OPEN

The Korea National Patient Safety Incidents Inquiry Survey: Feasibility of Medical Record Review for Detecting Adverse Events in Regional Public Hospitals

Min Ji Kim, PhD Cand, Hee Jung Seo, RN, MPH,* Hong Mo Koo, MD,* Minsu Ock, MD, PhD,†‡ Jee-In Hwang, RN, PhD,§ and Sang-Il Lee, MD, PhD‡*

Objectives: We aimed to examine the Korea National Patient Safety Incidents Inquiry conducted in the Republic of Korea; specifically, we assessed the validity of screening criteria, interreviewer reliability, quality of medical records, and the time required for reviewing medical records.

Methods: A 3-stage retrospective medical record review was performed. The sensitivity and positive predictive value of the screening criteria for the adverse events were calculated, and interreviewer reliability was verified using the overall agreement rate and κ value. In addition, the results of medical record quality assessment and time required for review were analyzed.

Results: There were a total of 4159 patients (55.5%) with at least 1 of the 41 screening criteria. In stage 1, the overall percent of agreement was 81.9% when all negatives from the 2 reviewers were included, and the κ value was 0.64 (95% confidence interval [CI], 0.61–0.66). In stage 2, 84.6% of cases were a perfect match, and 87.4% were a partial match. The κ values were 0.159 (95% CI, 0.12–0.20) and 0.389 (95% CI, 0.35–0.43), respectively. The mean quality assessment scores were 3.18 of 4 points in stage 1 and 3.05 of 4 points in stage 2. In stage 1, it took an average of 13.02 minutes to asses each patient file; in stage 2, it took an average of 5.06 minutes.

Conclusions: To increase the feasibility of medical record review for detecting adverse events, it is important not only to improve the reliability between reviewers but also to monitor the quality of medical records and the time required for review.

Key Words: adverse event, medical record review, validation study, feasibility study

(*J Patient Saf* 2022;18: 389–395)

To prioritize patient safety issues and determine the effects of various patient safety interventions, it is essential to examine the status of patient safety and follow up on patient safety trends.[1,2] Various methods, each with unique strengths and weaknesses, are used to examine the status of patient safety.[3,4] One representative example is the patient safety reporting and learning system that prepares and disseminates the strategy for improvement based on the details of reports on patient safety incidents from medical professionals. However, there may be issues in assessing patient safety solely based on this system because it relies on self-reports, and the incidence of reports varies depending on the medical institution's level of awareness of patient safety.[5]

As such, studies have aimed to review medical records to comprehensively identify the status of patient safety.[6] Because the Harvard Medical Practice Study was conducted in the United States to examine the incidence and characteristics of adverse events through a retrospective review of medical records of patients,[7,8] the method used in that study has been used worldwide, including in Australia, Canada, the Netherlands, and Sweden, to conduct a medical record review at the national level.[9–17] Although medical record reviews are regarded as the criterion standard for identifying adverse events, studies have indicated the importance of improving reliability among reviewers, including physicians and nurses.[6,18,19] In addition, improving the feasibility of medical record review for the identification of adverse events requires resolving issues of high cost, its time-consuming nature, and the importance of the quality of medical records.[3,4]

In the case of the Republic of Korea (hereinafter Korea), the Korea National Patient Safety Incidents Inquiry (KNPSII) survey was conducted to identify the incidence and characteristics of adverse events after the establishment of the first Comprehensive Plans for Patient Safety as a part of the 5-year plan to improve patient safety and quality of medicine. Specifically, a 3-stage retrospective medical record review was conducted in 15 regional public hospitals in Korea to determine the efficacy of using medical record review to identify adverse events.[18,20–22] The detailed characteristics of the adverse events identified in this particular survey have been discussed in a previous article.[23]

This study focused on the feasibility of the KNPSII survey. Specifically, the study aimed to examine the validity of the screening criteria, interreviewer reliability, the quality of medical records, and the amount of time needed to review medical records. Previous research has focused on the reliability of medical records[6,19] and rarely reported the aspects to be considered in terms of feasibility. We expect the results of this study to contribute to the improvement of the methodology of medical record review for the identification of adverse events to enhance feasibility.

