' (25) had prevalences of <10%.

There was a strong inverse association between the prevalence and the inter-rater agreement (Fig. 1). However, a

**Table 2** ICC and kappa, by case

| Case | ICC | Kappa | R (T) | Case | ICC | Kappa | R (T) | Case | ICC | Kappa | R (T) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | 33 (3) | 21 | 0.81 | 0.8 | 26 (4) | 41 | 0.42 | 0.41 | 18 (11) |
| 2 | | | 33 (3) | 22 | 0.5 | 0.49 | 26 (8) | 42 | 0.67 | 0.66 | 27 (9) |
| 3 | 0.34 | 0.34 | 26 (15) | 23 | 0.52 | 0.51 | 21 (9) | 43 | 0.6 | 0.59 | 26 (10) |
| 4 | 0.48 | 0.48 | 33 (10) | 24 | 0.81 | 0.8 | 21 (5) | 44 | 0.31 | 0.3 | 13 (9) |
| 5 | 0.6 | 0.59 | 30 (9) | 25 | 0.65 | 0.64 | 28 (6) | 45 | 0.69 | 0.68 | 24 (8) |
| 6 | 0.73 | 0.73 | 32 (9) | 26 | 0.43 | 0.42 | 21 (9) | 46 | 0.24 | 0.23 | 11 (11) |
| 7 | 0.24 | 0.24 | 16 (11) | 27 | 0.56 | 0.56 | 23 (7) | 47 | 0.64 | 0.63 | 27 (6) |
| 8 | 0.75 | 0.75 | 33 (9) | 28 | 0.68 | 0.67 | 28 (7) | 48 | 0.52 | 0.51 | 24 (9) |
| 9 | 0.64 | 0.63 | 28 (8) | 29 | 0.55 | 0.54 | 24 (10) | 49 | 0.34 | 0.34 | 15 (8) |
| 10 | 0.76 | 0.75 | 33 (6) | 30 | 0.54 | 0.53 | 19 (7) | 50 | 0.45 | 0.45 | 20 (10) |
| 11 | 0.72 | 0.71 | 28 (6) | 31 | 0.49 | 0.48 | 19 (7) | 51 | 0.28 | 0.27 | 16 (10) |
| 12 | 0.48 | 0.47 | 27 (8) | 32 | 0.47 | 0.46 | 21 (9) | 52 | 0.51 | 0.5 | 21 (10) |
| 13 | 0.43 | 0.42 | 25 (7) | 33 | 0.42 | 0.42 | 19 (10) | 53 | 0.44 | 0.43 | 17 (9) |
| 14 | 0.37 | 0.36 | 21 (8) | 34 | 0.64 | 0.63 | 27 (10) | 54 | 0.46 | 0.46 | 20 (7) |
| 15 | 0.41 | 0.4 | 20 (8) | 35 | 0.49 | 0.48 | 19 (9) | 55 | 0.55 | 0.54 | 24 (10) |
| 16 | 0.38 | 0.37 | 23 (9) | 36 | 0.5 | 0.49 | 24 (9) | 56 | 0.61 | 0.6 | 25 (8) |
| 17 | 0.72 | 0.71 | 32 (6) | 37 | 0.54 | 0.53 | 20 (6) | 57 | 0.55 | 0.54 | 21 (13) |
| 18 | 0.71 | 0.7 | 24 (8) | 38 | 0.36 | 0.35 | 18 (11) | 58 | 0.6 | 0.59 | 25 (11) |
| 19 | 0.59 | 0.58 | 29 (14) | 39 | 0.38 | 0.37 | 19 (9) | 59 | 0.55 | 0.54 | 21 (10) |
| 20 | 0.42 | 0.41 | 26 (13) | 40 | 0.51 | 0.5 | 25 (11) | 60 | 0.2 | 0.19 | 9 (8) |

Case: the analyses were made case by case. Cases comprise 58 patient safety incident cases that were classified by the raters. Cases 1 and 2 were pre-classified by the authors and used for instruction. ICC, intra-class correlation. The pairwise correlation between ICC and kappa is 0.999. R (T): number of raters (number of types used by all raters combined to classify the case).

group of 'residual' incident types [e.g. 'Other/not known' (8, 15, 29)] demonstrated both very low prevalences and very low ICCs and another group of incident types ['Medication' (17), 'Medical equipment' (18) and 'Buildings and infrastructure' (27)] demonstrated relatively high prevalences and relatively high ICCs.

