Detecting Adverse Events in Thai Hospitals Using Medical Record Reviews: Agreement among Reviewers

Pattapong Kessomboon, MD*,
Supasit Panarunothai, MD**, Pradit Wongkanaratanakul, MD***

** Department of Community Medicine, Khon Kaen University, Khon Kaen*
*** Center for Health Equity Monitoring, Naresuan University, Phitsanulok*
**** National Health Security Office, Nonthaburi*

Background: *Attention to the problem of adverse events increases worldwide. The present study aimed to evaluate inter-rater reliability of medical record reviews of adverse events in the Thai context.*
Material and Method: *A total of 279 inpatient records were reviewed by 23 clinical auditors. Each record was examined independently by 3 auditors using a standardised review form. Agreements on the occurrence of AEs among auditors were assessed using Kappa statistic (k).*
Results: *Agreement of the auditors in detecting at least 1 medical condition potentially related to AEs was moderate (k = 0.54, 95% confidence interval (CI) 0.42-0.65). Agreement on each item of the listed conditions ranged from slight to substantial (k = 0.08 to 0.79). There was a positive trend of the correlation between the rate of the conditions and the kappa statistic (Spearman rank correlation = 0.65, p-value = 0.058). Agreement on determining AEs occurrence was fair (weighted k = 0.34, CI = 0.22-0.45). Agreement on determining AEs preventability was fair (weighted k = 0.27, CI = 0.16-0.39).*
Conclusion: *The reliability of medical record review to detect AEs is influenced by the prevalence rate of AEs, as well as the variability among reviewers. The use of chart review to detect AE is reasonable if the rate of the event is sufficiently high.*

Keywords: *Adverse event, Medical record review, Agreement*

J Med Assoc Thai 2005; 88 (10): 1412-8
Full text. e-Journal: *http://www.medassocthai.org/journal*

Adverse event (AE) among hospitalised patient is one of the major healthcare problems receiving more attention in many countries[1-7]. Studies have shown that the rates of AEs could be reasonably estimated[1-6]. Information on types and related factors of AEs can also be uncovered leading to the development of relevant preventive measures[8,9]. In the United States the prevalence of adverse events and deficiencies of the malpractice litigation system have led to increasing interest in establishing no-fault compensation programs[10,11].

The newly enacted 2002 Universal Health Coverage Act in Thailand recognizes the problems of AEs and designates 1% of the universal health coverage budget available for compensating the victims of AEs to alleviate suffering during the delayed claim processes, or in case no one was found guilty[12]. Article 42 of the Act also mentions that once the compensation is paid, the National Health Security Office can charge the provider who was identified to be the cause of AEs[12]. The Article activated widespread protests among medical doctors mainly due to the lack of confidence regarding the judgement on identification of faults. The government finally promised to amend the Article in favour of the doctors, however, the promise has not yet been substantiated.

The experiences of lawsuits of malpractice among Thai physicians were not uncommon, about 5% of the physicians surveyed had been sued at least once by patients or relatives[13]. The Thai Medical

Correspondence to : Pannarunothai S, Centre for Health Equity Monitoring, Faculty of Medicine, Naresuan University, Phitsanulok 65000, Thailand. E-mail: supasitp@nu.ac.th

Council reported the increasing number of malpractice accusations on Thai physicians from 251 cases for the whole year in 2002, up to 332 cases in the first month of 2003[14]. These figures suggest an urgent need for Thailand to systematically evaluate the situation in order to understand the nature and related factors of the problem so that appropriate preventive measures and policies can be developed.

Experiences from other countries provide valuable lessons in terms of available methodologies for investigating AEs. Yet, there is no gold standard test for estimating adverse event rates. Most studies are based on review of medical records using guided implicit jugdement[1-4], with limitations and criticisms on validity and repeatability[15-17].