METHODS

Study Design

This study conducted a 3-stage retrospective medical record review in 15 regional public hospitals in Korea based on the methodology of previous studies.[7,16,20] It was conducted in accordance with the first Comprehensive Plans for Patient Safety and was exempted from ethical approval in accordance with Article 2 Paragraph 2 Subparagraph 1 of the Enforcement Rule of the Bioethics and Safety Act.[24]

From the *Korea Institute for Healthcare Accreditation, Seoul; †Department of Preventive Medicine, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan; ‡Department of Preventive Medicine, University of Ulsan College of Medicine; and §Department of Nursing Management, College of Nursing Science, Kyung Hee University, Seoul, Republic of Korea.

Correspondence: Minsu Ock, MD, PhD, Ulsan University Hospital, University of Ulsan College of Medicine, 877 Bangeojinsunhwando-ro, Dong-gu, Ulsan 44033, Republic of Korea (e-mail: ohohoms@naver.com).

*ORCIDs: M.J.K., https://orcid.org/0000-0002-5780-2341; H.J.S., https://orcid.org/0000-0001-8512-8051; H.M.K., https://orcid.org/0000-0002-0774-3240; M.O., https://orcid.org/0000-0001-9949-9224; J.-I.H., https://orcid.org/0000-0003-2284-5795; S.-I.L., https://orcid.org/0000-0002-1068-7542.

This research was funded by the Ministry of Health & Welfare, Republic of Korea.

The authors disclose no conflict of interest.

Medical Record Review

In stage 1 of the medical record review, 2 nurses independently verified the presence of 41 screening criteria in the electronic medical records of the patients. In stage 2 of the review, 2 physicians independently assessed the incidence of adverse events for the patients identified in stage 1 of the review to complete a detailed adverse event report. In stage 3, in accordance with suggestions from a previous study,[20] the results of stage 2 were reviewed further by a review committee, which consisted of individuals with experience in a previous related study or experts in the field of patient safety. The results for each institution were summarized based on details of the adverse events determined in stage 3, including the number of events, degree of harm, causality, and preventability, to determine the final adverse events by pooling the outcomes from each institution. An interim audit was conducted by the institution during the review to share review experience among the reviewers and discuss any differences in opinion so that consistent criteria can be used in the review. In case queries arose during the review, consistent answers were provided through an expert review through the management of the corresponding institution, which was also shared with other reviewers.

Measures

In the stage 1 review, data including reviewer information (duration of review, etc.), patient information (sex, duration of hospitalization, department of admission, route of admission, etc.), presence of screening criteria, and medical records quality assessment were collected. The screening criteria included the following 41 items in 7 areas in total: "general: admission, discharge, and transfer"; "events"; "drug, fluid, blood"; "test results"; "related surgery"; "obstetrics and neonates"; and "others." Any patient with at least one screening criterion identified by at least one reviewer in the stage 1 review progressed to stage 2 for further review.

In stage 2 of the review, reviewer information (duration of review, etc.), incidence of adverse events, detailed adverse event report (type of adverse event, degree of harm, causality, preventability, etc.), and medical record quality assessment were collected. The quality of medical records in stage 1 and stage 2 review was evaluated as a single comprehensive question on a 4-point scale as follows: 1 (very poor), 2 (poor), 3 (good), and 4 (very good).

Patient Selection

This survey was conducted at 15 general hospital-level medical institutions with at least 200 beds among 33 regional public hospitals in Korea. We used data from patients who were discharged in 2016. Using the random number table, data from 500 patients who were discharged on a randomly extracted date between January 1, 2016, and December 31, 2016, were reviewed for each institution, and a total of 7500 discharged patients were surveyed (500 per institution). The number of samples was calculated by institution based on similar studies.[21,25]

Data were included if the patient records were complete with all discharge summaries and coding, if the patient was in the hospital for at least 24 hours, and if the patient was 19 years or older. Data were excluded if the patient was admitted for the first time at the Department of Traditional Korean Medicine, Dental Care, Psychiatry, or Rehabilitation Care. The scope of review for medical records was 1 year before and after the discharge date of the index admission. For example, if September 1, 2016, was selected as the discharge date, the scope of review for medical records was between September 1, 2015, and September 1, 2017.