## Discussion

This pilot study of safety incident classification using a Danish version of the ICPS demonstrated a fair-to-good reliability of the ICPS classification with a mean inter-rater agreement kappa close to 0.5 [18]. The raters in this pilot project were volunteers who, while they were experienced at classifying hospital events, had not received formal training other than a short written introduction including two examples of pre-classified cases. In another safety classification system, the human factors analysis and classification system has reported kappa estimates as high as 0.7 [19–21], whereas others have found estimates of index of concordance as low as 0.2 [22]. However, in practice, most reporting systems will offer only minimal training to raters and so our results may be more reflective of the reliability of real practice in the field when such reporting systems are used.

Four factors serve to determine the inter-rater agreement [19]: the skill and motivation of the raters (raters' ability); the clarity of the cases (the nature of the items to be classified); the clarity of the operational definitions of the types and the directions that guide the coding process (the definition of types and the instructions) and the adequacy of the underlying classification scheme (the taxonomy).

### The skill and motivation of the raters

The raters were under artificial constraints. They were given the user guide documentation and the pilot cases at the same time, and were allowed only 1 week to return the completed ratings. Additional instruction and training of the raters may have improved inter-rater reliability. Not all raters completed the entire pilot set, with cases skipped throughout, and with some raters skipping the last 10–15 cases. Several raters noted that the rating exercise required much more than the expected 4 h. Therefore, a better estimate of the time needed to complete the pilot test or a shorter test with fewer cases could possibly have increased the inter-rater agreement, although the inter-rater agreement was not associated with and therefore did not decline with the presentation order of the 58 cases in the test material (data not shown).

### The clarity of the cases

The clarity, the complexity and especially the subject matter of the cases comprised the second group of factors determining the inter-rater agreement. The assessment of clarity and complexity is difficult, but the number of words in the case description may offer a crude measure of clarity. This is to some degree supported by the observation that the

**Table 3** Case descriptions of the three cases with lowest and highest inter rater agreement

| C | Case description |
|---|---|

60   *Event description*: patient escapes from a closed section. The patient leaves the closed section together with a maintenance worker while nobody was watching

     *Consequence of the incident*: the patient may need longer treatment

     *Type (no of raters)*: 11(3), 14(9), 15(1), 16(4), 26(3), 27(8), 28(8), 29(4). Kappa: 0.19

46   *Event description*: refuse collector blocks the door open for convenience while collecting the garbage. A dangerous person escapes from the closed section

     *Consequence of the incident*: course of treatment interrupted and thus worsening or prolongation of the acute psychotic episode, risk of vandalism or violence.

     *Type (no of raters)*: 1 (1), 3 (1), 8 (1), 11 (2), 14 (9),16 (2), 23 (1), 26 (2), 27 (10), 28 (11), 29 (6). Kappa: 0.23

7   *Event description*: a control X-ray of the chest that should have been carried out was not ordered

     *Consequence of the incident*: prolongation of hospital stay

     *Type (no of raters)*: 1 (9), 2 (10), 3 (12), 5 (1), 8 (2), 9 (2), 10 (11), 11 (2), 13 (7), 16 (16), 28 (1). Kappa: 0.24

10   *Event description*: a patient with inoperable cholangiocarcinoma and IDDM is prescribed a glucose drop as the patient is fasting and the patient potassium is high. The drop was set at ~22.45 yesterday. This morning we discover that the drop is 1 l sodium chloride/glucose added with Actrapid (insulin) and not a pure glucose solution. Overnight the patient had a need for extra sugar due to low blood sugar