In Thailand, the Centre for Healthcare Information (CHI), an independent public organization working in collaboration with major healthcare funding agencies, has extensive experience on medical record audits. Medical record review has been successfully used as mechanisms for audits on medical coding and reimbursement. The needs to extend to quality of care audits have been suggested[18]. This study was an attempt to explore the possibility of quality of care audit for the Thai health system. Rates and types of AEs are being explored.

Prior to jumping into a large-scale study,the authors need some experiences and deep understanding of the properties of the tools and procedures for studying AEs. The present study aimed to evaluate the reliability of medical review for determination of AEs which occurred among hospitalized patients by assessing the agreement among doctor auditors in determining AEs occurrence. Recommendations for further uses of the method in Thailand could then be formulated.

Material and Method

Medical record review by medical auditors, using guided implicit judgement form, was the main mode of data collection. 400 medical records were selected from a pool of medical records collected by the Centre for Healthcare Information.The authors purposively selected medical records of patients with high possibility of AE occurrence to focus the evaluation to the variability of reviewers' judgement and the process of detecting AEs. Inclusion criteria were medical records of patients who died during hospital stay, or having the relative weight by Diagnosis Related Group (DRG) of 3 or higher, or having hospital charge of 50,000 Baht (US$ 1,250) or more. Exclusion criteria were medical records that contained incomplete documents or low readability due to poor quality photocopying.

Twenty-three medical doctors were recruited for the reviews. All reviewers were regular auditors of the CHI. All have more than 10 years experience of medical practice at major regional hospitals in Thailand. Their specialty training backgrounds ranged from internal medicine, to general surgery, orthopaedics, paediatrics, obstetrics and gynaecology, neurosurgery and anaesthesiology.

A one-day training session was held for these reviewers. After having presented the objectives and detailed study procedures, case exercises were used for practicing of AE determination. The case exercise consisted of a case summary and a real medical record. Discussions among the reviewers led to suggestions for some adjustments of the guided review form, adding more examples of clinical conditions included into the list. The final version was a 2-page standardised review form containing 13 clinical conditions potentially related to AEs (Table 2).

The definition of AE in the present study was the same as in previous studies[1-3]. An *"adverse event"* was defined as "an unintended injury caused by medical management rather than by the disease process, this injury is sufficiently serious to lead to a prolongation of hospitalization or a temporary or permanent impairment or disability or death to the patient at time of discharge". The *"preventability of an AE"* is defined as "an error in management or the failure to follow the recommended practice at an individual or system level". The recommended practice is defined as "the current level of expected performance for the average practitioner or system that manages the condition in question".

Three reviewers independently reviewed each record, 23 reviewers made a total of 837 reviews on 279 records. Each auditor reviewed on average 36 medical records. The reviews were randomly divided into three groups to compare the agreement among three auditors, because no auditor could be taken as a gold standard as specified in other studies[1-3]. Degree of agreements made by auditors were calculated and presented.

After detecting the presence of clinical conditions listed in the guided review form, the reviewers were asked to make decisions whether any AE(s) had occurred according to the given definition. They rated their confidence on a scale of 1 to 6, where 1 was that the AEs were least likely to occur, and 6 was that AEs were the most likely to occur. The same scale was used for decision making on the preventability of the AEs.

Kappa statistic (k) was analyzed to evaluate inter-rater agreement for binary data. Weighted kappa (k_w) and intra-class correlation coefficient (ICC) were computed for ordinal data (for the 1-6 scale rating of confidence in making decisions about AEs occurrence and preventability)[19-24]. The weight (w_{ij}) was calculated according to the following formula [24]: $w_{ij} = 1 - \{(i - j)/(k - 1)\}^2$; where i is an index of row, j is an index of column and k is 6.

Sample size was calculated using a formula proposed by Norman and Streiner[22]: $N = Z^2 * P_o(1 - P_o) /d^2 (1 - P_e)^2$, where P_o and P_e were derived from a pilot study as 0.86 and 0.77 respectively. Z was 1.96 and d (accepted distance from k) was 0.15. The required sample size was 388; the present study began with selection of 400 cases sent to reviewers.