Education of Reviewers

For stages 1 and 2, two reviewers who were trained in the method of reviewing medical records were recruited internally from each institution. Reviewers were recruited from an externally affiliated institution if it was difficult to recruit internally. A total of 30 reviewers were selected for stages 1 and 2, and attended a class about patient safety–related terminology and the identification of adverse events through medical record review. Two experts with previous experiences in conducting medical record review studies to detect adverse events educated reviewers. The class took about 1.5 hours. Subsequently, reviewers took part in an approximately 1.5-hour practice review using 2 sample medical records and case studies on the score for causality and preventability by each adverse event. The case studies were discussed to ensure that consistent review criteria could be established; they mostly included cases in which there was confusion in the decision.

Statistical Analysis

Review results were summarized using Microsoft Office Excel 2013 (Microsoft Corporation, Redmond, Washington), and descriptive analyses were conducted using IBM SPSS 20 (IBM, Armonk, New York). The sensitivity and positive predictive values were calculated for the adverse events, and interreviewer reliability was verified using the overall agreement rate and the κ value. The reliability of the stage 1 results was based on whether there was a need for a second review by reviewing at least one screening criterion for a particular patient. Reliability of the stage 2 results was based on the agreement in the reviewers' opinions about the occurrence of an adverse event, the degree of harm, the causality, and the preventability. "Perfect match" was defined as agreement in all of the opinions for the occurrence of an adverse event, the degree of harm, the causality, and the preventability; "partial match" was defined as agreement on the decision for adverse event but disagreement in the degree of harm, causality, or preventability. In addition, the quality assessment and duration of review of medical records were compared and examined among the institutions and the stage 1 and 2 reviewers.

RESULTS

There were a total of 4159 patients with at least 1 of the 41 screening criteria, accounting for 55.5% of all patients. The highest incidence was found for "specialty consult," followed by "length of index was over 30 days," "transfusion or use of blood products," "in-unit procedure," and "decrease in hemoglobin or hematocrit of 25% or greater" (Table 1). The positive predictive value in stage 1 was 17.9%. The screening criterion with the highest positive predictive value was "mechanical ventilation greater than 24 hours post-op." The highest sensitivity was observed for "specialty consult," followed by "length of index was over 30 days" and "in-unit procedure."

In stage 1, the overall percents of agreement were 81.9% when all negatives from the 2 reviewers were included and 67.3% when all negatives from the 2 reviewers were excluded (Table 2). Here, the κ value was 0.64 (95% confidence interval [CI], 0.61–0.66). By institution, the minimum and the maximum percents of agreement when all negatives were included were 62.0% and 95.0%, respectively; when all negatives were excluded, the minimum and the maximum overall percents of agreement were 48.2% and 84.1%, respectively. In stage 2, 83.7% of cases were a perfect match, and 87.4% were a partial match (Table 3). The κ values were 0.159 (95% CI, 0.12–0.20) and 0.389 (95% CI, 0.35–0.43), respectively. By institution, the percentage of perfect matches ranged from 57.8% to 97.5%, and the percentage of partial matches ranged from 73.0% to 98.7%.

The mean quality assessment scores were 3.18 of 4 points in stage 1 and 3.05 of 4 points in stage 2 (Table 4). By institution, the minimum and maximum scores in stage 1 were 2.93 and