     *Consequence of the incident*: patient's blood sugar dropped during the night. As the patient has had diabetes for a long time, she was able to feel the blood sugar coming down and alerted the nurse. In the worst-case scenario, it could have ended up with coma or death. Unclear how frequent causes of the incident: it was a temporary nurse who never before had been here at this department, who made the dropped yesterday evening

     *Proposals for action*: more permanent staff, more carefulness with regard to medicine, etc

     *Type (no of raters)*: 1 (2), 11 (20), 16 (6), 17 (33), 18 (1), 28 (28). Kappa: 0.75

24   *Event description*: at ~5 p.m. the furnace guard got a message from NNIT (IT support) that the voltage of the UX 9, the network server for the intensive care unit, was 0. At the investigation the furnace guard found the switch for the UX 9 on the electrical board in the basement was turned off. The board is located in a locked room and the switch can only be turned off by a physical action. To re-establish the power, two other switches in the electrical cabinet had to be dismounted, which delayed the remediation of the situation

     *Consequence of the incident*: as far as known, no consequence for the patients. The staff of the department were very troubled. The possible consequences could have been fatal

     *Causes of the incident*: the investigation found that the switch for the UX 9 on the electrical board in the basement was turned off. The electrical board is located in a locked room and can only be turned off by a physical action

     *Proposals for action*: electrical cabinets should be locked with a key. Electrical cabinets should not be over-crowded and all switches should be freely accessible.

     *Type (no of raters)*: 15 (1), 18 (1), 26 (1), 27 (21), 28 (2). Kappa: 0.80

21   *Event description*: power outage in ~20 min where the back-up power also failed to operate

     *Consequence of the incident*: no consequence. Could have had serious consequences, if we, at that time have had very unstable patients

     *Causes of the incident*: new buildings: electricity work was being carried out

     *Proposals for action*: back-up power should always work. In situations (ex. construction work) where there is a risk that departments may experience power failure, departments should be warned

     *Type (no of raters)*: 21 (2), 27 (26), 28 (2), 29 (1). Kappa: 0.80

C, case number. The types used by the rates for classification of the case and the number of rates using each type are given at the bottom of each cell. Refer to Table 1 for type titles.

number of words in the case description was strongly and positively associated with the inter-rater agreement. Presumably, more words reduce ambiguity. Furthermore, as can be inferred from Table 3, short-case descriptions can lead raters to speculate and to select incident types that may be unwarranted based on the case description. For example,

**Table 4** ICC, kappa and prevalence, by type

| Type | | ICC | Kappa | Prevalence |
|---|---|---|---|---|
| | Administrative processes | | | |
| 1 | Handovers/shift changes/sector changes/referral | 0.30 | 0.33 | 0.69 |
| 2 | Appointment | 0.32 | 0.35 | 0.24 |
| 3 | Waiting list/waiting time/continuity break | 0.20 | 0.22 | 0.64 |
| 4 | Admissions/reception | 0.26 | 0.29 | 0.24 |
| 5 | Discharge | 0.50 | 0.54 | 0.14 |
| 6 | Patient identification | 0.74 | 0.77 | 0.14 |
| 7 | Informed consent | 0.71 | 0.74 | 0.10 |
| 8 | Other/not known | 0.03 | 0.03 | 0.22 |
| | Clinical processes | | | |
| 9 | Screening/prevention/routine check-up | 0.06 | 0.07 | 0.50 |
| 10 | Diagnosis/examination/assessment | 0.26 | 0.28 | 0.60 |
| 11 | Treatment/intervention/monitoring | 0.35 | 0.38 | 0.79 |
| 12 | Care/rehabilitation | 0.51 | 0.55 | 0.38 |
| 13 | Test/survey/test results | 0.39 | 0.42 | 0.41 |
| 14 | Detention/fixation | 0.47 | 0.51 | 0.14 |
| 15 | Other/not known | 0.01 | 0.01 | 0.14 |
| 16 | Professional communication and documentation | 0.24 | 0.27 | 0.90 |
| 17 | Medication | 0.81 | 0.83 | 0.26 |
| 18 | Medical equipment | 0.63 | 0.66 | 0.29 |
| 19 | Infection | 0.38 | 0.41 | 0.17 |
| 20 | Blood and blood components | 0.60 | 0.64 | 0.10 |
| 21 | Gases and air for medical use | 0.58 | 0.62 | 0.09 |
| | Self-harm, suicide attempts or suicide | | | |
| 22 | Self-harm | . | . | . |
| 23 | Suicide attempt | 0.94 | 0.94 | 0.07 |
| 24 | Suicide | 1.00 | 1.00 | 0.03 |
| | Patient accident | | | |
| 25 | Fall | 0.93 | 0.94 | 0.07 |
| 26 | Other | 0.53 | 0.57 | 0.14 |
| 27 | Buildings and infrastructure | 0.58 | 0.62 | 0.28 |
| 28 | Resources and organization | 0.28 | 0.31 | 0.97 |
| 29 | Other incident type | 0.09 | 0.11 | 0.12 |