Interpretation of the kappa statistic is based on the criteria proposed by Landis and Kock[23]. The agreement was defined as *poor* if kappa was less than 0, *slight* if kappa was 0-0.19, *fair* 0.2-0.39, *moderate* 0.4-0.59, *substantial* 0.6-0.79 and *almost perfect* 0.8-1.0.

Results

Altogether, 279 medical records (70%) were completed with 3 reviews. The patients' median age was 67 years with interquartile range (IQR) of 55-75 years. 53.8% of them were male. The rate of AEs was as high as 35% of total (AE was positive if at least 2 out of 3 reviewers agreed). The high rate reflected the purposive selection criteria of the study. The drug AE was only 2.5% while AEs resulted in death was 15%. The rate of preventable AEs was as high as 19% (Table 1).

Clinical conditions listed in the guided review form and detected by at least 2 reviewers are described in Table 2. Nosocomial infection was the most frequent condition found (30% of the total). Deaths were found among 19% of total (but only 15% were related to AEs). Four clinical conditions in the list not found in the present study were operation on the wrong organ, fall, obstetrics or gynaecological complication and APGAR score at 5 minutes less than 6.

Agreement on detecting listed conditions

Agreements among 3 reviewers in detecting the clinical conditions listed in the guided review form are shown in Table 3. Substantial agreement was found for detecting death condition (k = 0.79), but was only slight for detecting the removal of an organ (k = 0.16). The overall agreement of detecting at least one clinical condition was moderate (k = 0.54).

Table 1. The occurrence of AEs

Occurrence*	No.	% of 279 (95%CI)
AE	98	35.1 (29.5-40.8)
Drug AE	7	2.5 (0.7-4.4)
AE Death	43	15.4 (11.1-19.7)
Preventable	53	19.0 (14.4-23.6)

*The occurrence was defined as positive if detected by at least 2 reviewers

Table 2. Clinical conditions detected

Item	Clinical conditions*	No. (%)
1	Nosocomial infection	86 (30.5)
2	Death	53 (19.0)
3	Treatment complication	41 (14.7)
4	Pressure sore	12 (4.3)
5	Drug adverse event	10 (3.6)
6	Cardiac arrest	8 (2.9)
7	Neurological deficit	4 (1.4)
8	Removal of organ	1 (0.4)
9	Others	9 (3.2)
10	Operation on wrong organ	0
11	Fall	0
12	OB-GYN complication	0
13	APGAR score < 6 at 5 min.	0

* The occurrence was defined as positive if detected by at least 2 reviewers

Spearman rank correlation of kappa statistic (of data in Table 3) and the rate of clinical conditions (data in Table 2) was 0.65 (p-value = 0.058) indicating trend of positive correlation between the two variables.

Agreement on determining AEs occurrence and preventability

The agreements on medical audit between 3 reviewers are presented in Table 4. The weighted kappa(s) of agreement on determining AE occurrence and preventability were *fair* (k_w = 0.34 and 0.27 respectively), but intra-class correlation coefficient(s) of both decisions were *moderate* (ICC = 0.60 and 0.47 respectively).

Discussion and Conclusion

The present study was conducted to explore the methodology for detecting medical adverse events by independent medical reviews. Four hundred medical records were sent out but the returns could match 279 cases with complete 3 review results. The response rate of 70% should somewhat be considered sufficient.