TABLE 1. Frequency and Sensitivity of Screening Criteria

Screening Criteria	Patients of Second Stage Review, n (%)	Sensitivity, %	Positive Predictive Value, %
[General: admission, discharge, and transfer]			
1. The index admission was an unplanned admission related to previous healthcare management	765 (5.1)	10.2	13.1
2. Length of index was more than 30 d	1667 (11.2)	34.9	28.6
3. Unplanned readmission after discharge from index admission	798 (5.4)	14.5	17.1
4. Revisiting emergency department within 72 h after discharge from index admission	180 (1.2)	3.9	18.7
5. Unplanned transfer to another acute care hospital	320 (2.2)	5.2	18.3
6. Temperature higher than 38.3°C at the point of discharge	44 (0.3)	1.1	24.2
7. Unplanned transfer from general care to intensive care	181 (1.2)	7.4	41.0
8. Specialty consult	2262 (15.2)	42.0	19.4
[Events]			
9. Cardiac or respiratory arrest, rapid response team activation	152 (1.0)	5.2	34.2
10. Death in hospital	661 (4.4)	10.1	22.3
11. Hospital incurred patient injury	442 (3.0)	29.3	69.4
12. Oversedation/hypotension	268 (1.8)	9.4	33.5
13. Restraint use	439 (3.0)	11.1	28.8
14. Acute dialysis	11 (0.1)	0.3	25.0
15. In-unit procedure	1329 (8.9)	30.6	26.1
16. Treatment of organ damage after an invasive procedure	19 (0.1)	1.3	71.4
17. Acute myocardial infarction, cerebrovascular accident, or pulmonary embolus during or after an invasive procedure	7 (0.0)	0.4	42.9
[Drug, fluid, blood]			
18. Transfusion or use of blood products	1,463 (9.8)	31.7	30.0
19. Peniramin injection (chlorpheniramine maleate 4 mg) or Chlorpheniramine Maleate Injection Huons (chlorpheniramine maleate 4 mg) use by intramuscular or intravenous route	170 (1.1)	7.4	43.0
20. Abrupt medication stop	853 (5.7)	25.0	28.4
21. Antidotes use	86 (0.6)	0.5	6.6
22. Adverse drug reaction	230 (1.5)	14.4	59.4
[Test result]			
23. Health care–associated infection	351 (2.4)	15.0	43.8
24. Decrease in hemoglobin or hematocrit of 25% or greater	917 (6.2)	26.2	28.9
25. Hypoglycemic symptom	185 (1.2)	5.0	27.2
26. Bleeding tendency	52 (0.3)	2.0	32.6
27. Rising BUN or serum creatine >2 times baseline	441 (3.0)	9.5	19.3
28. *Clostridium difficile*–positive stool	70 (0.5)	4.0	63.8
[Related surgery]			
29. Postoperative troponin level greater than upper normal limit	14 (0.1)	0.7	45.5
30. Mechanical ventilation >24 h postoperatively	14 (0.1)	1.1	80.0
31. Unplanned return to the operating theater	36 (0.2)	2.4	64.3
32. Unplanned removal, injury, or repair of organ during surgery	2 (0.0)	0.1	50.0
33. Intraoperative epinephrine, norepinephrine, naloxone, or romazicon	127 (0.9)	2.8	22.1
34. Unplanned change in procedure or surgery	22 (0.1)	0.7	29.4
35. Intubation, reintubation, BiPap in postanesthesia care unit	2 (0.0)	0.0	0.0
36. X-ray in postanesthesia care unit	1 (0.0)	0.0	0.0
[Obstetrics and neonates]			
37. β_2 Agonist use in obstetrics	0 (0.0)	—	—
38. Oxytocic agents in obstetrics	116 (0.8)	0.3	3.1
39. Complications of neonatal, abortion, amniocentesis or labor and delivery	0 (0.0)	—	—
[Others]			
40. Documentation or correspondence indicating litigation, dissatisfaction	49 (0.3)	1.6	28.6
41. Any other undesirable outcomes not covered above	125 (0.8)	7.9	51.3
Total	14,871 (100.0)	—	17.9

TABLE 2. Reliability of First Reviewers by Medical Institution

Medical Institution	Overall Percent Agreement (Including All Negative Findings)	Overall Percent Agreement (Excluding All Negative Findings)	κ Value	95% CI
A	92.4	83.4	0.84	0.76–0.93
B	62.0	48.2	0.31	0.24–0.38
C	90.8	68.9	0.75	0.67–0.84
D	85.8	76.1	0.72	0.63–0.80
E	86.0	74.2	0.72	0.63–0.81
F	71.4	49.6	0.42	0.34–0.51
G	74.4	59.2	0.49	0.41–0.58
H	83.8	64.8	0.66	0.57–0.74
I	78.6	68.8	0.57	0.48–0.65
J	83.0	63.4	0.64	0.56–0.73
K	83.2	70.2	0.66	0.58–0.75
L	85.4	76.2	0.71	0.62–0.79
M	72.0	49.6	0.45	0.37–0.53
N	95.0	84.1	0.88	0.79–0.97
O	84.0	81.0	0.57	0.49–0.65
Total	81.9	67.3	0.64	0.61–0.66

3.52, respectively, and the scores in stage 2 were 2.72 and 3.30, respectively. The main reason for a negative assessment in stage 1 was insufficient description for identifying patient condition and inadequate medical records, incorrect entry on procedure and treatment, and differences in details between medical records. The majority of reviewers in stage 2 responded that identifying adverse events was difficult because of insufficient description of patient conditions and inadequate medical records.