ICC, intra-class correlation. The pairwise correlation between ICC and kappa is 0.999. The prevalence of an incident type was defined as used by 'at least one' rater to classify the 58 cases.

the brief description of Case 7 ('A control x-ray of the chest that should have been carried out was not ordered') led to the use of 11 different incident types by raters although only 'Handovers/shift changes/sector changes/referral', 'Waiting list/waiting time/continuity break', 'Test/survey/test results' and 'Professional communication and documentation' seem warranted based on the interpretation by the authors. Although participants were instructed to select incident types based only on the case description without further speculation about the context, raters may have found it difficult to adhere to this instruction when a short case description was presented. Ten raters (30%) used 'Diagnosis/examination/assessment' to classify Case 7 perhaps misinterpreting this as a case descriptor rather than an incident type, highlighting the importance of communicating the distinctions among incident types to raters in advance. The number of words in the case description is a potentially modifiable factor. Reporters of incidents should be encouraged not to be too terse in their case description.

In this pilot test, two patient safety incident cases per incident type were selected to maximize the variance in the cases to improve the validity of the estimate of the inter-rater agreement. However, whether this choice would increase or decrease the estimates of inter-rater agreement (compared with a random sample of cases) is not clear.

## The clarity of the operational definitions of the types

The third group of factors determining the inter-rater agreement is the clarity of the operational definitions of the types and the user guide. From Fig. 1 several clusters of types emerge. The types in the lower right hand side of the figure, having low inter-rater agreement and high prevalence,
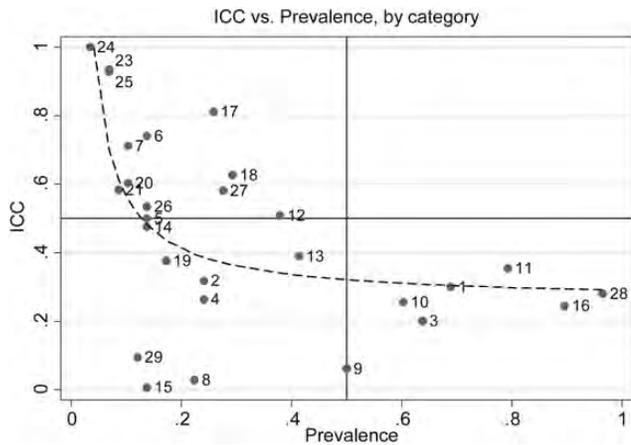
**Figure 1** ICC and the prevalence of the 29 incident types in 58 cases. The dotted line represents the best fit regression line (ICC = 0.0304/prevalence + 0.2607), showing a strong inverse association between ICC and prevalence. The numbers are the numbers of the incident types in the mandatory part of the classification, see Table 4 for type titles. The prevalence of an incident type was defined, somewhat arbitrarily, as used by 'at least one' rater to classify the 58 cases.