Table 3. Agreement on detecting each of the listed conditions

Item*	Kappa	SE
Nosocomial infection	0.51	0.059
Death	0.79	0.059
Treatment complication	0.25	0.058
Pressure sore	0.26	0.059
Drug adverse event	0.20	0.059
Cardiac arrest	0.36	0.059
Neurological deficit	0.24	0.059
Removal of organ	0.16	0.056
Others	0.08	0.059
Operation on wrong organ	n.a.	n.a.
Fall	n.a.	n.a.
OB-GYN complication	n.a.	n.a.
APGAR score < 6 at 5 min.	n.a.	n.a.
At least 1 condition detected	0.54	0.059

* Item description is the same as Table 2

Table 4. Agreement on determining AEs occurrence and preventability

Compared groups of reviews*	k_w	SE
AE occurrence		
Group1 vs Group2	0.30	0.059
Group1 vs Group3	0.39	0.059
Group2 vs Group3	0.32	0.059
Combined kappa	0.34	0.059
ICC = 0.60 (95%CI 0.52-0.68)		
Preventability		
Group1 vs Group2	0.24	0.059
Group1 vs Group3	0.34	0.059
Group2 vs Group3	0.24	0.059
Combined kappa	0.27	0.059
ICC = 0.47 (95%CI 0.17-0.68)		

* 837 Reviews of 279 records by 23 auditors were randomly divided into 3 groups for the purpose of comparisons

Moreover, the present study used 3 reviewers for each record instead of 2; this led to a sophisticated agreement analysis as present in method section. The present study has revealed important findings regarding the properties of the medical review for determining AE occurrence when used in the Thai context.

Detection of clinical conditions potentially related to AEs

The overall rate of AEs occurrence, determined by at least 2 reviewers, was high (35%) reflecting the selection criteria in which high-risk cases were purposively selected for this study.

Other studies gave the lower rates of AEs ranging from 3.7 to 16.6%[1-4]. The reasons can be explained by the differences of selection procedures and the specific review processes of each study as well as the contextual variations.

The precision of the estimation of AEs occurrence using medical record reviews has been challenged[15,25,26]. However, the evidence suggests that the problem of AEs is sizable and warrants serious attention.

A recent study in France has demonstrated that a retrospective medical record review provided similar estimates of rates of AEs derived from a prospective study design was better than the result from a cross-sectional study. It has also some advantages in terms of costs and related workload[6].

Correlation between rate of AE and agreement

The present study showed a large variation of the rates of each of the listed clinical conditions potentially related to AEs, ranging from 0 to 30.5%. The Kappa statistic (k) for the agreement of reviewers on the detection of each clinical condition also varied from 0.08 to 0.79.

A striking finding was the positive correlation between the Kappa statistic and the rate of each clinical condition. This suggests that the prevalence rate of the condition under study potentially influences the agreement among reviewers.

Agreement on determination of AE occurrence and preventability of the AEs

The present study has shown that the agreement of medical record review among Thai doctor auditors on making decisions whether the listed conditions detected indicating the occurrence of AEs and whether they are preventable were *fair* to *moderate*. Determination on the preventability of an AE is less reliable than determining AE occurrence.

Previous studies using medical reviews achieved similar or a little bit better agreements with k of AE determination = 0.31-0.57, however k of AE preventability = 0.19-0.33[3,6,15].

It has been demonstrated that better reliability of determination on preventability can be achieved if prospective mode of data collection is employed[6,7] and the scope of conditions or events under study is limited, e.g. only drug adverse events[9].

A combined method of data collection using prospective telephone interview, chart review on read-

mission and provider interview has revealed a substantial reliability (k = 0.61) of determination of AEs occurring among patients after hospital discharge[7]. The prospective mode is also superior for identifying preventable AEs (k = 0.44)[6,7]. However, it has limitations for a large scale implementation.

For the present study, the investigators have attempted to standardize the guided review processes by holding a training session with case exercises. However, the reliability among the Thai auditors was still low. A previous study has shown that discussion among reviewers does not improve the reliability[17]. Because the medical review is based on implicit judgement of reviewers, therefore it has a tendency of low agreement. However, the approach has advantages. It is appropriate for detecting a wide range of conditions that needs expertise and unexpected conditions can be better detected than the rigid criteria[27]. Detecting AEs lies in this situation.

For the present study, this was the first experience of quality audit for the reviewers. The authors believe that more experiences gained should contribute to a better audit. Exchange of knowledge and experiences gained from the first round among reviewers across different disciplines and among the investigators should also enhance the learning curve, particularly if the processes are conducted in the context of care improvement rather than detection of faults.