In stage 1, it took an average of 13.02 minutes to review one patient record, with a range of 6.35 to 52.50 minutes (Table 5). Regardless of the completion of a detailed adverse event report, the mean review time for in stage 2 was 5.06 minutes, with a range of 0.62 to 22.5 minutes. In addition, the time taken to complete a detailed adverse event report was 9.47 minutes on average, with a range of 3.39 to 29.04 minutes.

DISCUSSION

This study examined the feasibility of the KNPSII survey conducted in accordance with the Patient Safety Act of Korea.

TABLE 3. Reliability of Second Reviewers by Medical Institution

	Patients Who Progressed to Second Stage Review	Perfect Match*			Partial Match†		
Medical Institution	n	Overall Percent Agreement	κ Value	CI	Overall Percent Agreement	κ Value	CI
A	229	79.5	0.12	−0.03 to 0.27	83.8	0.40	0.25 to 0.55
B	367	83.7	0.09	−0.02 to 0.20	85.8	0.27	0.14 to 0.41
C	148	80.4	−0.08	−0.13 to −0.04	87.2	0.45	0.24 to 0.65
D	297	94.3	0.08	−0.11 to 0.27	95.3	0.34	0.08 to 0.60
E	271	81.9	0.21	0.06 to 0.36	86.0	0.46	0.31 to 0.60
F	284	82.4	0.05	−0.08 to 0.18	83.8	0.18	0.03 to 0.33
G	314	80.3	0.13	0.00 to 0.27	82.8	0.30	0.16 to 0.44
H	230	57.8	0.01	−0.09 to 0.11	73.0	0.41	0.29 to 0.53
I	343	83.4	0.06	−0.06 to 0.19	84.5	0.17	0.03 to 0.31
J	232	79.3	0.38	0.24 to 0.52	84.5	0.57	0.45 to 0.70
K	282	97.5	−0.01	−0.02 to 0.00	97.5	−0.01	−0.02 to 0.00
L	307	80.8	0.09	−0.03 to 0.21	82.1	0.19	0.06 to 0.32
M	278	95.0	0.28	0.02 to 0.54	95.3	0.36	0.10 to 0.63
N	157	96.8	0.27	−0.17 to 0.71	98.7	0.79	0.52 to 1.07
O	420	90.5	0.36	0.21 to 0.52	91.7	0.48	0.33 to 0.62
Total	4159	84.6	0.16	0.12 to 0.20	87.4	0.39	0.35 to 0.43

*Agreement in all of the opinions for the occurrence of an adverse event, the degree of harm, the causality, and the preventability.

†Agreement on the decision for adverse event but disagreement in the degree of harm, causality, or preventability.

TABLE 4. Quality of Medical Record by Medical Institution

Medical Institution	First Stage Review, Mean ± SD	Second Stage Review*, Mean ± SD
A	3.12 ± 0.33	3.21 ± 0.41
B	2.93 ± 0.27	2.72 ± 0.48
C	3.06 ± 0.24	3.30 ± 0.56
D	3.51 ± 0.50	3.05 ± 0.38
E	3.10 ± 0.31	3.01 ± 0.11
F	3.37 ± 0.48	3.10 ± 0.37
G	3.06 ± 0.34	2.90 ± 0.30
H	3.52 ± 0.50	3.25 ± 0.46
I	3.00 ± 0.05	2.99 ± 0.11
J	3.00 ± 0.15	2.86 ± 0.37
K	3.39 ± 0.54	2.67 ± 0.50
L	3.02 ± 0.22	3.00 ± 0.00
M	3.48 ± 0.50	3.05 ± 0.21
N	3.24 ± 0.44	3.00 ± 0.00
O	3.00 ± 0.03	2.99 ± 0.11
Total	3.19 ± 0.42	3.05 ± 0.39

*Detailed adverse event reports only.