are characterized by being either general or unspecific, e. g. 'Handovers/shift changes/sector changes/referral'(1) and 'Waiting list/waiting time/continuity break'(3) (numbers in parentheses refer to the type numbers in Table 4 and Figure 1. These are very broad types, and moreover, problems with transfer of care and continuity break have achieved high attention in the recent years, which might have inflated the prevalence. The high prevalence of 'Diagnosis/ examination/assessment' (1) and 'Treatment/intervention/ monitoring' (11) probably indicates the high prevalence of these problems in practice. However, in defining the comprehensiveness of the mandatory part of the classification, a more detailed sub-classification was not thought to be fruitful. In the Danish version, the ICPS-WHO incident type 'Professional documentation' was complemented with communication (oral, written), and may also cover communication failures. To encourage more precision, this type has subsequently (and based on this pilot test), been expanded with several subtypes also in the mandatory part of the classification. The prevalence of 'Resources and organization' was so high that the type becomes useless, since a prevalence of near 100% means that there is little useful discrimination. Problems with resources and organizational factors may always play a role in safety incidents (relative to other incident types that are more specific). To address this, the general type has subsequently been redefined to include several subtypes that are more specific and hence discriminating.

Not surprisingly, incident types with low prevalence and high ICC, (e.g. 'Suicide' and 'Falls') are well defined and specific incident types, but rare relative to other incident types. 'Medication', 'Medical equipment' and 'Buildings and infrastructure' are similarly specific and well defined but have a very high prevalence relative to their high ICC. This cluster suggests that the inter-rater reliability of some incident types

in the lower left quarter of the plot in Fig. 1 could be improved by refining definitions and improving the user guide. Finally, in the lower left quarter of the plot in Fig. 1, the cluster of 'Other/unknown' types (8, 15, 29) are 'residual' incident types that could be expected to have a low ICC and fortunately in this instance also have a low prevalence.

## The adequacy of the classification scheme

The fourth group of factors determining the inter-rater agreement is related to the adequacy of the underlying classification scheme. The low prevalences of the 'residual' incident types (e.g. Other/not known) indicate that the underlying classification scheme is reasonably adequate and exhaustive. Others have noted that the ICPS-WHO is a conceptual framework rather than a real classification system [10, 23]—a point which is also acknowledged by the authors behind the WHO World Alliance for Patient Safety [1]. Moreover, Schulz *et al.* [23] argue that the framework ought to be regarded, not as a taxonomy or a classification system, but as an 'information model' or 'template' since it violates a number of characteristics of a proper classification system, e.g. the ICPS lacks identifiers or codes, central terms are used with apparently different meanings and there is no linkage between its 'Key Concepts' and the classification itself [23]. Nevertheless, the class of incident types, (being one of the 10 classes of the ICPS) offers a detailed classification system that seems to capture the variety of medical task contexts in which adverse events occur; and the results of our pilot test seem to indicate that when slightly modified it is a useable and reasonably reliable tool when put into practice.

## Conclusion

Judged from the kappa and ICC statistic, the overall inter-rater agreement in this study was 'fair' to 'good' (using the labels proposed by Fleiss [10]), which is a surprisingly high level considering that participants had no training and performed the classification after a short, uncontrolled reading of a short guideline. The pilot test gives some directions about how inter-rater agreement could be improved. Reporters of incidents should be encouraged to be precise but avoid being too terse when describing a case. The broad incident types with high prevalence and low ICC should be subdivided into more specific subtypes, or their use should be restricted by strict definitions and further instruction in the user guide. Raters should be instructed to use only types warranted by the information in the case description and to refrain from speculating, and in general, raters should be trained in applying the type definitions.