Further efforts to improve the reliability should be attempted to differentiate the study designs by limiting the range of potential clinical conditions to be included in the review form particularly the serious ones, limiting types of settings such as studying only operation related AEs. However, the primary purpose of any attempt must first be clarified. Administrators at a national organization might be more interested in seeing the big picture of the problem rather than a subset one.

Studies to compare different modes of data collection, e.g. prospective vs. retrospective, not only in terms of reliability but also on costs, effects on providers' attitudes and organizations' behaviours related to quality of care in each country's context should be supported.

In conclusion, the present study revealed the property of medical record review used to determine AEs among hospitalized patients in Thailand. The reliability of the process was somewhat comparable to other studies. The rates of conditions seem to influence the agreement among reviewers.

Acknowledgements

The authors wish to thank are grateful to all supporters and funders of the project. All reviewers, Professor Ammar Siamwala, Dr. Wiput Poolcharoen, Health Systems Research Institute, Centre for Healthcare Information, WHO Thailand, Department of Central Accounting, colleagues at the Department of Community Medicine, Khon Kaen University, Centre for Health Equity Monitoring, Naresuan University and Epidemiology Unit, PSU. The first author is supported by the Thailand Golden Jubilee Programme for PhD study under the Thailand Research Fund.

References

1. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalised patients. Results of the Harvard medical practice study I. N Engl J Med 1991; 324: 370-6.
2. Leape LL, Brennan TA, Laird NM, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalised patients: results of the Harvard medical practice study II. N Engl J Med 1991; 324: 377-84.
3. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. Med J Aust 1995; 163: 458-71.
4. Vincent C, Neale G, Woloshynowych. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001; 322: 517-9.
5. David P, Lay-Yee R, Schug S, Briant R, Scott A, Johnson S, et al. Adverse events regional feasibility study: indicative findings. N Z Med J 2001; 114: 203-5.
6. Michel P, Quenon JL, Sarasquesta AM, Scemama O. Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. BMJ 2004; 328: 199.
7. Foster AJ, Clark HD, Menard A, Dupuis N, Chernish R, et al. Adverse events among medical patients after discharge from hospital. CMAJ 2004; 170: 345-9.
8. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. Surgery 1999; 126: 66-75.
9. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events: implications

for prevention. JAMA 1995; 274: 29-34.
10. Brennan TA, Localio AR, Leape LL, Laird NM, Peterson L, Hiatt HH, et al. Identification of adverse events occurring during hospitalization: a cross-sectional study of litigation, quality assurance, and medical records at two teaching hospitals. Ann Intern Med 1990; 112: 221-6.
11. Brennan TA. Medical injuries: International perspectives. Med J Aust 1995; 163: 475-6.
12. National Health Insurance Act. Office of the National Health Insurance. Nontaburi. 2002.
13. Ungprapan V. Problems of malpractice law suits in Thailand. 2000. In Thai Medical Council. A summary report of the committee on medical malpractice. 14 June-31 October 2000.
14. Thai Medical Council. Annual Report. The Thai Medical Council. Nontaburi. 2004.
15. Thomas FJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record reviews for estimating adverse event rates. Ann Intern Med 2002; 136: 812-6
16. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001; 286: 415-20.
17. Hofer TP, Berstein SJ, DeMonner S, Hayward RA. Discussion between reviewers does not improve reliability of peer review of hospital quality. Medical Care 2000; 38: 152-61.
18. Veera Ingkapasakorn. Medical record audit: foundation for clinical quality in Thailand. J Health S 2001; 10: 42-53.
19. Cohen J. A coefficient of agreement for nominal scales. Educational Psychological Measurement 1960; 20: 37-46.
20. Cohen J. Weighted kappa: Nominal scale agreement with provision for scale disagreement or partial credit. Psychological Bulletin 1968; 70: 213-20.
21. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York: John Wiley & Sons. 1981.
22. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 2nd ed. Halmilton: B.C. Decker Inc. 2000; 220-2.
23. Landis JR, Kock GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74. Cited by Feinstein AR. Principles of Medical Statistics. Boca Raton: Chapman & Hall/CRC. 2002.417.
24. Feinstein AR. Principles of Medical Statistics. Boca Raton: Chapman & Hall/CRC. 2002, 417-8.
25. Brennan TA. The Institute of Medicine report on medical errors - Could it do harm? N Engl J Med 2000; 342: 1123-5.
26 McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors: Are they exaggerated in Institute of Medicine report? JAMA 2000; 284: 93-5.
27 Lilford RJ, Mohammed MA, Braunholtz D, Hofer TP. The measurement of active error: methodological issues. Qual Saf Health Care 2003; 12(Suppl II): ii8-ii12.