We examined the validity of the screening criteria, interreviewer reliability, medical record quality assessment, and the time taken for medical record review. For the method of reviewing medical records to identify adverse events to remain the criterion standard in examining patient safety status, it must be able to yield valid and reliable results; this study is significant in that it provides relevant comprehensive data about this.

In the first national-level KNPSII survey conducted in Korea, 41 screening criteria were used. This is greater than the 18 criteria used in Harvard Medical Practice Study–related studies but less than the number used in the Global Trigger Tool.[19] Increasing the number of screening criteria may not only increase the sensitivity of identifying adverse events but also increase the workload for stage 1 reviewers; therefore, it is important to maintain an appropriate number of screening criteria.[18] Thus, it is necessary to exclude screening criteria that are less useful. The usefulness of screening criteria must be determined based on the sensitivity and positive predictive value in addition to incidence. In this study, despite the low incidence and sensitivity of some screening criteria, those with high positive predictive values cannot simply be eliminated based on low incidence and sensitivity (e.g., treatment of organ damage after an invasive procedure). However, exclusion from the list of screening criteria or the revision of details of the criterion can be considered for the screening criteria with low incidence, sensitivity, and positive predictive value (e.g., antidotes use).

The reliability among stage 1 reviewers found in this study was similar to that in previous studies,[19] and the variation in the reliability among reviewers by institution also does not seem to be great, suggesting that the results show considerable reliability. In the KNPSII survey, the need for a stage 2 review was determined by pooling the results from the independent review by stage 1 reviewers; if considerable reliability among reviewers can be ensured, it seems possible to split the medical records among reviewers to reduce the workload for stage 1 reviewers.

Nevertheless, in this study, the reliability among stage 2 reviewers was lower than among stage 1 reviewers. Although the overall percent agreement on adverse events was similar to that from a previous study (≥80%),[19] the reliability for overall percent agreement, including the cases with no adverse events, could be overestimated. In other words, it would be more valid to use the κ value rather than the overall percent agreement to evaluate the reliability in this case, and the κ value of this study (including partial match) was found to be less than 0.4, which is not very good.[26] Specifically, it was very rare to see all agreements among the decisions regarding adverse events, degree of harm, causality, and preventability. Although agreement in the assessment of preventability of adverse events has been reported in previous studies,[19] no studies, to our knowledge, have reported reliability that takes the assessment of causality or the degree of harm of adverse events into account. It would be important to prepare measures to improve reliability among stage 2 reviewers. It is believed that the reliability between reviewers can be increased by conducting more training using real cases and including the interim audit in the review process. Furthermore, it seems necessary to add another step for discussion among stage 2 reviewers or introduce stage 3 reviews.[27,28]

For medical record review to be used to identify adverse events,[3,4] the medical records must be completed in full. However, very few studies examining medical record reviews for the identification of adverse events have reported the quality of medical records. In this study, stage 1 and 2 reviewers were asked to assess the quality of medical records on a scale of 4, and the overall score was relatively good with a mean score of 3.19 and 3.05 for the stage 1 and 2 reviewers, respectively. However, some reviewers gave negative assessments and the reasons included incorrect entry on procedure and treatment, inconsistencies in the details of medical records, and inadequate information in medical records. Furthermore, the score from stage 2 reviewers was lower than that from stage 1 reviewers, which seems to suggest that the quality of medical records has a direct effect on the determination of adverse events. Therefore, future studies need to assess the quality of medical records in more detail. First, questions that can measure the essential requirements for medical records to identify adverse events must be developed.[29,30] In future studies, it would be meaningful to examine whether there is a difference in the percentage of adverse events identified and in the distribution of preventability scores between the medical records of certain quality and those that were not.