The class of incident types of the ICPS appears adequate, exhaustive and well suited for classifying and structuring incidents reports. At the same time, since incident types represent adverse events at levels that are clinically meaningful, safety managers and other healthcare professionals should find them to be useful in classifying and retrieving incidents.

## Acknowledgements

## Funding

## References

1. World Alliance for Patient Safety. International Classification for Patient Safety (ICPS), 2011. Taxonomy website: http://www.who.int/patientsafety/taxonomy/en/.

2. World Alliance for Patient Safety. The conceptual framework for the International Classification for Patient Safety (v1.1). Final technical report, 2011. http://www.who.int/patientsafety/taxonomy/icps_full_report.pdf.

3. Runciman W, Hibbert P, Thomson R *et al*. Towards an International Classification for Patient Safety: key concepts and terms. *Int J Qual Health Care* 2009;**21**:18–26.

4. Thomson R, Lewalle P, Sherman H *et al*. Towards an International Classification for Patient Safety: a Delphi survey. *Int J Qual Health Care* 2009;**21**:9–17.

5. Battles JB, Stevens DP. Adverse event reporting systems and safer healthcare. *Qual Saf Health Care* 2009;**18**:2.

6. Hutchinson A, Young TA, Cooper KL *et al*. Trends in health-care incident reporting and relationship to safety and quality data in acute hospitals: results from the National Reporting and Learning System. *Qual Saf Health Care* 2009;**18**:5.

7. Department of Health Expert Group. An organisation with a memory: report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. London: Department of Health, 2000.

8. Kohn LT, Corrigan J, Donaldson MS *et al*. *To Err is Human*. Washington, DC: National Academy Press, 2000.

9. Secker-walker J, Taylor-Adams S. Clinical incident reporting. In: Vincent C (ed.). *Clinical risk management: Enhancing patient safety. 2nd edition*. London: BMJ Books, 2001, 419–38.

10. Taib IA, McIntosh AS, Caponecchia C *et al*. A review of medical error taxonomies: a human factors perspective. *Saf Sci* 2011;**49**:607–15.

11. Rasmussen J. Human errors. A taxonomy for describing human malfunction in industrial installations* 1. *J Occup Accid* 1982;**4**:311–33.

12. World Health Organization. WHO draft guidelines on adverse event reporting and learning systems, Geneva, Switzerland: World Health Organization Press, 2005. http://www.who.int/patientsafety/events/05/Reporting_Guidelines.pdf (December 2012, date last accessed).

13. Pham JC, Gianci S, Battles J *et al*. Establishing a global learning community for incident-reporting systems. *Qual Saf Health Care* 2010;**19**:446.

14. Danish Regions' website, 2011. http://www.regioner.dk/In+English.aspx.

15. Bagian JP, Lee C, Gosbee J *et al*. Developing and deploying a patient safety program in a large health care delivery system: you can't fix what you don't know about. *Jt Comm J Qual Improv* 2001;**27**:522–32.

16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.

17. StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP [computer program], 2009.

18. Fleiss JL, Levin BA, Paik MC. Statistical methods for rates and proportions. 3rd edition. Hoboken, NJ: Wiley-Interscience, 2003.

19. Shappell SA, Wiegmann DA. Applying reason—the human factors analysis and classification system (HFACS). *Hum Factors Aerospace Saf* 2001;**1**:59–86.

20. Milligan FJ. Establishing a culture for patient safety-the role of education. *Nurs Educ Today* 2007;**27**:95.

21. ElBardissi AW, Wiegmann DA, Dearani JA *et al*. Application of the human factors analysis and classification system methodology to the cardiovascular surgery operating room. *Ann Thorac Surg* 2007;**83**:1412–9.

22. Olsen NS, Shorrock ST. Evaluation of the HFACS-ADF safety classification system: inter-coder consensus and intra-coder consistency. *Accid Anal Prev* 2010;**42**:437–44.

23. Schulz S, Karlsson D, Daniel C *et al*. Is the International Classification for Patient Safety a classification? *Stud Health Technol Inform* 2009;**150**:502.