การค้นหาภาวะไม่พึงประสงค์ในโรงพยาบาลประเทศไทยด้วยการทบทวนเวชระเบียน: ความตรงกันระหว่างผู้ทบทวน

ปัตพงษ์ เกษสมบูรณ์, ศุภสิทธิ์ พรรณารุโณทัย, ประดิษฐ์ วงษ์คณารัตนกูล

ความเป็นมา: *ความสนใจต่อปัญหาภาวะไม่พึงประสงค์มีมากขึ้นในประเทศต่าง ๆ ทั่วโลก การวิจัยนี้เพื่อประเมินความตรงกันระหว่างผู้ทบทวนเวชระเบียน เพื่อค้นหาภาวะไม่พึงประสงค์ในโรงพยาบาลประเทศไทย*

วัสดุและวิธีการ: *แพทย์ 23 คน ทบทวนเวชระเบียนผู้ป่วยใน 279 ฉบับ แต่ละฉบับมีแพทย์ทำการทบทวนอย่างอิสระต่อกัน 3 คน โดยใช้แบบคัดกรองที่เป็นมาตรฐาน ความตรงกันของการสรุปว่าเกิดภาวะไม่พึงประสงค์ระหว่างแพทย์ผู้ทบทวนเสนอด้วยสถิติ Kappa (k)*

ผลการศึกษา: *ความตรงกันระหว่างแพทย์ผู้ทบทวน ภาวะทางคลินิกอย่างน้อย 1 ชนิดที่มีโอกาสทำให้เกิดภาวะไม่พึงประสงค์ ความตรงกันอยู่ในระดับ ปานกลาง (k = 0.54, 95%CI 0.42-0.65) ความตรงกันของแต่ละภาวะทางคลินิก มีการแปรผันตั้งแต่ น้อย จนถึง มาก (k = 0.08 to 0.79) พบว่ามีแนวโน้มทางบวกของความสัมพันธ์ระหว่างอัตราการพบภาวะนั้น กับค่าสถิติ kappa (Spearman rank correlation = 0.65, p-value = 0.058) ความตรงกันในการสรุปว่าเกิด ภาวะไม่พึงประสงค์อยู่ในระดับ พอใช้ (weighted k = 0.34, CI = 0.22-0.45) ความตรงกันในการสรุปว่าภาวะไม่พึงประสงค์นั้นป้องกันได้ อยู่ในระดับ พอใช้ (weighted k = 0.27, CI = 0.16-0.39)*

สรุป: *ความน่าเชื่อถือของการทบทวนเวชระเบียนในการค้นหาภาวะไม่พึงประสงค์ มีปัจจัยอิทธิพลจากความชุกของภาวะไม่พึงประสงค์ และ ความแปรปรวนระหว่างแพทย์ผู้ทบทวน ดังนั้นการนำวิธีทบทวนเวชระเบียนไปใช้ในการค้นหาภาวะไม่พึงประสงค์ควรใช้ในกลุ่มที่อุบัติการณ์การเกิดอยู่ในระดับสูง*