TABLE 5. Review Time of First and Second Stage Review by Medical Institution

Medical Institution	First Stage Review, Mean ± SD	Second Stage Review, Mean ± SD	Detailed Adverse Event Reports, Mean ± SD
A	7.98 ± 12.90	3.96 ± 45.23	9.57 ± 8.59
B	6.35 ± 8.18	4.50 ± 5.86	6.54 ± 5.48
C	8.46 ± 21.15	2.02 ± 3.73	7.18 ± 6.39
D	10.68 ± 21.83	0.68 ± 4.10	11.47 ± 10.51
E	12.57 ± 12.61	3.05 ± 5.22	7.17 ± 9.76
F	8.92 ± 17.30	0.57 ± 1.41	3.39 ± 1.66
G	9.54 ± 16.23	22.50 ± 31.56	29.04 ± 45.31
H	17.96 ± 30.82	2.87 ± 4.75	5.45 ± 5.44
I	16.75 ± 20.61	5.51 ± 18.19	20.01 ± 49.53
J	8.32 ± 6.92	5.41 ± 9.93	8.87 ± 7.58
K	6.84 ± 6.35	0.62 ± 2.42	6.44 ± 3.13
L	9.71 ± 9.98	3.65 ± 6.66	9.29 ± 5.05
M	16.43 ± 33.48	14.95 ± 41.84	18.00
N	52.50 ± 130.40	1.71 ± 3.00	6.30 ± 4.50
O	8.39 ± 11.03	2.22 ± 4.94	6.60 ± 4.36
Total	13.02 ± 37.63	5.06 ± 19.10	9.47 ± 20.12

The time required for medical record review reported in this study could be used in deciding the number of reviewers needed in similar future studies, in addition to being a realistic consideration for determining the sample size for future medical record reviews. Because there are no previous studies, to our knowledge, that have reported the time taken for medical record review to identify adverse events, it is difficult to determine if our results are representative. However, it seems realistically feasible that more medical records can be reviewed per day than recommended by the World Health Organization, which suggests that no more than 30 cases be reviewed per day per reviewer.[31] However, it will be necessary to monitor not only the fatigue of the reviewer but also the reliability within the reviewer.

This study has several limitations. First, the KNPSII survey was conducted among the regional public hospitals in Korea with very few obstetrics and newborn patients, requiring caution when determining the usefulness of the screening indices related to obstetrics and neonates. For example, although "oxytocic agents in obstetrics" also showed low sensitivity and positive predictive value in this study, it seems that additional results from hospitals specialized for delivery may be necessary to determine its usefulness. Second, because the reviewers were medical professionals at the corresponding institutions, it is possible that the quality of medical records was overestimated. It is necessary for independent reviewers to check interrater reliability in future studies. Third, because patients who did not have any of the screening criteria in the stage 1 review were not reviewed in stage 2, it is difficult to analyze the screening criteria, sensitivity, and specificity of stage 1. Fourth, although the reviewers were instructed to start the review after first entering the reviewer information into the KNPSII survey system, some reviewers mentioned that this instruction was not followed. The review time was measured starting at the time of the first entry, so when the instructions were not followed, it is possible that the overall review time was underestimated.

CONCLUSIONS

In this study, the validity of screening criteria used in the process of medical record review for the identification of adverse events, interreviewer reliability, assessment of quality of medical records, and the time taken for the review of medical records was examined. The results provide foundational data that can be used to improve how medical records are reviewed for the identification of adverse events. Specifically, the results can be used in selecting screening criteria and improving reliability when conducting similar studies in other countries.

REFERENCES

1. Bates DW, Larizgoitia I, Prasopa-Plaizier N, et al. Research Priority Setting Working Group of the WHO World Alliance for Patient Safety. Global priorities for patient safety research. *BMJ*. 2009;338:b1775.
2. Baines R, Langelaan M, de Bruijne M, et al. How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time. *BMJ Qual Saf*. 2015;24:561–571.
3. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *J Gen Intern Med*. 2003;18:61–67.
4. Murff HJ, Patel VL, Hripcsak G, et al. Detecting adverse events for patient safety research: a review of current methodologies. *J Biomed Inform*. 2003; 36(1–2):131–143.
5. Stavropoulou C, Doherty C, Tosey P. How effective are incident-reporting systems for improving patient safety? A systematic literature review. *Milbank Q*. 2015;93:826–866.
6. Klein DO, Rennenberg RJMW, Koopmans RP, et al. A systematic review of methods for medical record analysis to detect adverse events in hospitalized patients. *J Patient Saf*. 2021;17:e1234–e1240.
7. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324:370–376.
8. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med*. 1991;324:377–384.
9. Wilson RM, Runciman WB, Gibberd RW, et al. The quality in Australian Health Care Study. *Med J Aust*. 1995;163:458–471.
10. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ*. 2004;170:1678–1686.
11. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ*. 2001;322: 517–519.
12. Davis P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals I: occurrence and impact. *N Z Med J*. 2002;115:U271.
13. Michel P, Quenon JL, Djihoud A, et al. French national survey of inpatient adverse events prospectively assessed with ward staff. *Qual Saf Health Care*. 2007;16:369–377.
14. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Murillo J, et al. Incidence of adverse events related to health care in Spain: results of the Spanish National Study of Adverse Events. *J Epidemiol Community Health*. 2008; 62:1022–1029.
15. Sousa P, Uva AS, Serranheira F, et al. Patient and hospital characteristics that influence incidence of adverse events in acute public hospitals in Portugal: a retrospective cohort study. *Int J Qual Health Care*. 2018;30: 132–137.
16. Smits M, Langelaan M, de Groot J, et al. Examining causes and prevention strategies of adverse events in deceased hospital patients: a retrospective patient record review study in the Netherlands. *J Patient Saf*. 2021;17: 282–289.
17. Nilsson L, Borgstedt-Risberg M, Soop M, et al. Incidence of adverse events in Sweden during 2013–2016: a cohort study describing the implementation of a national trigger tool. *BMJ Open*. 2018;8:e020833.
18. Ock M, Lee SI, Jo MW, et al. Assessing reliability of medical record reviews for the detection of hospital adverse events. *J Prev Med Public Health*. 2015;48:239–248.
19. Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open*. 2016;6:e011078.
20. Kim S, Shin HG, Jo AEJ, et al. Variation between hospitals and reviewers in detection of adverse events identified through medical record review in Korea. *Int J Qual Health Care*. 2020;32:495–501.
21. Hwang JI, Chin HJ, Chang YS. Characteristics associated with the occurrence of adverse events: a retrospective medical record review using the Global Trigger Tool in a fully digitalized tertiary teaching hospital in Korea. *J Eval Clin Pract*. 2014;20:27–35.
22. Hwang JI, Kim J, Park JW. Adverse events in Korean traditional medicine hospitals: a retrospective medical record review. *J Patient Saf*. 2018;14: 157–163.
23. Kim MJ, Seo HJ, Koo HM, et al. The Korea National Patient Safety Incidents Inquiry (KNPSII): Characteristics of adverse events identified through medical record review in regional public hospital. *J Patient Saf*. In press.
24. Patient Safety Act. Available at: https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=167782&viewCls=engLsInfoR&urlMode=engLsInfoR#0000. Accessed on Oct 4, 2021.
25. Korea Institute for Healthcare Accreditation. *Development of Korean Patient Safety Incident Information Collection Model and Study on Demand Analysis of Medical Safety R&D Service Report*. Seoul, Korea: Korea Health Industry Development Institute; 2019.

26. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33:363–374.

27. Soop M, Fryksmark U, Köster M, et al. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. *Int J Qual Health Care*. 2009;21:285–291.

28. Klein DO, Rennenberg RJMW, Koopmans RP, et al. Adverse event detection by medical record review is reproducible, but the assessment of their preventability is not. *PLoS One*. 2018;13:e0208087.

29. Hammer A, Wagner A, Rieger MA, et al. Assessing the quality of medication documentation: development and feasibility of the MediDocQ instrument for retrospective chart review in the hospital setting. *BMJ Open*. 2019;9:e034609.

30. Halfon P, Staines A, Burnand B. Adverse events related to hospital care: a retrospective medical records review in a Swiss hospital. *Int J Qual Health Care*. 2017;29:527–533.

31. World Health Organization. *Assessing and Tackling Patient Harm: A Methodological Guide for Data-Poor Hospitals*. Geneva, Switzerland: WHO; 2010.